JOURNAL OF CYBERSECURITY

Journal of Cybersecurity, 2025, tyaf028
https://doi.org/10.1093/cybsec/tyaf028

Research Paper

Pseudonymization and reporters' protection by design in the EU whistleblower directive

Mirosław Kutyłowski and Gabriel Wechta*

Cryptology Department, NASK National Research Institute, Kolska 12, 01-045 Warsaw, Poland

*Corresponding author. Gabriel Wechta, Cryptology Department, NASK National Research Institute, Kolska 12, 01-045 Warsaw, Poland. E-mail: gabriel.wechta@nask.pl

Received 27 February 2025; revised 11 September 2025; accepted 1 October 2025

Abstract

The EU Whistleblower Directive aims to create a framework where the persons reporting breaches of EU law are protected against retaliation. In contrast to GDPR, it is mainly based on trust assumptions and not on the concept of privacy and security by design. As we are explicitly dealing with problems of unlawful behavior, this is a critical issue. In this paper, we analyze the role of pseudonymization, the main technical tool promoted in the GDPR, within the Whistleblower Directive. To see the real impact of the Directive, we analyze how these issues are reflected in the national law in Germany and Poland. We show that the current law does not take advantage of the opportunities given by pseudonymization and does not create a clear legal framework that can be converted to problem-relevant technical requirements. Even worse, it allows the Member States to ban anonymous reports. On the other hand, we show that so far, no pseudonymization tool developed within official ID management frameworks addresses all threats to reporting systems.

Keywords: whistleblower; protection; EU Directive; anonymity; pseudonymity

1 Introduction

Information about law violations within organizations often comes to light thanks to whistleblowers. However, these individuals face a significant risk of retaliation after submitting a report. Unfortunately, there are numerous examples of reporters being punished for reporting non-compliance, negligence, or corruption [1,2].

The "Directive (EU) 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons reporting infringements of Union law" [3] (hereinafter referred to as the Whistleblower Directive or simply the Directive) aims to address gaps in the existing legislation of the Member States and obliges them to pass laws protecting whistleblowers.

The primary goal of the Directive is to improve the public confidence that whistleblowing can be carried out confidentially and avoid any form of retaliation, and consequently increase the share of reported cases in the total number of law violations. However, while fully acknowledging the righteous arguments for and appreciating the general goals, the real impact of the Directive depends on legal and technical details, as well as the creativity of parties attempting to neutralize the positive effects of the Directive or misuse it. A closer look reveals that the Directive follows a very conservative approach and reflects the technological situation from at least half a century ago.

Its protection measures are effectively limited to banning all forms of retaliation against whistleblowers based on the silent assumption that at least one reporting channel is run honestly.

A different approach would be to focus on processes that are *secure by design and default*, taking advantage of remarkable advances in Privacy Enhancing Technologies (PETs), but being aware of the existing limitations. Remarkably, this approach could be effective in the case of many attack and misuse vectors. The main techniques of PETs are pseudonymization and anonymization. They do not automatically solve all possible problems, but the potential for improving whistleblower protection against retaliation is significant in many areas.

The key questions addressed in this paper are: To what extent do the Directive and the resulting national regulations permit the use of pseudonymization and anonymization techniques? Are there any legal obstacles to their implementation? Furthermore, assuming such use is allowed, are tools and systems readily available on the market that can be utilized immediately, without significant adjustments,

© The Author(s) 2025. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

while fully compliant with the regulations and adhering to high standards of security and privacy by design?

Previous work

To date, analyses of whistleblowing, including the recent Whistleblower Directive [3], have focused on the legal, administrative, and sociological perspectives. Prior to the Directive, research mainly focused on defining whistleblowers' roles within societal and legal contexts. Studies have highlighted the adverse effects of whistleblowing on individuals, including negative impacts on psychological well-being [4] and future employment prospects [5]. Furthermore, many papers [6–8] have examined skeptical attitudes toward whistleblowing, particularly in countries where historical contexts contribute to what is often termed an "ethical climate" hostile to whistleblowers. A common perspective across the cited works highlights that whistleblowing is a process with many subtle nuances and resists a "one-size-fits-all" approach. [9] expands on this, emphasizing that the unconditional protection of whistleblowers may cause more harm than good. Specifically, in business sectors vulnerable to secret disclosure, the level of protection should be in line with the state's secrecy laws to mitigate potential risks of reckless protection misuse.

The complexities of whistleblowing across different cultures and legal systems are studied in [10], where Poland, a former Eastern Bloc country, is used to illustrate how historical mistrust rooted in the role of "informers" under oppressive regimes has fostered an enduring cultural skepticism toward whistleblowing. Pacella's work [11] examines U.S. whistleblowing laws within the context of cybersecurity-associated breaches and threats. Pacella notes that narrow judicial interpretations of U.S. whistleblowing programs will likely leave whistleblowers unprotected.

Following the publication of the Directive [3] in October 2019, researchers began analyzing its legislative impact and potential pitfalls. Kaufmann et al. [12] assessed its impact on the German legal system in depth. Teichmann et al. in [13] examined the implications of the Directive for company-employee relations, arguing that the Directive's protections may be overly generous and could enable misuse by employees while simultaneously allowing an organization to create a false appearance of guaranteeing anonymity, a viewpoint we revisit in subsequent sections. Scheu [14] further criticizes conceptual weaknesses in the Directive, highlighting issues such as the expansion of the list of protected subjects by Member States, which prevents uniform implementation in the EU.

While the Directive and the discussion surrounding it focuses on protection against retaliation from the reported organization, we are not aware of any publication concerning the problem of retaliation from third parties, including possible retaliation for *not reporting*. In fact, an individual can be coerced (for example, for political reasons) or paid to submit a report for the benefit of a third party. Such and similar scenarios were discussed in the e-voting literature (see, e.g., [15]), but they are also relevant to whistleblowing systems.

Pseudonymization is strongly advocated by GDPR [16] as an effective data management tool to support the compliance of data protection obligations.[1] In 2025, the European Data Protection Board published official guidelines on pseudonymization [17], further clarifying its role and technical implications in data protection frameworks.

1 The GDPR explicitly recognizes that pseudonymized data are not anonymous and therefore must be appropriately protected. In particular, GDPR refers to (secret cryptographic) data that enables re-identification.

Despite its central role in privacy protection, the use of pseudonymization specifically for whistleblower protection has received limited attention in the existing literature. An exception is the work of Berendt and Schiffner [18], which discusses the potential role of anonymization and pseudonymization in processing whistleblowing reports and points to known limitations of anonymous communication techniques. The paper discusses the interplay of anonymity, ethics, and governance issues. This constitutes an important starting point, but there remains a gap in systematically exploring pseudonymization as a dedicated technical safeguard for protecting whistleblowers, particularly in the context of balancing accountability with confidentiality.

So far, no widely accepted formal model for personal data protection in the whistleblowing systems has been presented in the literature. Due to conflicting requirements, creating such a model is a nontrivial issue. Perhaps, it is even harder than modeling e-voting, as the anonymity of a voter is unconditional, while the identity of a whistleblower must be revealed in certain situations.

How to approach legislative impacts of the Directive

The EU legislators have chosen to cover the topic of whistleblower protection by a directive and not by a regulation. Therefore, the Member States are responsible for the legal framework of whistleblowers' protection. While the general framework defined in the Directive is mandatory, the decision on how to implement it is left to the Member States. In particular, the Directive does not preclude higher standards of protection. Moreover, the Directive does not force the use of any specific technical solution.

This approach has certain advantages and disadvantages. On the positive side, it enables adjustment of the rules and smooth embedding into the national legal and cultural context. Roles and responsibilities may be assigned to already existing institutions in a cost-effective way. On the downside, the readiness to implement the Directive may vary from state to state, with some implementing the rules with significant delays and quite reluctantly (see [19] for the adaptation timeline). In practice, the authorities of each Member State might be tempted to implement the Directive in a minimalistic way, just to reduce costs given budget limitations and other tasks perceived as more important (as emphatically illustrated by the Polish implementation, see Section 'Pseudonymization in German and Polish implementations of the directive'). Apart from limited willingness to combat violations of European law, there are significant disadvantages on the technical side. First, each Member State must organize and implement a technical ecosystem to support reporting channels while bearing the costs of the venture. As we shall see, there are no on-the-shelf solutions to be deployed, so there are many technical challenges ahead, with many alternatives to solve the problems. This may have further negative consequences, since design decisions might be irreversible for a long time. In particular, technical incompatibilities created by different design approaches can last very long. Note that even for the technically easy problem of online authentication of individuals, this approach resulted in wasted efforts and delayed deployment in Europe.

Given that situation, we aim to analyze the situation from the point of view of an advocate of the devil, taking into account not only good intentions but also what can go wrong. One might argue that the Member States could choose to adopt national laws that provide stronger and wider security guarantees than the Directive requires. However, this is likely to remain wishful thinking (a concern also noted in [20]).

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

Our contribution

Our contribution is twofold. First, we scrutinize the Directive and its two national implementations from the perspective of identity management rules (Section 'The Whistleblowing Directive'). The goal is to analyze the real-world applicability of identity-hiding techniques.

- In Section 'Pseudonymization: Pros and cons', we examine the Directive in the context of protecting the whistleblower's identity and demonstrate why identity-hiding is critical for the proper functioning of any whistleblowing system.
- We evaluate the advantages and disadvantages of identity-hiding methods. We conclude that, in general, pseudonymization is the most suitable approach for whistleblowing systems.
- We further address the question of whether pseudonymization is actually permissible under the Directive by comparing pragmatic requirements and what the Directive actually allows. We find that, despite dealing with arguably higher-stakes scenarios, the Directive does not rely on the privacy by design paradigm but instead relies on trust assumptions and legal obligations. As the Directive does not explicitly propose any identity management framework, we systematically infer what is allowed and what is not based on its recitals and articles (Section 'Pseudonymization issues in the directive'). Finally, we provide a comprehensive synthesis of the technical requirements related to whistleblowers' identity, as derived from the Directive's recitals and articles.
- From the same point of view, we analyze the German and Polish implementations of the Directive (Section 'Pseudonymization in German and Polish implementations of the directive'). We reveal that identity-hiding is not actively encouraged in those two state laws and is even effectively discouraged in Poland.

Second, we check whether the intended (and compliant with the Directive!) pseudonymization goals can be achieved with widely available tools without the necessity to create custom solutions. We demonstrate that *none* of the available solutions, whether cryptographically strong primitives (Section 'Cryptographic primitives and protocols associated with pseudonymization') or industry-recognized identity-hiding tools (Section 'Available pseudonymization tools and their suitability'), can be directly used to fulfill the requirements of a whistleblowing system.

Impact.

We believe that advanced solutions, analogous to those used in e-voting [15], can be created for whistleblowing systems. To achieve this, we must:

(1) assess the current legal situation (as done in this paper),
(2) define the functional and non-functional requirements for a secure whistleblowing system (current ongoing work),
(3) develop technical solutions that meet these requirements (a challenge for the privacy, cryptography, and security community).

In addition, the lessons learned could serve as valuable feedback for reshaping the technical and legal frameworks of whistleblowing systems.

The Whistleblowing Directive

Background

Whistleblowing is the act of reporting unethical, illegal, or harmful activities within an organization. Typically, whistleblowers are insiders who disclose misconduct related to fraud, corruption, safety violations, environmental damage, or breaches of public trust.

Whistleblowing has deep historical roots, the first documented whistleblower case in US history occurred in 1777 [21]. Ten marines reported misconduct by the Commander-in-Chief of the Continental Navy, accusing him of torturing British prisoners of war. These whistleblowers faced retaliation for their action; however, they set a critical precedent—the Continental Congress passed the first whistleblower protection law in 1778, affirming that it was the "duty of all persons in service of the United States" to report misconduct. This legislation marked a foundational moment in the history of American whistleblower protections.

In the 20th century, consumer advocate Ralph Nader became a pivotal figure in the modern whistleblower movement, starting from his 1965 book, "Unsafe at Any Speed" [22], which exposed hazardous practices in the automobile industry and led to significant consumer protection reforms. Nader not only acted as a whistleblower himself but also championed legislative efforts [23] to protect others who expose corporate or governmental misconduct.

