Position Paper: The Past, Present, and Future of Sanitizable and Redactable Signatures

Arne Bilzhauser*, Henrich C. Pöhls† and Kai Samelin‡
*Chair of IT-Security and Institute of IT-Security & Security Law (ISL), University of Passau, Passau, Germany
†IBM Research – Zurich, Rüschlikon, Switzerland and TU Darmstadt, Darmstadt, Germany
Email: {ab.hp@sec.uni-passau.de}, ksa@zurich.ibm.com

Abstract—Sanitizable signature schemes (SSS), as well as redactable signature schemes (RSS), gained a lot of attention in the recent past. In a nutshell, both types of signature schemes allow to alter signed data in a controlled way by a, potentially semi-trusted, third party. The resulting signatures still verify. Thus, the authenticity of the subsequently modified content is preserved. In this position paper, we discuss the state-of-the-art, and highlight potential future research opportunities. We hope this work gives rise to additional ideas, real-life use-cases, and interesting upcoming research, helping to bring both primitives into