Legal protections for whistleblowers differ significantly across countries and often depend on factors such as the jurisdiction where the original wrongdoing occurred. To date, several countries have enacted comprehensive whistleblower protection laws that establish formal reporting mechanisms and safeguard whistleblowers from retaliation [24]. One of the most prominent legal initiatives is the European Union's attempt to harmonize legislation across all Member States through the Whistleblowing Directive.

Short overview of the directive

The Directive's primary goal is to enforce creating reporting systems, define formal requirements regarding their operation, and protect whistleblowers from possible retaliation. The Directive applies solely to violations in specific areas of EU law [12] including public health, environmental protection, product safety, financial services and markets, and data privacy [Art. 2]. However, the Member States have the discretion to extend its scope.

Reporting system

The mission of a reporting system is to enable submission and processing of whistleblower reports. A reporting system must be accessible to all individuals who acquired information on breaches in a work-related context. Apart from the employees in the standard sense, this includes at minimum several additional entities (see [Art. 4(1)]). Notably, the Directive does not include certain groups, such as Ph.D. students (for example, in Poland, Ph.D. students are not categorized as employees, even if they have standard teaching duties). Similarly, patients of public health services are not protected by the Directive despite the potentially significant consequences of retaliation. The list describing "work-related context" is open and allows for extensions by the Member States.

A reporting system must provide the whistleblower with a receipt of receiving a report,[2] investigate the reported issue, provide follow-up information and feedback to the whistleblower, and if justified, transfer the case to another reporting system or to a law enforcement authority (see Fig. 1). A reporting system must be designed to ensure the confidentiality of whistleblowers and any third party mentioned in the report.

2 Except for some extraordinary cases concerning, in particular, whistleblower's security. See [Art. 11].

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

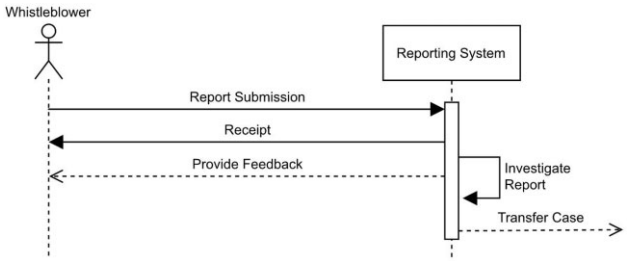

Figure 1. An example of whistleblowing process flow. Solid and dashed lines represent obligatory and conditional procedures, respectively.

Reporting paths

The Directive defines three paths for processing a report:

(1) **Internal reporting**: This path is managed individually by the entity where a breach occurs. Member States have to encourage whistleblowers to report breaches internally whenever the breach can be effectively addressed internally, and the whistleblower considers that there is no risk of retaliation [Art. 7].
(2) **External reporting**: This path is managed by competent authorities designated by a Member State. A whistleblower can use an external reporting path at any time, in particular after submitting a report internally [Art. 10].
(3) **Public disclosure**: Whistleblowers may also choose to disclose breaches publicly. However, to enjoy legal protection from retaliation, the whistleblower must meet specific requirements [Art. 15].

Most legal entities in the private and public sectors are required to establish internal reporting channels. However, certain small entities are exempted from this obligation or can share resources of internal reporting paths [Art. 8]. Member States are obliged to establish external reporting channels [Art. 11].

Example Workflow: An example workflow for processing a report can look as follows:

(1) An employee of an organization witnesses a breach.
(2) The employee estimates whether the breach can be adequately addressed internally and how severe the risk of retaliation is.
(3) The employee decides to report the breach using the external reporting path.
(4) The employee submits the report over a selected channel and receives a receipt.
(5) The report is assigned to an investigator, an impartial person or department.
(6) The investigator examines the reported case. If necessary, additional information may be requested from the whistleblower and other parties involved.
(7) Within 3 months, the employee receives feedback concerning the issues raised in the report.

Reporting channels and their confidentiality

The Directive defines six reporting channels for submitting a report. There are three channels via written communication (postal mail, physical complaint boxes, online platform) and three oral channels (telephone, voice messaging system, in-person meeting). Note that the whistleblower's identity can be easily inferred through most of these channels. For telephone and voice messaging systems, a whistleblower's voice can be recognized by the recipient of the report. Even if anonymization techniques are employed, vocal characteristics or speech patterns may still provide clues to the whistleblower's identity [25,26]. In the case of postal mail and physical complaint boxes, if reports are handwritten, the whistleblower's identity can be deduced through graphology. If reports are printed, one can frequently recognize the printer. Furthermore, whistleblowers who submit reports to complaint boxes can be identified through witnesses, CCTV footage, or other surveillance measures. Obviously, an in-person meeting cannot guarantee anonymity. An online platform, depending on its design, may protect a whistleblower's identity by concealing it from unauthorized parties or, on the contrary, can provide opportunities for illicit circulation of the whistleblower's personal data.

Beyond confidentiality issues, many channel-specific vulnerabilities can critically endanger reporting channels. In particular, ineffective authentication of a report's author can be exploited for impersonation or flooding attacks. For example, a written report can be submitted with a forged signature. Similarly, a voice-messaging system can be misused using an AI system to provide reports with a perfectly synthesized voice of an alleged whistleblower.

Among the available channels, arguably the only one that might be capable of ensuring a high level of confidentiality and resistance to impersonation attacks is the online platform. Therefore, for the remainder of this paper, we will focus on reporting systems based on *online platforms*.

Pseudonymization: pros and cons

In personal data protection, concealing explicit identities during data processing is a fundamental best practice. In particular, if an explicit identity is not required for a justified purpose, it should be removed from the dataset in accordance with the data minimization principle [16]. However, in the context of whistleblowing, the situation is more complex. Although replacing an individual's explicit identity with an identity-hiding alias enhances confidentiality and follows the data minimization principle, it also introduces new challenges and risks. Therefore, we briefly examine various identity-concealing approaches, considering their respective advantages and disadvantages within the whistleblowing framework.

A reporting system envisioned in the Directive comprises three primary parties: whistleblowers, report receivers, and organizations processing the reports. According to the Directive, a report receiver should be impartial. In the case of internal reporting, this receiver may be an individual or a department within the organization or, alternatively, an external third-party entity. The Directive assumes a separation of interests between the report receiver and the organization. This is clearly an overly optimistic assumption, as the receiver is often an employee of the organization and may be subject to various pressures, regardless of the Directive's provisions. This is especially likely if the reported organization violates the EU law. In this situation, the security of a whistleblower can a whistleblower's security may depend on the behavior of the report receiver. The receiver may face ethical dilemmas, be susceptible to bribery, experience pressure from the organization, or fear that the reported issue could lead to organizational shutdown and job losses. Furthermore, the report receiver might inadvertently or purposefully disclose the whistleblower's identity to other employees. Thus, the report receiver might be the critical weak point in the chain of trust. Hiding the whistleblower's explicit identity can mitigate this risk and reduce the trust requirements for the report receiver.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

Explicit identification

Mandatory association of the report with the reporter's identity, such as resulting from the requirement to sign the report, has notable implications for whistleblowing.

Disadvantages:

- **Identity exposure:** The whistleblower's identity could be exposed to third parties through negligence, malicious behavior, etc. In this case, many forms of retaliation are possible, due to the limited scope of the Directive.
- **Chilling effect:** Identity disclosure can have a chilling effect on a potential whistleblower and effectively stop them from filing a report.
- **Whistleblower's protection misuse:** A report may be filed not to expose violations of EU law but for the whistleblower's personal benefit, such as preventing dismissal or improving chances for a job promotion. This is quite easy since, on the one hand, the report may concern petty issues or reflect only *reasonable concerns or suspicions*; on the other hand, the organization concerned would have to prove that the report did not influence the organization's decisions. Proving it beyond any doubt might be nearly impossible.

Advantages:

- **Resistance to defamation and impersonation attack:** If the identity of the person submitting a report is verified during the submission process, it becomes nearly impossible to submit a false report under the name of another employee. This measure prevents individuals from misleading the organization and attempting to incite retaliatory action against the impersonated individual.
- **Resistance to sybil and flooding attacks:** An individual cannot file multiple reports purporting that they are being submitted from different people (a technique known as a *Sybil attack*). Similarly, a denial-of-service attack (flooding the reporting channel with a large number of reports to block the system) is mitigated, as such attempts can be traced back to the offending person(s).

As we can see, while an explicit identification approach has advantages in preventing whistleblowers' misbehavior. At the same time, it might discourage reporting. Consequently, it may result in a system in which the only reports submitted would be those that misuse protection mechanisms.

Anonymization

There are many ways how to understand *anonymizatiom*. In this paper, we follow the GDPR [16], according to which anonymization is, technically speaking, a one-way function, that is, one can convert an individual identifier to an anonymous identifier, but linking an anonymous identifier back to its preimage is infeasible. In particular, there is no cryptographic trapdoor to invert the conversion. Although anonymization addresses confidentiality concerns associated with explicit identification, it introduces new issues.

Disadvantages:

- **Authorization Challenges:** Deciding whether a report has been submitted by a person entitled to whistleblowing in the given organization (ref. [Art. 4]) might be a technical challenge. Theoretically, the issue can be solved with anonymous credentials technology [27] (see Section 'Anonymous credentials'). However, the system would be very complex. Note that applying for anonymous credentials and gathering the documents to be presented to the issuer of anonymous credentials is already a sensitive issue. In particular, a malicious issuer of anonymous credentials may create a list of potential whistleblowers. This list may be sold to employers who want to avoid hiring "troublemakers".
- **Susceptibility to Sybil and Flooding Attacks:** Anonymization provides a perfect camouflage for attackers, allowing them to use multiple unlinkable anonymous identities to corrupt the system. The mathematically provable one-way nature of anonymization hinders detecting both types of attacks.
- **Hindering Legal Proceedings:** The reported case may require follow-up legal proceedings. For example, criminal charges cannot be filed against a whistleblower who deliberately misled the authorities. Similarly, an accused person can be deprived of the right to defend themselves due to the impossibility of interrogating the witness.

Advantages:

- **Verifiable confidentiality by design:** Anonymization ensures that the reporting channel will not disclose the identity of a whistleblower. No trust assumptions are necessary; confidentiality follows from the properties of anonymization mechanisms.
- **Low organizational overhead:** The administrative and technical costs for maintaining confidentiality are substantially reduced; the effort is on the side of a whistleblower who creates an anonymous identifier. A suitable software application can easily handle this task.
- **Resilience to human errors:** Human-related disclosure risks are very hard to mitigate. After anonymization, no mistake can lead to an identity disclosure.

Although anonymization has substantial advantages, its disadvantages disqualify the use of anonymization for whistleblowing systems.

Pseudonymization

According to GDPR [16], pseudonymization is a mapping of identifiers to pseudonyms. In contrast to anonymization, pseudonymization can be reversed using specific secret data.[3] The most obvious realization of pseudonymization is with a cryptographic one-way function with a trapdoor. In any case, pseudonymization should enable controlled de-pseudonymization through a trusted entity or a group of cooperating entities. Pseudonymization has some minor disadvantages common to all contemporary IT systems based on cryptographic techniques.

Disadvantages:

- **Quantum and other Cryptanalytic Threats:** With the advancement of cryptanalysis, including quantum attacks, certain cryptographic methods used for pseudonym generation could become vulnerable [28]. Advances in this field are frequently unpredictable.
- **Human-Factor Risks:** Pseudonymization typically requires secure storage of credentials or passwords by the whistleblower. avoiding If these credentials are lost, the user may lose access; if passwords are compromised, unauthorized access may occur.

Pseudonymization enables avoiding the critical problems of explicit identities and anonymization.

3 In cryptographic literature, pseudonymization is also called *revocable anonymization*.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

Advantages:

- **No Identity Exposure:** Protecting a whistleblower's identity reduces to protection of data necessary for de-pseudonymization (e.g., cryptographic keys). Strict access control to these data should prevent any illegitimate identity disclosure. Many effective access control techniques exist; as a rule, they are significantly more reliable than any organizational trust-based method.
- **No Chilling Effect:** For a strong pseudonymization system with a strict de-pseudonymization policy supported by appropriate technical mechanisms, a whistleblower may believe to be sufficiently protected.
- **No Protection Misuse:** As long as the reported organization cannot independently de-pseudonymize a whistleblower, a person attempting to misuse the system cannot raise the argument of retaliation being exerted.
- **Resistance to Sybil and Flooding Attacks:** If a pseudonymization scheme ensures that only a single pseudonym can be created for an individual within the context of reporting for a given organization (so-called *domain pseudonyms* [29], see Section 'Cryptographic primitives and protocols associated with pseudonymization'), then Sybil and flooding attacks will fail. If this property does not hold for a pseudonymization scheme, then the Sybil and flooding attacks are technically possible, but de-pseudonymization would expose the offenders. Therefore, the threat is not so critical.
- **Controlled Identity Disclosure:** Identity disclosure can be controlled cryptographically by applying a threshold scheme where only an authorized group of entities can perform de-pseudonymization.

There are specific requirements for a pseudonymization scheme to be used in a well-designed reporting system. We point to three important features:

- **One Pseudonym per Domain:** An individual must be assigned only one pseudonym within an organization (i.e., a *domain-specific pseudonym)*, ensuring that Sybil attacks are infeasible.
- **Conditional De-Pseudonymization:** Under conditions predefined by an appropriate policy, linking a pseudonym with the real identity must be possible.
- **Inter-Domain Unlinkability:** An individual's pseudonyms across different organizations must be unlinkable.

Due to the recognized advantages of pseudonymization, we believe it to be the best-suited identity-hiding technique for whistleblowing systems. The key aspects are the availability of domain-specific pseudonyms and enabling the controlled identity disclosure, which enables calling a witness if the case goes to a court and serves as a safeguard against system misuse by users.

Pseudonymization issues in the directive

When analyzing legal acts, it is essential to distinguish between articles and recitals. Articles establish binding legal obligations, while recitals express intentions and objectives, serving to guide interpretation by Member States and courts. However, in the case of the Whistleblowing Directive, a document deeply concerned with identity protection,[4] mapping legislative language to concrete security and technological requirements usually cannot be done unambiguously. To improve this process, by understanding the underlying intentions of legislators (i.e. its *ratio legis*), we do not restrict the reading to articles but we also reach to its recitals. Accordingly, this section examines the recitals and articles of the Directive that affect the processing of personal identity data within reporting systems, derives guiding principles, and supports them with references to the relevant articles (as the normative basis) and recitals (as interpretative guidance). These recitals and articles can be categorized as shown in Table 1.

Table 1. We identified themes related to identity management and mapped the corresponding recitals and articles to each theme.

Theme	Recitals	Articles
Risks of abuse and overprotection	1, 32, 93	—
The role of identity hiding	30, 32, 34, 76, 82	6(2), 6(3), 16
Technical feasibility and design requirements	5, 60, 78, 83	12, 17
Internal vs external reporting	62, 69	7.2, 9, 11
Retaliation and protection mechanisms	37, 56, 57, 91, 93	21
Interoperability and fragmentation risks	76, 78, 104	—

Risks of abuse and overprotection

The Directive seeks a balance between protecting whistleblowers and preventing abuse, as overly strong anonymity protections may hinder accountability and encourage misuse [Rec.1]. The risk of malicious reporting is explicitly acknowledged as individuals who deliberately submit false or misleading reports are not entitled to protection [Rec.32], thus ruling out irreversible anonymization, which would prevent accountability. In retaliation cases, the burden of proof is shifted to the potentially retaliating party [Rec.93]. This is intended to protect whistleblowers but may risk making them practically untouchable, since it can be nearly impossible to prove that actions made by party accused of retaliation were unrelated to a report. To mitigate these risks, protection mechanisms must ensure accountability while also preventing overexposure and misuse.

The role of identity hiding

The Directive differentiates between contexts in which strict anonymity is crucial to safeguard physical integrity and those where only limited protection is intended [Rec.30]. Drawing this demarcation line highlights the intention to build reporting systems with a moderate protection level, which is inadequate, say, in the case of a large-scale embezzlement of public funds. This approach is reinforced by the principle that the authors of deliberately false reports should not be protected, which, again, rules out irreversible anonymization that would prevent prosecution [Rec.32] in such cases. Based on [Rec.34], Member States may even choose to disregard anonymous reports altogether, which may preserve the status quo but could deter reporting in high-corruption cases. At the same time, the Directive recognizes identity protection [Rec.76]. To fulfill this requirement, it is usually recommended that pseudonyms replace all personal identifiers in a dataset storing reports and that a separate secure component performs the identity-hiding process and identity disclosure. Though leaving implementation choices to Member States risks fragmented and incompatible systems. Identity disclosure may become mandatory during investigations to safeguard the rights of defense

4 Notably, the processing of personal identity data is not addressed directly; instead, it appears primarily as a secondary issue in discussions of other aspects of whistleblowing.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

[Rec.82], effectively making anonymization unsuitable while permitting pseudonymization.

The Directive does not impose an obligation to accept reports submitted anonymously [Art. 6(2)], and leaves it up to Member States to decide in this regard. This provision appears to be the result of a political compromise; however, it significantly lowers the actual minimum protection level. On the other hand, it may follow from the belief that pseudonymization techniques are not mature enough to risk creating legal obligations. Also, an unconditional obligation to process reports submitted anonymously would enable Sybil and flooding attacks. Conversely, whistleblowers who initially report anonymously may still qualify for protection if they can later prove authorship [Art. 6(3)].[5] Finally, the Directive establishes strict rules on identity protection, stipulating that the identity of the reporting person cannot be disclosed to anyone other than the authorized staff responsible for receiving or following up on reports, unless the reporting person has given explicit consent [Art. 16].

Technical feasibility and design requirements

The Directive admits technical requirements that extend beyond legal obligations, underlining the importance of robust reporting system design. It encourages common minimum standards, though it remains unclear whether these standards explicitly apply to technical platforms [Rec.5]. From a technical perspective, harmonized solutions (such as potential integration with the European Digital Identity Wallet [31]) are preferable, since fragmented approaches are prone to failure. Reports must be submitted easily and in full confidentiality [Rec.60, Rec.73]. While encryption of internet traffic is now standard practice, communication metadata often remains exposed. Achieving genuine confidentiality may therefore require anonymous communication channels, such as those enabled by TOR technology [32]. Procedures should be periodically reviewed and good practices exchanged [Rec.78], but such informal measures cannot substitute for standardized technical requirements. This is especially true in cryptography, where reliance on best practices alone frequently leads to incompatibility, as illustrated by challenges in eID frameworks. System design is further bound to data protection rules, as the Directive explicitly reaffirms the applicability of the GDPR [16] [Rec.83], referencing to Article 25 of GDPR, which mandates data protection by design and by default, including measures such as pseudonymization.

External reporting channels must also guarantee completeness, integrity, and confidentiality, preventing unauthorized access [Art. 12]. Thus, although the Directive does not prescribe detailed technical specifications, it implicitly demands robust security architectures. The Directive further reinforces the applicability of GDPR, emphasizing secure systems by design and default [Art. 17]. It is worth recalling that GDPR explicitly highlights pseudonymization as a recommended technique for protecting personal data.

Internal vs external reporting

The Directive differentiates between internal and external reporting, noting that breaches of confidentiality are more likely in internal reporting [Rec.62]. This risk justifies the use of external channels in many cases, though it could arguably be mitigated if internal systems implemented robust pseudonymization. The Directive also refers to existing channels established by EU authorities [Rec.69], which are intended to guarantee confidentiality of reporting persons' identities. However, this assurance appears overly optimistic as closer examination reveals that currently deployed systems are not secure by design (see Section 'Available pseudonymization tools and their suitability').

For internal reporting (encouraged when there is no risk of retaliation) [Art. 7(2)], the Directive requires appropriate procedures, including follow-up mechanisms [Art. 9]. A crucial element is the presence of a feedback channel to the whistleblower. Here, pseudonymization can play an important role in reducing risks perceived by the whistleblower. However, in the case of anonymous reports, this requires an anonymous two-way communication channel. The Directive also regulates external reporting procedures, allowing competent authorities to withhold receipts and feedback if disclosure might compromise identity protection [Art. 11]. This limitation can be addressed technically through cryptographic tokens or e-voting-inspired protocols, where whistleblowers verify receipt and follow-up messages encrypted with their public keys. Such systems ensure that investigations proceed regardless of whistleblower engagement with the feedback, which is crucial in severe cases of law violations.

Retaliation and protection mechanisms

The Directive introduces a broad framework to protect whistleblowers from retaliation, which directly impacts the technical design of reporting systems. Protection extends to the broadest possible range of persons, implying that technical solutions must be non-discriminatory [Rec.37]. For example, systems tied to EU citizenship through identity cards would be impermissible as they exclude non-citizens. Directive recognizes that authorities handling reports must be independent and free of conflicts of interest [Rec.56]. However, this is not verifiable from the whistleblower's perspective. Pseudonymization can mitigate this concern by preventing personnel from knowing the whistleblower's identity, thereby reducing the risk of abuse. The Directive also requires that whistleblowers receive follow-up information [Rec.57], which necessitates anonymous communication channels when pseudonymization is used. The option for whistleblowers to opt out of feedback highlights the lawmakers' recognition of persistent threats, although pseudonymization can reduce some of these risks. According to [Rec.91], Directive's rules prevail over obligations such as non-disclosure agreements; pseudonymization further strengthens this by preventing the identification of the whistleblower, leaving no target for harassment. In retaliation cases the burden of proof is shifted to the accused party [Rec.93], creating potential for misuse, as it is often impossible to prove that no link exists to the report. Strong pseudonymization provides a solution by making retribution technically infeasible. Furthermore, protection against retaliation is codified by requiring Member States to deploy technical protections proportional to the risks faced by whistleblowers [Art. 21]. Therefore, systems that offer inadequate protection could be seen as violating the Directive's requirements.

Interoperability and fragmentation risks

Importantly, the Directive creates room for significant fragmentation in the implementation of technical protection mechanisms. Member States are recognized as having the authority to adopt adequate procedures to protect the identity of all individuals mentioned in a report [Rec.76]. Although pseudonymization is a recommended approach, leaving technical implementation to individual Member States risks generating incompatible solutions. Without harmonization, interoperability issues are likely to emerge. The Directive encourages review and exchange of good practices [Rec.78] among the

5 This could be achieved, for instance, by means of pseudonymous signature schemes [30].

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

competent authorities. However, this informal process cannot substitute for standardized and coordinated technical solutions. Since the Directive only establishes minimum standards [Rec.104], Member States are allowed to adopt more advanced protective measures if desired. While this flexibility opens the door for stronger privacy-preserving technologies, it also exacerbates the risk of fragmentation and lack of technical coherence across the EU.

Conclusions

Based on the recitals and articles of the Directive, we come to the following conclusions regarding the identity management of the envisioned reporting system:

(1) The accessibility aspect of whistleblowing plays a central role in the Directive. However, certain reporting channels (e.g., in-person meetings) fail to provide confidentiality guarantees for the whistleblower's identity. This, in turn, weakens the overall security of whistleblowing systems.
(2) Many provisions mirror solutions from the GDPR but do not refer to principles like *privacy by design and default* or *accountability*. Thereby, the Directive is not so focused on privacy and security like GDPR, despite that the whistleblowing case requires much more attention. While the GDPR recommends pseudonymization multiple times, this is not the case in the Whistleblower Directive. Many rules, such as processing data only by authorized personnel, reflect only what has already been enforced by GDPR.
(3) The Directive leaves it up to the Member States to decide whether to reject anonymous reports. The only apparent justification for this provision is to enable defense against Sybil and flooding attacks. Unfortunately, due to high retaliation risks in the most drastic cases, this may eliminate reporting, where it is mostly needed. Note that a pseudonymous identity framework would solve all these problems, but it is effectively discouraged by the Directive.
(4) Pseudonymization, in the spirit of the Directive, is at best tolerated. Notably, Member States may oblige whistleblowers to disclose their identities during reporting.
(5) The Directive's provisions for whistleblower protection are mainly based on a legal obligation to refrain from retaliation against a whistleblower in work-related matters. This is insufficient. Obligatory pseudonymization would solve most of these issues, but the Directive fails to go in this direction.
(6) The Directive enables to protect a whistleblower by restricting the information flow. First, if a competent authority processing a report considers that issuing feedback may jeopardize the whistleblower's security, it can withhold such feedback ([Art. 11(b)]). Furthermore, according to [Art. 11(b)], the Directive allows external report recipients to withhold a receipt from the whistleblower if they "reasonably believe that acknowledging receipt of the report would jeopardize the reporting person's identity." This is controversial, as recipients could exploit this provision to avoid providing a receipt, leaving whistleblowers without any proof and vulnerable to retaliation. Again, a pseudonymous identity framework and anonymous communication channels (e.g., using TOR [32]) would solve the problems without creating opportunities for misbehavior mentioned above.
(7) On the other hand, whistleblowers may abuse the legal protections granted by the Directive. For example, submitting a report on a minor issue may be used to block the termination of employment. These strategies of misusing the whistleblower's protection would fail in the case of pseudonymous identities. As long as they effectively hide the whistleblowers' identity, targeted discrimination or retaliation would be impossible. Therefore, it may happen that the Directive will have a limited impact on fighting violations of European law, but instead, it will open doors for certain pathologies of the job market.
(8) The Directive focuses on low-risk cases, neglecting high-risk scenarios and systemic corruption e.g., involving state institutions. It cannot be safely assumed that external reporting solves the problem in the case of severe charges against governments of some Member States for violating European law. A higher level of protection is necessary in these extreme cases. Some ideas from the anonymous witness programs and identity protection schemes could be shared with whistleblowing systems.
(9) While the Directive anticipates both cross-border and domestic interoperability, including legal harmonization, it fails to provide pragmatic mechanisms to achieve this. For example, [Art. 14] defines the concept of "experience exchange." However, "experience" is gathered when the systems are developed, which might be too late for any substantial changes in the system architecture. For example, if identity management is designed differently (see Section 'Pseudonymization in German and Polish implementations of the directive'), systems will likely require significant effort to interoperate. In the past, even in relatively simple cases such as implementing qualified electronic signatures, the effort to harmonize independently developed systems was surprisingly high.
(10) Ambiguities arise regarding the supervision of reporting systems. Under [Art. 14], competent authorities (for external reporting) are required to review their reporting processes. While Member States are expected to assign this responsibility to a supervisory entity, the Directive does not clearly define the legal framework governing this entity. It remains unclear what the entity's authority limits are; for example, whether it can enforce actions by the competent authorities or is restricted to merely providing recommendations. This lack of clarity could result in bad practices going unnoticed for years, and even when exposed, bureaucratic delays could indefinitely postpone necessary changes. A remedy to this situation might be relying on technologies that address some threats by design and enable easy conclusions in the case of an audit.
(11) In practice, the Directive may be interpreted so that a whistleblower has to prove a work relationship with the report's subject. However, the Directive lacks proposals how it could be implemented. Proving such a relationship without explicitly revealing one's identity is far from being trivial.
(12) The Directive does not mandate a framework of approved schemes helping to maintain and demonstrate compliance with the Directive (as notably can be done under GDPR). This lack of standardization creates a situation with no consistent assurance that identity management systems operate as intended. Consequently, whistleblowers may lack confidence in the system's protections.
(13) Most critically, the Directive does not guard against implementers to slither their way through the Directive, exploiting loopholes and creating poorly designed or maliciously implemented systems that meet compliance requirements while failing to provide meaningful protections for whistleblowers. This

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

most likely would discourage potential whistleblowers to report law violations.

Pseudonymization in German and Polish implementations of the directive

To spotlight how differences in implementation of the Directive can effectively change its impact on national law, we discuss German [33] and Polish [34] acts.[6]

Comments on the German law

The Directive was implemented by the so-called HinSchG Act [33], effective from May 31, 2023. From the perspective of pseudonymization, the following aspects of HinSchG should be noted:

(1) §10 of HinSchG covers the rules of processing personal data. It states that processing them by competent authorities is allowed only when necessary to perform the tasks defined within HinSchG. Thus, one can claim that storing any plaintext identification data that is not needed for explicitly defined tasks, such as feedback purposes, is forbidden. Referring to §10, a whistleblower may challenge any request to disclose their full identity data as long as pseudonyms are sufficient for purposes of HinSchG. In particular, a whistleblower may attempt to use the pseudonymous identification function (Restricted Identification, see Section 'Restricted identification') that has been implemented on the German personal identity card. On the other hand, according to §16.1 and §27.1, the competent authorities are not obliged to support anonymous report submission. Effectively, in the case where the competent authority does not provide any anonymous communication channel, finding a way to submit a report anonymously is an issue left to be solved by a whistleblower.

(2) §11 states that a whistleblower has the right to review and correct the protocol, as well as to confirm it: "durch ihre Unterschrift oder in elektronischer Form zu bestätigen" (in Eng.: *confirm with a signature or in electronic form.*). It is important to note that *electronic form* refers to an electronic document signed with a qualified signature. In this context, the German lawmakers disregarded the concept of Pseudonymous Signature developed by the German federal authority BSI [30]. Note that a qualified electronic signature provides strong cryptographic proof of the document's origin for the third parties. However, this can pose a problem in the event of a report's leakage, where the qualified signature may create more threats than it mitigates. Specifically, revoking a corresponding certificate does not prevent the disclosure of a whistleblower's identity; it only impacts the legal standing of the signature (written form equivalence).

(3) In the German system of *competent authorities*, the competencies are distributed among various authorities. If a report is submitted to a wrong authority, it is forwarded to the appropriate competent authority *while maintaining the confidentiality of the whistleblower's identity* (in German: "unter Wahrung der Vertraulichkeit der Identität der hinweisgebenden Person"). If this is technically impossible, then the whistleblower must be informed by the authority.

This procedure technically allows a whistleblower to submit a report to an untrusted authority that does not support anonymization tools. In this case, the whistleblower's strategy might be to submit the secured report to a different, trustworthy authority, thereby ensuring that the untrusted authority receives the report in a secured and protected form.

(4) In Germany, it is obligatory to process anonymous reports (§16, §27).

Comments on the polish law

A different approach, less favorable for the technical protection of whistleblowers, is represented by Polish law [34]:

(1) Unlike in Germany, processing anonymous reports is not obligatory. [Art. 7] of [34] states that competent authorities have the right to accept them, but this is not obligatory (in Eng.: *[competent authorities] may accept reports submitted anonymously*). Moreover, there are no specific rules defining the conditions for decisions on anonymous reports leaving room for arbitrary decisions. In particular, anonymous reports might be rejected to reduce the workload of competent authorities.

(2) According to [Art. 8(4)], competent authorities process only the personal data of the whistleblower that are necessary for report submission and follow-up actions. Personal data that are no longer necessary or were collected accidentally must be removed immediately (where "immediately" effectively means within 14 days after determining the data are irrelevant to the case).
These rules could potentially create a defense opportunity for whistleblowers. If a competent authority does not provide an anonymous reporting channel, the whistleblower could use an anonymous account for follow-up and request deletion of all data that reveal their identity.
However, this strategy is useless. According to [Art. 45], a registry of an external reporting channel must contain a record enabling the identification of the whistleblower. Notably, if the report is submitted anonymously but via a non-anonymous channel, the whistleblower's identity data will be retained in the registry! Moreover, the whistleblower's data should be stored for one full calendar year *after* the case is transferred to other authorities for follow-up procedures. This effectively results in a mandatory retention of the whistleblower's identity data (up to two years after transmitting the data to other authorities).

(3) [Art. 27] introduces obligations concerning non-disclosure of the personal data of a whistleblower. Violation of these rules is punishable by imprisonment. This rule may be regarded as a substitute for privacy by design by technical means. The intention is that the risk of criminal charges and severe penalties should minimize the threat of data leakages. Unfortunately, the regulation may turn against the protected persons: in the case of a security breach or an intended data leakage, the competent authority may attempt to deny the source of the problem and claim that the whistleblower must have disclosed data themselves.

(4) Lack of unconditional pseudonymization by technical means creates a threat of intentional misuse of protection measures defined in [Art. 12]. A person facing the threat of, say, a disciplinary dismissal may submit a report as a whistleblower and enjoy the protection against "retaliation." After being

6 We chose these two implementations because we can read them in the language that they were published, reducing the risk of translation errors.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

dismissed, this person may present an acknowledgment of report submission (according to [Art. 36], issuing an acknowledgment is obligatory) and claim that the dismissal is a retaliation for the report. In Polish legal and social reality, this is a likely scenario.

(5) According to [Art. 36], an external report can be submitted orally or in writing. In the latter case, it can be submitted as a paper document or an electronic document. [Art. 36(2)] refers to a *document form*, which, by definition, must enable identification of the document author (Art. 77.2 of Civil Code). In this way, any anonymous report submitted over an electronic channel should be rejected due to formal reasons. This effectively bans all advanced solutions with cryptographic pseudonymization that offer a high protection level.

(6) On the other hand, [Art. 43] states, in particular, that the competent authority shall guarantee that the submission process and the following processing ensure the confidentiality of the whistleblower's identity. The scope of protection is broad. It concerns the data enabling whistleblowers to be identified directly or indirectly. Thus, the Polish regulation obliges the competent authority to collect sensitive identity data (even if they are not to be used) and, at the same time, puts the burden of data protection on this authority. pragmatic approach to reduce the threats would be immediate data pseudonymization in the sense of GDPR (thereby, reversible).

(7) [Art. 45] defines the contents of a registry keeping records on report submissions. One of the fields is in Eng.: "personal data of the whistleblower and the person to whom the report relates, necessary to identify them". So, the question is whether a pseudonym can be used instead of an explicit identifier as "personal data necessary to identify" the whistleblower. Strictly speaking, a pseudonym itself does not point to an "identifiable person." The explicit identity can be revealed only with additional data that, according to the state-of-the-art, should be stored elsewhere in a secure IT component. There was a long discussion on whether pseudonymized data falls to the category of personal data, however the recent Opinion of Advocate General Spielman from 6 February 2025 [35] states that pseudonymized data without "additional data" are not "personal data" anymore. So, to implement [Art. 45], it would be necessary to keep pseudonymized data together with data enabling de-pseudonymization. This would be nonsense from the technical point of view.

Conclusions

The approach taken by the German legislators is cautious: while anonymous submission of reports is allowed, there is no obligation for competent authorities to implement such channels. This approach aligns with our findings in Section 'Cryptographic primitives and protocols associated with pseudonymization' on limited maturity of available pseudonymization tools.

On the other hand, the Polish law contains controversial points. It lacks a mandatory requirement to process anonymous reports, leaving decisions arbitrary. Although provisions protect personal data and require their deletion when no longer necessary, loopholes weaken these safeguards. Severe penalties for breaches of whistleblower confidentiality aim to ensure privacy but effectively shift responsibility onto whistleblowers. Although confidentiality is required, competent authorities are obliged to collect identity data. From the legal point of view, it is unclear whether full pseudonymization can be applied.

Cryptographic primitives and protocols associated with pseudonymization

In this Section, we examine cryptographic pseudonymization primitives and protocols that potentially can be used for whistleblowing systems.

Blind signatures

A blind signature [36] is a digital signature where the signed message is hidden (blinded) from the signer. The user first blinds the message M to create M' and sends it to the signer. The signer creates a signature s' over M' and returns it to the user, who then converts s' into a valid signature s for the original message M. The core property is unlinkability: given a set of signatures created in this way, together with the corresponding messages, the signer cannot link the final signature s to the intermediate signature s' that was returned to the user.

In the context of reporting systems, blind signatures can be used as follows: An external third party T signs data on behalf of whistleblowers. An individual I sends a blinded version M' of their report M to T. If required, I proves their eligibility to report on O.[7] T signs M' with a key dedicated to O and returns the blinded signature s'. I then converts s' into a valid signature s for M and submits both M and s to the reporting channel.

This method requires minimal trust in T, as it does not access the reports. As there is a single authority T responsible for signing and verification process, no complex PKI infrastructure is required. It also guards against Sybil and flooding attacks.

However, blind signatures are too strong cryptographically. If multiple reports have been submitted, T and the reporting channel cannot link the authors to their reports. Disclosing the report's author might be necessary for a criminal prosecution. What is actually needed is a blind signature with a trapdoor, where authorized parties can link a signature s to one of the original signing requests s' in the cases defined by law.

Restricted identification

The Restricted Identification (RI) [37] is a mechanism for creating pseudonymous passwords by a trusted tamper-proof cryptographic device separately for each *domain* providing a service (in fact, a server or a group of servers). The protocol has been designed by the German authority BSI for personal identity cards.

A user's personal ID card D stores a single secret key sk used by the RI protocol. To create a password, RI utilizes the domain's public key dpk validated through a special certificate. The core cryptographic procedure derives a password from dpk and sk. The point is that the passwords derived for different public keys dpk', dpk'' are unlikable. That is, it is infeasible to detect whether they were created by the same device. Moreover, for a given dpk, device D always generates the same password during its lifetime.

A whistleblower can use RI to create pseudonyms when submitting reports. This approach offers immediate advantages, including resilience against Sybil and flooding attacks, as it ensures that a whistleblower appears under only one pseudonym within an organization.

However, RI also presents some limitations for whistleblowing purposes. First, RI cannot be used to sign a report. Second, it lacks a way to verify whether the individual is a legitimate whistleblower.

7 Later in this section, we investigate how one can demonstrate entitlement without explicitly revealing their identity.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

Yet, the main concern is de-pseudonymization of a whistleblower. The only option is for the ID card issuer and the organization to use their private keys to recompute the pseudonyms of all potential individuals one by one and compare them with the pseudonym used by the whistleblower. This exhaustive search is inevitable due to the hash function employed in the final step of pseudonym derivation.

Domain pseudonyms and signatures

Domain signatures extend the Restricted Identification concept by enabling to sign using a pseudonym. A user holds a private key sk (e.g., implemented on the personal identity document) and can: (1) derive a *domain-specific pseudonym* for any domain, (2) create *domain-specific signatures* corresponding to this pseudonym. The following properties are crucial for potential applications:

(1) A user can create exactly one pseudonym for a given domain (resilience to Sybil attacks).
(2) Given a set of pseudonyms and signatures from different domains, it is computationally infeasible to determine which pseudonyms correspond to the same user (unlinkability).

A nationwide domain signature infrastructure, such as one based on electronic identity documents, could offer strong protection for whistleblowers. For this purpose, each organization should be assigned a public key available in a public registry. Assume that an individual I aims to submit a report M about an organization O. The following steps could be executed:

(1) I fetches dpk of O and derives their pseudonym $P_{I,O}$ for that domain.
(2) I signs the report M with $P_{I,O}$, creating $s_{P_{I,O}}$.
(3) Both M and $s_{P_{I,O}}$ are anonymously submitted to the reporting channel for O.

Domain signatures protect report integrity and prove that the report originates from the owner of the pseudonym $P_{I,O}$. Domain signatures have certain advantages. First, a single pseudonym per domain prevents Sybil and flooding attacks. Second, cross-domain unlinkability protects whistleblowers from intersection attacks, where a whistleblower is identified by learning to which organization they belongs: if it is known that two reports, one on O_1 and one on O_2, originate from the same individual, then one can narrow down the set of potential whistleblowers to individuals in relation to both O_1 and O_2. Third, some domain signature schemes allow a domain-specific pseudonym to be linked to the real identity of its owner by a relevant group of authorities.

However, domain signatures and pseudonyms are not free from problems. First, there is no mechanism to prevent an unrelated individual from reporting on an organization. Second, a nationwide framework of domain signatures (e.g., within the European Digital Identity Wallet [31]) would be needed. Using it solely to protect whistleblowers would most likely prove to be infrastructural overkill, as arguably most whistleblowers will not experience multiple reasons to report during their lifetime. On the other hand, there are compelling reasons to develop a domain signature ecosystem. Apart from analogous applications (e.g., anonymous witness protection in court proceedings), many hard problems of personal data protection would be solved by design.

Domain signatures can be realized relatively easily, as shown by by the German federal authority BSI [30]. Unfortunately, these schemes require tamper-proof devices on the user's side. Moreover, there is no anonymity with respect to the issuer, and numerous sophisticated data leakage techniques could lead to deanonymization [38]. Other constructions, such as those based on pairing groups, are currently much more difficult to deploy due to unavailability of commercial products.

Proving entitlement via zero-knowledge proofs

According to [Art. 4] of the Directive, any work-based relationship qualifies for whistleblowing. The challenge is to verify the right to whistleblow while keeping the whistleblower's identity confidential. System-level pseudonymization will be useless if other data provided by the whistleblower indicates their identity.

Since it is impossible to predict all evidence forms (e.g., contracts, invoices, emails, etc.) witnessing a work-based relationship, integrating them into a formal verification system is a hard challenge. Creating a separate system for this purpose would be difficult due to the diversity of cases, the costs, and organizational overhead.

Manual verification would require a high level of trust, so a cryptographic alternative could be a rescue. One possible approach is to use mechanisms that prove possession of a digital document signed by an issuer while selectively revealing its contents.[8]

Theoretically, Zero-Knowledge Proofs (ZKP) could fulfill this purpose. A ZKP allows one party (the prover) to convince another party (the verifier) that a statement is true without revealing anything beyond the truth of the statement itself. For example, assume that a whistleblower W holds a digital document D issued and signed by an organization O. Then the rights for whistleblowing on O should be granted to W if they could prove the following statements about D:

(1) D is a work-related contract,
(2) D contains a valid signature of organization O,
(3) D contains a valid signature of whistleblower W,
(4) the secret key used in W's signature is in possession of the party presenting the proof.

An example of ZKP of possession of a digital signature, applicable to any standard signature scheme, is given in [39]. To verify the proof, the verifier must have access to the hash of the message. Thus, the remaining issue is to prove that the hash used in the ZKP corresponds to the document's hash. This task, however, is more complicated because we aim to conceal certain parts of the document (e.g., appearances of individuals' names). Thus, the document cannot be presented to the verifier as is usually done for signature verification. We identify two methods to achieve this goal.

The first method treats the entire document as a single bit string, applying a hash function directly. In this case, we may use general zk-SNARK constructions (see [40] for an accessible introduction). In particular, zk-SNARKs enable to prove that a specific message hashes to a given value without disclosing the entire message. Moreover, zk-SNARKs can support designated verifier proofs [41], allowing only the designated verifier to validate the provided zk-SNARK. Nonetheless, it is essential to note that zk-SNARKs impose significant computational overhead on the prover's side. This could pose challenges when designing a whistleblowing system intended for use by individuals with very simple tools. As of today, there are no such zk-SNARKs in consumer products.

8 We limit our consideration to documents with a level of "seriousness" sufficient to warrant an organization's signature. Such documents are referred to here as work-related contracts. It is important to note that documents lacking robust authenticity mechanisms (i.e., digital signature), cannot be regarded as valid entitlements, as they are susceptible to forgery.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

Alternatively, a document can be divided into smaller chunks based on its structure, each chunk hashed separately, with a final proof generated via a hash aggregation technique like a Merkle Tree [42].[9] The rest of the protocol could be achieved using selective disclosure of chunks. This approach is not universal, as it requires the document to be stored in a structure-oriented file format, such as XML, which is uncommon in practice. Ideally, the tool would integrate with PDF readers, support standard signed PDFs, and remain user-friendly without needing additional technical knowledge. Although it is possible to build such a software tool, we are not aware of any product of this kind on the market.

Proving entitlement via attribute certificates

Attribute certificates (AC) [43] are digital certificates that associate specific attributes or privileges with a user or entity, rather than establishing their identity like traditional certificates. ACs demonstrate eligibility for actions without revealing the user's identity. In the context of reporting systems, the right to submit a report on an organization is analogous to having such an attribute. However, this solution faces the transferability problem: an individual who has lawfully acquired an attribute certificate can transfer it to a third party.

One proposed solution is to embed valuable secrets (e.g., a private key that gives access to the bank account) [44] within the certificate, ensuring that any transfer leaks the secret, thereby deterring the user from sharing. Nonetheless, this approach may fail if the colluding parties are willing to accept the consequences or if no valuable secret can be leveraged within the system, as is likely in the context of an EU-governed process, such as a reporting system. As a result, the issue of transferring attribute certificates continues to pose a significant challenge.

Distributed ledger technology and blockchain

Distributed ledger and blockchain technologies offer key benefits for anonymous communication between a whistleblower and a receiver of the report. They can serve as an anonymous bulletin board for submitting reports, acknowledgments, and feedback, with the blockchain's append-only feature protecting against data manipulations and deletions.

Although a distributed ledger may be publicly accessible, data can be encrypted so that only the whistleblower and the reporting channel can access the plaintext version. Symmetric encryption using static Diffie-Hellman key exchange with the reporting channel's static key and the whistleblower's ad hoc public key, playing the role of a pseudonym, is an easy and pragmatic option.

This approach offers advantages such as asynchronous and distributed access to the communication channel, making it difficult to identify the whistleblower based on the traffic data. However, it poses challenges regarding hosting and visibility of changes. If we assume that the blockchain is publicly available and hosted by the network (i.e., a permissionless blockchain [45]), then an observer can infer information such as the frequency of report submissions.

"Cryptographers' feedback on eIDAS"

Recently, a group of cryptographers specializing in Anonymous Credentials (AC; refer to Section 'Anonymous credentials') published an opinion [46] on the eIDAS 2.0 regulation [31]. They highlighted that the cryptographic technique currently proposed in Architecture and Reference Framework [47] for revealing attributes in the EU Digital Identity Wallet (EDIW), namely, the use of a salted hash, fails to meet the privacy requirements outlined in eIDAS 2.0, and particularly [31, Art. 5a]. Their analysis is relevant to our context, as an adequately implemented EDIW could serve as a component used by a whistleblower to submit a report in a privacy-preserving way. The authors of [46] advocate for the adoption of AC and suggest redesigning the EDIW with AC in mind. However, they appear overly optimistic about the feasibility of integrating full-fledged AC into EDIW. There are practical challenges that have hindered a widespread real-world deployment of AC systems so far.

Key concerns highlighted in the feedback include the lack of hardware support and challenges related to non-transferability. The proposed AC system relies on digital signature schemes that require pairing-friendly elliptic curve groups such as BBS/BBS+, which are not yet widely supported by trusted hardware available on the market. Additionally, the non-transferability of credentials is enforced through a combination of an interactive protocol and the requirement for the cryptographic key to be stored in tamper-proof hardware. This is a critical challenge for the application of such AC systems for whistleblowing purposes, as providing secure devices to the citizens exclusively for this single purpose is unrealistic. Apart from that, all disadvantages of AC systems from the point of view of whistleblowing apply, as pointed out in Section 'Anonymous credentials'.

Available pseudonymization tools and their suitability

Given the Directive's goals and existing threats to identity management, our next step is to examine which commonly used or ready-to-deploy pseudonymization tools are suited for whistleblowing systems.

Privacy pass

The primary goal of Privacy Pass [48] is to provide a reliable and user-friendly alternative to CAPTCHA. Instead of solving CAPTCHA, the user presents a token issued by a trusted third party and given only to known and authenticated users. The main feature of this solution is strong privacy protection—The issuer cannot link used tokens back to the clients who obtained them. Thus, by presenting a token from an issuer A, a user only demonstrates eligibility to receive a token from A (for example, as a holder of an email account provided by A), without revealing any further identity information. The mechanism of Privacy Pass relies on blind signatures discussed above. Privacy Pass tokens are single-use only.

Unfortunately, Privacy Pass is not ideal for generating pseudonyms for whistleblowers. First, a whistleblower can obtain multiple tokens, enabling flooding and Sybil attacks, which Privacy Pass cannot detect. Second, since tokens are for one-time use, follow-up communication is impossible. That said, Privacy Pass could be useful as an auxiliary tool in limited cases, such as verifying that a reporting person meets the criteria outlined in [Art. 4].

FIDO2

The Fast Identity Online (FIDO2) [49] authentication framework, introduced by the FIDO Alliance, prioritizes privacy protection. Its core component, the *authenticator*, is a secure unit (hardware or software)

9 To mitigate dictionary attacks on fields with small domains (e.g., birth dates), salted inputs should be used.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

that interacts with a client (usually a browser) and a server (with the client acting as a proxy). FIDO2 uses a simple challenge-response protocol, where the authenticator signs certain values including the hash of the challenge sent by the server.

For each server, the authenticator undergoes a separate registration process during which a dedicated key pair for digital signatures is created. These keys are used subsequently in the challenge-response protocol with this server. The authenticator signs the public key with an *attestation key* and sends both to the server. The role of the signature created with the attestation key is to witness that the public key obtained by the server originates from a legitimate authenticator.

FIDO2's main advantages are flexibility, vendor independence, widespread deployment, and universal applicability. In particular, possessing a FIDO2 authenticator does not imply that its holder is a whistleblower.

For reporting purposes, a dedicated server can be established to handle whistleblower registration and report submissions and forward them to the appropriate destination. A small twist in the FIDO2 protocol would enable the authenticator to sign a report as a response to the server's challenge. Note that in this case, the signature itself would not disclose the user's explicit identity, as the public key is randomly generated for each organization. In this scenario, pseudonymization becomes straightforward: the registration server acts as a trusted third party but knows the identity of the authenticator and its owner. The advantages of this approach include resilience to flooding and Sybil attacks, independence of public keys generated by a single authenticator, and, therefore, resilience to cryptanalysis aiming to link the pseudonyms.

However, the critical disadvantage of FIDO2 is its dependence on a trusted party. Additionally, according to FIDO2 specification [49], attestation keys may be shared among multiple devices, making de-pseudonymization unfeasible in such cases. If the authenticator is a black-box device, there is also a potential risk that the device provider could embed kleptographic[10] code to enable device identification during protocol execution.

Privacy coins

Another prospective tool for protecting whistleblowers' identities is the technology behind certain privacy-focused cryptocurrencies such as Monero [51] and Zcash [52]. With some moderate adjustments, their anonymity mechanisms can be reused for reporting systems. We take Monero as an example.

In Monero, a user is identified by long-term public keys (K^v, K^s), where K^v is a public *view key* and K^s is a *spend key*, with corresponding private keys k^v and k^s, respectively. Importantly, transactions are not addressed to the receiver's public key pair but to a one-time public key K^o, derived from the receiver's public keys and the randomness chosen by the issuer. Only the receiver can calculate the corresponding private key k^o, enabling them to claim ownership of the crypto assets transferred. In other words, the Monero transaction structure allows the issuer to choose a one-time public key for the receiver.

This mechanism can be applied to a reporting system. An individual I contacts a trusted party T, providing public keys and proof of eligibility to report on organization O. If I is eligible, T derives a one-time public key K^o for I, creates a certificate C with K^o and relevant details, and publishes it on its blockchain.[11] I then calculates the corresponding private key k^o.

With C and knowledge of k^o, I can demonstrate their eligibility to report on O. Specifically, during report preparation, I signs the report using k^o. During verification, O verifies the signature using data published on T's blockchain. Multiple reports from the same whistleblower can be linked through Monero's double spending prevention mechanism. Sharing the private view key k^v with T allows verification of transactions without enabling T to create the reports on behalf of the users. The presented mechanism supports de-pseudonymization while at the same time does not reduce protection against malicious organizations.

On the downside, this technique shares some disadvantages with blockchain-based solutions (Section 'Distributed ledger technology and blockchain'), making it difficult to integrate seamlessly into a whistleblowing system and requires a trusted party T to issue credentials.

EU sanctions whistleblowing tool

Recital 17 of [53] states that "[...] the importance of insider reporting in detecting competition law infringements has already been recognized in the leniency policy pursued by the Commission under Article 4a of Commission Regulation (EC) No 773/2004 as well as with the recent introduction of an anonymous whistleblower tool by the Commission." The tool mentioned in Recital 17 appears to be the whistleblowing system[12] managed by the European Commission [54], primarily devoted to reporting EU sanctions violations.

The EU Sanctions website [54] offers two reporting methods: direct email and anonymous whistleblowing tool with an option to provide personal information. The former non-anonymous method is preferred: "Providing your personal details will give your statement more credibility and increase the likelihood of actionable outcomes."

The "About anonymity" section[13] provides information on how the tool handles anonymity. When filing a report, a user can choose to remain anonymous or provide personal details used solely for communication purposes (at least, according to the documentation). Importantly, users are advised not to include personal details within the body of the report. If identifying information is included inadvertently, the Commission will remove it before sharing the report with third parties for investigation. Finally, if the whistleblower's anonymity cannot be guaranteed despite these measures and there is no explicit consent of the report's author to disclose their identity, the report will not be shared with third parties running the investigation.

A secure and anonymous communication channel is claimed to be provided via a *secure inbox*, accessible with a tool-generated *case number* and user-selected *passcode*. The tool also advises on maintaining anonymity, such as manually typing the URL to prevent spoofing and avoiding any contact via the reported organization's network.

We conclude that in the provided solution "anonymity" is understood differently than in GDPR [16]. It is not even "pseudonymity," but rather refers to a trusted intermediary obliged to remove identifying data before forwarding them for investigation. This is an

10 Kleptography studies how to design cryptographic systems that secretly leak information to their creators, see [50].

11 From Monero's perspective, C is a *transaction* with a ring size of 1. Refer to [51] for an introduction to CLSAG signatures.

12 https://eusanctions.integrityline.com/frontpage

13 https://eusanctions.integrityline.com/app-page;appPageName=Anonymity

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

interesting deviation from the main line of the Directive, that is whistelblower's protection is not based on outlawing any kind of retaliation but rather by hiding the identity of a whistleblower to prevent retaliation.

Furthermore, in the case of the tool discussed, identity hiding is not controlled by whistleblowers (as could be the case for some tools discussed previously). Instead, this task has been entrusted to an intermediary server (a server running Integrity Line software provided by a private company EQS Group.[14]).

Paid black box private company tools

Many paid whistleblowing compliance solutions are offered by private companies[15]. However, their claims of confidentiality are not supported by any technical evidence, as these applications function as web interfaces to opaque, black-box infrastructures. So far, we have not encountered any technical documentation or whitepapers that would be publicly available. Consequently, these solutions rely solely on the manufacturer's assertions and trust rather than verifiable security measures.

Anonymous credentials

A credential system consists of users and organizations. A user obtains credentials from an issuer and presents them to a verifier. If the verification is successful, the verifier can grant the user certain privileges. An Anonymous Credentials (AC) system is a credential system providing several privacy-enhancing features. In an AC system, a user can be associated with multiple pseudonyms, each for a different role (e.g., for interaction with a given organization or within a separate process). In particular, an organization may know a user only via a pseudonym and a set of attributes revealed by the user and confirmed anonymously by a third party (technically, the pseudonym is also an attribute). An issuer can issue a credential concerning a set of attributes to a user, and the user can later prove possession of a credential while disclosing only a selected set of attributes and their values. No information about other attributes can be deduced by the verifier.

Credentials can be one-show or multiple-show, depending on the desired functionality. Pseudonyms in AC systems are unlinkable, which means that different pseudonyms of the same user cannot be linked as long as they do not appear in the same proof. Similarly, multiple demonstrations of multi-use credentials or separate credential demonstrations by the same user are also unlikable, if not linked by the revealed attributes. These features form the core functionalities of an AC system. By combining user privacy with strong authentication, AC systems are particularly valuable in contexts requiring strong privacy guarantees. Additional requirements may sometimes be introduced, such as *conditional anonymity revocation* and strong *non-transferability* that are especially important in the context of whistleblowing systems.

At first glance, AC systems may seem to be perfectly suited for whistleblowing. They offer pseudonyms, active authentication, and a secure presentation of attributes that can be used to evaluate the trustworthiness of a report. For example, if a report concerns the public healthcare system, revealing that a whistleblower is a medical professional may increase report's credibility. AC systems may also be used to prove entitlement to whistleblowing while preserving privacy.

On the other hand, AC systems are powerful but complex frameworks. From a pragmatic perspective, deployment requires a broad range of use cases since whistleblowing is rather a niche application.

In the following, we review two AC systems from major vendors: Microsoft's U-Prove [55] and IBM's Idemix [56]. These systems serve as a reference point to review the strengths and limitations of currently available AC products in the context of whistleblowing.

U-Prove

Since 2010, Microsoft has been developing an AC system called U-Prove [55,57]. A distinctive feature of U-Prove is accommodating application-specific requirements through the content and structure of attributes (values encoded in the credential by the issuer and the user). Its core functionality revolves around the *U-Prove token*, issued to a user (referred to as the prover) by an issuer through an *issuance* protocol. The user can then selectively disclose a subset of attributes from the token (credential) to a verifier using the *presentation* protocol. During issuance, the prover generates a token-specific asymmetric key pair. The issuer then signs the token attributes and the prover's public key using the issuer's private key, while the prover's public key remains hidden from the issuer. The issuer's public key is widely known.

When running the presentation protocol, the prover presents the chosen token attributes, the token-specific public key, and the issuer's signature. The verifier sends a challenge, which the prover responds to by applying the token-specific private key. The one-time structure of the challenge ensures that a token cannot be replayed.

This simple framework underpins U-Prove's basic functionality. Advanced privacy-oriented features, such as untraceability, unlinkability, and selective attribute disclosure, are achieved using three types of attribute encoding techniques:

- *Token information field*: Encoded by the issuer, this field is always disclosed when the prover uses the token. Typical uses include encoding metadata like expiration dates or usage restrictions.
- *Attribute fields*: Encoded by the issuer, attribute fields are selectively disclosed by the prover when they use the token. Because multiple presentation proofs can be generated using the same U-Prove token, a prover can conceal the contents of some attribute fields from certain verifiers while revealing them to others.
- *Prover information field*: Encoded by the prover, this field is always invisible to the issuer but always disclosed during token use.

Combining these encoding techniques, alongside additional cryptographic schemes, enables various functionalities, such as on-demand or long-lived tokens and identity domains.

Despite its promising design, some authors indicate certain limitations of U-Prove. First, it lacks rigorous security proofs and detailed design justifications, only semi-formal arguments are given in [57]. Second, the relatively high computational requirements make it inapplicable on standard smart cards and weak embedded devices. This is critical for whistleblowing systems, as citizens cannot be expected to obtain high-performance devices. Third, the official documentation [55,57] is vague, while the system's complexity increases the risk of insecure implementations and overlooked privacy threats by third-party vendors.

Another critical issue highlighted in [58] relates to the attribute revocation mechanism. Any universal identity token to be used on a massive scale must be easily revocable. The latest cryptographic specification [57, rev. 5, March 2023] only mentions that token revocation mechanisms can be integrated but provides no

14 https://www.eqs.com/en-us/

15 See e.g., https://www.speakup.com/blog/top-whistleblowing-software-tools.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

concrete guidance or references. Meanwhile, the technology overview [55, rev. 3, March 2023] advertises three privacy-preserving revocation methods, inconsistently referring back to the cryptographic specification. Another key issue for whistleblowing systems is that U-Prove prevents de-anonymization, which is essential in our contexts (see Section 'Pseudonymization: Pros and cons'). U-Prove on Microsoft's GitHub repository includes three public repositories: a C# SDK[16], a client-side JavaScript companion SDK[17], and TypeScript reference implementation[18]. Despite being open-source, contributions are minimal, with only two contributors and commits dating back 1 to 3 years. Documentation is sparse; the README file for the SDKs explicitly lacks usage instructions. The implementation supports basic U-Prove functionality and some extensions; however, attribute revocation relies on a previously mentioned insecure technique, as pointed out in 2015 [58].

We conclude that at the moment, many issues are preventing U-Prove from being recognized as a product that fits the needs of a whistleblowing system.

Idemix

In 2001, J. Camenisch and A. Lysyanskaya proposed an AC system with key features like non-transferability, unlinkability, and optional revocation [27]. Later, Camenisch et al. refined the protocol into high-level primitives, practical interfaces, and a prototype [56]. Later, the project was developed with IBM [59] and integrated into its Bluemix cloud platform and later Hyperledger Fabric (HLF) for anonymous identity verification in a permissioned blockchain network [60]. This integration allows blockchain participants to prove their identities anonymously while ensuring that their transactions remain unlinkable. Idemix has also been used in the EU projects PrimeLife [61] and ABC4Trust [62].

In Idemix, a user U interacts with an issuing organization O_I to acquire a credential C that confirms specific attributes $\texttt{attr}$. The credential C is issued for a pseudonym N, computed jointly by U and O_I. Later, when presenting attributes (contained in C) to a verifying organization O_V, the user can select a subset $\texttt{attr}'$ of $\texttt{attr}$ to disclose. The system involves three primary interactive protocols between the user and the respective organization: pseudonym registration, credential issuance, and credential verification. The pseudonym N and the credential C are linked to the user's secret key sk_U. During attribute verification, the user performs a zero-knowledge proof of two statements: possession of a credential containing the attributes $\texttt{attr}'$ issued by O_I, and knowledge of the key sk_U corresponding to this credential. The zero-knowledge property of this process ensures that multiple showings of the credential to the same organization do not reveal any additional information.

Two mechanisms are proposed to ensure non-transferability. First, transferring a credential to another party would allow the recipient to retrieve a secret external to the system, such as a bank password. Second, it would grant access to all credentials issued to the same pseudonym.

Idemix also supports conditional de-anonymization of users. This task is entrusted to two trusted parties: a de-anonymization organization O_D, which acts as a judge, and CA, an entity possessing a mapping from pseudonyms to identities. Before a credential presentation, a verifying organization O_V and a user U agree on de-anonymization conditions, which are then sent to O_D. Later, if U's behavior meets these conditions, O_V can present a proof of U's misconduct. In response, O_D provides the necessary information enabling CA to retrieve the identity of U. This identity can be subsequently shared with O_V. Note that all these steps are performed on the cryptographic layer. Roughly speaking, during the credential presentation, the user encrypts their pseudonym using O_D's public key. This encryption is verifiable, meaning that O_V can confirm that the received ciphertext indeed contains the user's pseudonym but cannot decrypt it.

Idemix shares some features with U-Prove but offers functionalities that make it more suitable for whistleblowing. Unlike U-Prove, Idemix inherently supports unlinkability. Idemix also provides more advanced privacy features, enabling finer control over selective disclosure, useful for proving "work-related relation" anonymously. The most significant distinction is Idemix's native support for conditional de-anonymization, a feature U-Prove lacks.

Despite addressing many challenges, Idemix faces significant infrastructure and technical complexities that can hinder its market success and social acceptance. These include the need for a Public Key Infrastructure, handling lost or sold devices, reliance on less common cryptographic primitives like CL signatures, and posing standardization and hardware support challenges. Similarly to U-Prove, Idemix's inherent heavy computation makes it unsuitable for weak devices.

However, the main concern over applicability of Idemix to whistleblowing is the current state of the project. The most relevant and up-to-date project named Idemix that we could locate is the official open-source repository maintained by IBM on GitHub [63], which remains a work in progress. This project serves as a dedicated Hyperledger Fabric (HLF) enables an anonymous blockchain membership service [60].

It appears that the AC system initially presented in [44] and later abstracted and referred to as Idemix in [56] has undergone significant internal development by IBM. These improvements aimed to provide richer functionalities and a more succinct system, as seen in the technical report [59]. After a decade, a subset of (improved) Idemix's functionalities was envisioned as suitable for integration into HLF. This led to the creation of a new open-source project [63], intended solely as an SDK for HLF, which is also called Idemix. These two projects continue to be developed independently (see https://github.com/hyperledger/fabric/issues/3356). Considering that the repository [63] implements only a small subset of AC's procedures tailored for HLF and does not encompass the full Idemix [56], we conclude that [63] cannot be regarded as a reference implementation of the original Idemix [56] for settings such as whistleblowing. Most importantly, [63] does not implement conditional de-anonymization, which, as we have argued, is a necessary requirement for a whistleblowing system. We conclude by stating that we are unaware of any available original Idemix [56] implementation.

Problems with anonymous credentials

Although using AC for whistleblowers' protection sounds attractive (as also indicated in [46]), its success is conditioned by many critical factors: availability, market adoption, open licensing, freedom from patents, social acceptance, etc. While Idemix and U-Prove meet the patent condition (the core patents have expired, and the products are licensed under MIT and Apache 2.0), progress on the first two factors is slow.

One important practical challenge lies in the computational overhead of anonymous credential systems, which makes them unsuitable for low-end devices. Furthermore, revocation mechanisms,

16 https://github.com/microsoft/uprove-csharp-sdk
17 https://github.com/microsoft/uprove-javascript-sdk
18 https://github.com/microsoft/uprove-node-reference

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

Table 2. Summary of issues in discussed pseudonymization primitives and tools.

Technology	Issues
Blind signatures	Reports from the same person cannot be linked. No support for de-psudonymization.
Restricted identification	No protection against report manipulation. No support for de-psudonymization.
Domain pseudonyms and signatures	Need for dedicated infrastructure. Missing support for algebraic groups with pairings.
Attribute certificates	Possible transfer of certificates.
Zero-knowledge proofs	A concrete scheme is not available (for both proposed cases).
Distributed ledger technology and blockchain	Some information can be inferred from observing the ledger.
Privacy pass	Vulnerable to flooding and Sybil attacks. Follow-up contact is not handled.
FIDO2	Depends on a trusted third party. No support for de-psudonymization when used on multiple devices.
Privacy coins	Depends on a trusted third party. Some information can be inferred from observing the ledger.
EU sanctions whistleblowing tool	Pseudonymity is not supported.
Paid black box private company tools	No information about internal implementation. Depends on a trusted third party.
U-prove	Inapplicable on standard smart cards and weak embedded devices. Available implementation at a low maturity level. Lack of design justification. No support for de-psudonymization.
Idemix	Inapplicable on standard smart cards and weak embedded devices. Available implementation at a low maturity level.

essential for real-world applications, remain either underdeveloped or unavailable at this stage.

Finally, the emergence of the EDIW [31] can become a game changer. One possibility is the integration of AC into EDIW. However, given the tight implementation timeline for EDIW, it is likely that the Member States will favor simpler solutions. This approach could risk rendering AC systems obsolete if EDIWs succeed in delivering a substantial portion of the desired functionalities. Table 2 lists all aforementioned technologies as well as their recognized issues.

Conclusions

In the field of IT security requirements, there is a trend of shifting from obligation-based security measures and penalties to the concepts of security by design and verifiability. So far, this approach has not been widely adopted for protecting whistleblowers despite the critical importance for the individuals involved. We believe that pseudonymization is an ideal solution, providing unconditional protection against retaliation that could persist even in the event of security breaches or misconduct.

Our analysis shows that the Directive and the implementing regulations do not follow the approach of security and privacy by design. They rather rely on penalties and prohibiting retaliation. Moreover, the doors are open for rejecting reports, if a whistleblower attempts to stay anonymous. There is no road map for improving security and trustworthiness based on strong technical mechanisms.

A good compromise between the target security, ease of use, and implementation costs should be found. If possible, the solutions intended for whistleblowing should be reused for other purposes to reduce the investment effort and make it easier for individuals to learn. For this reason, we have analyzed various cryptographic pseudonymity tools available on the market or designed for the electronic identity of the individuals. However, we conclude that all of them suffer from inherent flaws that make them not fully suitable for whistleblowing.

We emphasize that due to the absence of concrete security requirements (inter alia whistleblower's identity management system), from the industry's perspective, the call for solutions (e.g., paid third-party compliance whistleblowing tools) appears vague. As a result, attempts to develop compliant systems may vary significantly concerning the target protection level. Specifically, a system can formally comply with the Directive and appear effective while, in reality, it can have fundamental security and privacy flaws.

To summarize, we revisit the core questions posed in Section 'Introduction'. Our answer is that such a system adhering to the privacy by design principle is feasible but requires a more sophisticated approach. This likely involves integrating the aforementioned tools in an innovative and strategic manner. Once more, we emphasize that a strong focus on privacy by design principles is essential for implementing a successful, large-scale, EU-wide whistleblowing system.

In light of the results of our analysis, the future of whistleblowing systems may appear rather pessimistic. However, both our analysis and assumptions are based on scenarios where institutions (including governmental bodies) adopt a neglectful or even adversarial stance toward whistleblowers. It is worth emphasizing that whistleblowing practices existed in the EU even before the adoption of the Directive, and in the coming years perhaps we will observe a positive impact of the new legal protections for whistleblowers.

According to Article 27(3), the EU Commission will present a report on the impact of the Directive's implementation in December 2025. By then, the report's findings should be confronted with the results of feasibility studies on security by design via pseudonymization. Addressing this need remains an important direction for future work.

We believe that even a partial implementation of the security by design paradigm can significantly decrease overall system costs and seriously decrease potential whistleblower concerns, effectively increasing the number of submitted reports. Last but not least, a secure technical framework developed for whistleblowers' protection may be reused for purposes such as witness protection in judicial proceedings, public discussions in democratic societies, etc.

Author contributions

Mirosław Kutyłowski (Conceptualization [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Project administration [equal],

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

Supervision [equal], Validation [equal], Writing – original draft [equal], Writing – review & editing [equal]), Gabriel Wechta (Conceptualization [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Project administration [equal], Supervision [equal], Validation [equal], Writing – original draft [equal], Writing – review & editing [equal])

Supplementary data

Supplementary data is available at *Cybersecurity Journal* online.

Conflict of interest: None declared.

Funding

This research was funded by the National Science Centre, Poland under the OPUS call in the Weave programme [2023/51/I/ST6/02770]. For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

References

1. Nie JB, Elliott C. Humiliating Whistle-Blowers: Li Wenliang, the Response to Covid-19, and the Call for a Decent Society. *J Bioethic Inq* 2020;**17**:543–7. https://dx.doi.org/10.1007/s11673-020-09990-x
2. National Whistleblower Center. Whistleblowers You Should Know: Learn more about courageous individuals who have come forward to expose wrongdoing. Washington, D.C., USA: National Whistleblower Center, 2024. https://www.whistleblowers.org/whistleblowers
3. European Parliament and Council of the European Union. Directive (EU) 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report breaches of Union law. Brussels, Belgium: European Union, 2024. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L1937
4. Pacella JM. Making Whistleblowers Whole. *UC Irvine Law Rev* 2022;**12**:1291–1330
5. Eisenstadt LF, Pacella JM. Whistleblowers Need Not Apply. *Am Bus Law J* 2018;**55**:665–719. https://dx.doi.org/10.1111/ablj.12131
6. Cheliatsidou A, Sariannidis N, Garefalakis A. *et al*. Exploring attitudes towards Whistleblowing in relation to sustainable municipalities. *Adm Sci* 2023;**13**:1–16. https://dx.doi.org/10.3390/admsci13090199
7. Ziółkowska K. The (Legal) status of a Whistleblower in Poland. Selected Issues. *Studies in Logic, Grammar and Rhetoric* 2023;**68**:573–88. https://dx.doi.org/10.2478/slgr-2023-0033
8. NAVEX. Whistleblowing in Italy: The EU Whistleblower Protection Directive; Portland, OR, USA: NAVEX, 2021. https://cdn.navex.com/image/upload/v1635985206/media/EU_Directive_Article_Italy_Viewpoint_ENG.pdf
9. Teichmann FMJ, Wittmann C. Cultural aspects of the EU-Whistleblowing Directive. *J Financ Crime* 2022;**30**:999–1005. https://dx.doi.org/10.1108/jfc-04-2022-0091
10. Kun-Buczko M. Poles' attitudes to the concept of Whistleblowing. historical and present background. *Białostockie Studia Prawnicze* 2022;**27**:121–34. https://dx.doi.org/10.15290/bsp.2022.27.04.08
11. Pacella JM. The cybersecurity threat: compliance and the role of Whistleblowers. *Brooklyn J Corporate, Financ Comm Law*. 2016;**11**:39–70
12. Kaufmann J, Häferer K, Grimhardt K. The new EU Whistleblowing directive. *Comput Law Rev Int* 2019;**21**:14–7.
13. Teichmann FM, Wittmann C. Whistleblowing: procedural and dogmatic problems in the implementation of directive (EU) 2019/1937. *J Financ Regul Comp* 2022;**30**:553–66. https://dx.doi.org/10.1108/JFRC-12-2021-0118
14. Scheu H. Pitfalls in implementing the EU Whistleblower directive. *Wroclaw Review of Law, Administration & Economics* 2024;**14**:52–68. https://dx.doi.org/10.2478/wrlae-2024-0002
15. Juels A, Catalano D, Jakobsson M. Coercion-Resistant Electronic Elections; New York, NY, USA: ACM, 2002. Cryptology ePrint Archive, Paper 2002/165.
16. European Parliament. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). Brussels, Belgium: European Parliament, 2016.
17. European Data Protection Board. Guidelines 01/2025 on Pseudonymisation. Brussels, Belgium: European Data Protection Board, 2025. Public consultation period: 17 January–14 March 2025.
18. Berendt B, Schiffner S. Whistleblower protection in the digital age–why 'anonymous' is not enough. From technology to a wider view of governance. *Int Rev Inf Ethics* 2022;**31**:1–16. https://dx.doi.org/10.48550/ARXIV.2111.02825
19. Schiffner Agata Malik-Bosak. Whistleblower law across Europe—country list. 2024. https://wemoral.com/blog/whistleblower-law-across-europe-country-list. (18 July 2025, date last accessed).
20. Staufer D, Pallas F, Berendt B. Silencing the risk, not the whistle: a semi-automated text sanitization tool for mitigating the risk of Whistleblower re-identification. 2024. https://dx.doi.org/10.48550/ARXIV.2405.01097
21. University of South Carolina Audit and Advisory Services America's First Whistleblowers. Columbia, SC, USA: University of South Carolina Audit and Advisory Services, 2021.
22. Nader R. *Unsafe at any speed: the designed-in dangers of the American Automobile*. New York, USA: Grossman Publishers, 1965.
23. Whistleblower Network News. 'An Unheard of Dream': Ralph Nader's 50 Years in Whistleblowing. Washington, DC, USA: Whistleblower Network News, 2021.
24. Banisar D. Whistleblowing: international standards and developments. In: Sandoval I. (ed.), *Corruption and Transparency: Debating the Frontiers between State, Market and Society*. Washington, DC, USA: World Bank–Institute for Social Research, UNAM, 2011.
25. Chen Y, Zheng S, Wang H. *et al*. 3D-Speaker-Toolkit: an open source Toolkit for multi-modal speaker verification and diarization. Piscataway, NJ, USA: IEEE, 2024.
26. Jung Jw, Zhang W, Shi J. *et al*. *ESPnet-SPK: full pipeline speaker embedding toolkit with reproducible recipes, self-supervised front-ends, and off-the-shelf models*. Laxenburg, Austria: ISCA, 2024.
27. Camenisch J, Lysyanskaya A. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. Online Cryptology ePrint Archive, 2001.
28. Chen L, Jordan S, Liu YK. *et al*. *Report on post-quantum cryptography*. Gaithersburg, MD, USA: NIST, 2016. https://dx.doi.org/10.6028/nist.ir.8105
29. Kutyłowski M, Shao J. *Signing with multiple ID's and a single key*. Piscataway, NJ, USA: IEEE, 2011, 519–20. https://dx.doi.org/10.1109/CCNC.2011.5766527
30. Bundesamt für Sicherheit in der Informationstechnik. *Technical Guideline TR-03110 v2.21–Advanced Security Mechanisms for Machine Readable Travel Documents and eIDAS Token*. Bonn, Germany: Bundesamt für Sicherheit in der Informationstechnik, 2016.
31. The European Commission. *Regulation (EU) 2024/1183 of the European Parliament and of the Council*. Luxembourg: Publications Office of the European Union, 2024. https://eur-lex.europa.eu/eli/reg/2024/1183/oj. (12 November 2024, date last accessed).
32. The Tor Project. The Tor Project. Online The Tor Project. 2023. https://www.torproject.org/. (12 November 2024, date last accessed).
33. Bundestag. *Act for better protection of whistleblowers (Whistleblower Protection Act - HinSchG) (German)*. Berlin, Germany: Bundestag, 2023. (29 July 2024, date last accessed).
34. Polish Parliament. *ACT of June 14, 2024 on the protection of whistleblowers. (Polish)*. Warsaw, Poland: Polish Parliament, 2024. (6 August 2024, date last accessed).
35. Court of Justice of the European Union. OPINION OF ADVOCATE GENERAL SPIELMANN, European Data Protection Supervisor v

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

Single Resolution Board. Court of Justice of the European Union, 2025. https://curia.europa.eu/juris/document/document.jsf?text=&docid=295078&pageIndex=0&doclang=en&mode=req&dir=&occ=first&part=1&cid=28925561
36. Chaum D. Blind signatu`res for untraceable payments. In: Chaum D, Rivest RL, Sherman AT, (eds), *Advances in Cryptology*. Boston, MA: Springer US, 1983, 199–203.
37. Kutyłowski M, Krzywiecki Ł, Kubiak P. *et al.* Restricted identification scheme and diffie-hellman linking problem. In: Chen L, Yung M, Zhu L, (eds), *Trusted Systems*. Berlin, Heidelberg: Springer, 2012, 221–38.
38. Kutyłowski M, Hanzlik L, Kluczniak K. Pseudonymous signature on eIDAS token - implementation based privacy threats. In: Liu JK, Steinfeld R, (eds), *Information Security and Privacy - 21st Australasian Conference, ACISP 2016, Melbourne, VIC, Australia, July 4-6, 2016, Proceedings, Part II. Vol. 9723 of Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, 2016, 467–77. https://dx.doi.org/10.1007/978-3-319-40367-0_31
39. Nguyen KQ, Bao F, Mu Y. *et al.* Zero-Knowledge Proofs of Possession of Digital Signatures and Its Applications. In: Varadharajan V, Mu Y, (eds), *Information and Communication Security, Second International Conference, ICICS'99, Sydney, Australia, November 9-11, 1999, Proceedings. Vol. 1726 of Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, 1999, 103–18. https://dx.doi.org/10.1007/978-3-540-47942-0_9
40. Nitulescu A. zk-SNARKs: A Gentle Introduction. Online Self-published. 2020. https://www.di.ens.fr/~nitulesc/files/Survey-SNARKs.pdf
41. Li C, Zhang F. Designated-Verifier zk-SNARKs Made Easy. Online Cryptology ePrint Archive. Paper 2024/1153. 2024.
42. Merkle RC. A Digital Signature Based on a Conventional Encryption Function. In: *Advances in Cryptology — CRYPTO '87*. Berlin Heidelberg: Springer, 1987, 369–378.
43. Turner S, Farrell S, Housley R. An Internet Attribute Certificate Profile for Authorization. Mountain View, CA, USA: IETF, 2010. RFC 5755. https://dx.doi.org/10.17487/RFC5755
44. Pape S. A survey on non-transferable anonymous credentials. In: Matyáš V, Fischer-Hübner S, Cvrček D, Švenda P, (eds), *The Future of Identity in the Information Society*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, 107–18.
45. Wang W, Hoang DT, Hu P. *et al.* A survey on consensus mechanisms and mining strategy management in Blockchain Networks. *IEEE Access* 2019;**7**:22328–70. https://dx.doi.org/10.1109/ACCESS.2019.2896108
46. Baum C, Blazy O, Camenisch J. *et al.* Cryptographers' Feedback on the EU Digital Identity's ARF. Technical report. Technical University of Denmark, École Polytechnique, Dfinity, Karlstad University, Radboud University, Brown University, Hasso-Plattner-Institute, Johannes Kepler University Linz, Northeastern University, Universität der Bundeswehr München, University of Washington, Partisia, EPFL, 2024.
47. European Commission. *European Digital Identity Wallet: Architecture and Reference Framework*. Luxembourg: Publications Office of the European Union, 2024. https://eu-digital-identity-wallet.github.io/eudi-doc-architecture-and-reference-framework/1.4.0/arf. Version 1.4.0. (25 November 2024, date last accessed).
48. Davidson A, Goldberg I, Sullivan N. *et al.* Privacy pass: bypassing internet challenges anonymously. *Proceedings on Privacy Enhancing Technologies* 2018;**2018**:164–80. https://dx.doi.org/10.1515/popets-2018-0026
49. FIDO Alliance. *Overview of the FIDO alliance*. Foster City, CA, USA: FIDO Alliance, 2023. https://fidoalliance.org/fido2/. (13 November 2024, date last accessed).
50. Young A, Yung M. The dark side of "black-box" cryptography or: should we trust capstone?. In: Koblitz N, (ed.), *Advances in Cryptology—CRYPTO '96*. Berlin, Heidelberg: Springer, 1996, 89–103.
51. Koblitz Koe, Alonso KM, Noether S. Zero to Monero: Second Edition. Online Self-published. 2020. https://www.getmonero.org/library/Zero-to-Monero-2-0-0.pdf
52. Ben-Sasson E, Chiesa A, Garman C. *et al.* Zerocash: decentralized anonymous payments from Bitcoin. Online Self-published. 2014. https://www.allcryptowhitepapers.com/zcash-whitepaper/
53. European Parliament and Council of the European Union. *Directive (EU) 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report breaches of Union law*. Brussels, Belgium: European Union, 2019. (14 November 2024, date last accessed).
54. The European Commission. *EU sanctions compliance reporting platform*. Luxembourg: Publications Office of the European Union, 2024. (14 November 2024, date last accessed).
55. Paquin C, Zaverucha G. *U-prove technology overview*. Redmond, WA, USA: Microsoft Corporation, 2024. https://github.com/microsoft/uprove-node-reference/blob/main/doc/U-Prove%20Technology%20Overview%20V1.1%20Revision%203.pdf. (18 November 2024, date last accessed).
56. Camenisch J, Herreweghen EV. Design and implementation of the idemix anonymous credential system. In: *Proceedings of the 9th ACM Conference on Computer and Communications Security*. New York, NY, USA: ACM, 2002, 21–30. https://dx.doi.org/10.1145/586110.586114
57. Paquin C, Zaverucha G. *U-Prove Cryptographic Specification*. Redmond, WA, USA: Microsoft Corporation, 2014. (18 November 2024, date last accessed).
58. Hanzlik L, Kubiak P, Kutyłowski M. *Tracing attacks on U-Prove with revocation mechanism*. New York, NY, USA: ACM, 2015.
59. IBM Research. *Specification of the Identity Mixer Cryptographic Library Version 2.3.0*. Yorktown Heights, NY, USA: IBM Research, 2009. (13 December 2024, date last accessed).
60. Hyperledger Fabric Project. *MSP Implementation with Identity Mixer*. San Francisco, CA, USA: The Linux Foundation, 2024. https://hyperledger-fabric.readthedocs.io/en/latest/idemix.html. (27 November 2024, date last accessed).
61. ERCIM. *PrimeLife Project*. Zurich, Switzerland: PrimeLife consortium, 2011. http://primelife.ercim.eu/. (13 December 2024, date last accessed).
62. CryptoExperts. *ABC4Trust Research Project*. Paris, France: CryptoExperts. https://www.cryptoexperts.com/research/projects/abc4trust/. (13 December 2024, date last accessed).
63. IBM. Idemix: Identity Mixer. 2024. https://github.com/IBM/idemix. Commit: 7140f9249264b1860f6ca307ca927c1ef419c0be.

Downloaded from https://academic.oup.com/cybersecurity/article/11/1/tyaf028/8294137 by guest on 12 May 2026

© The Author(s) 2025. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.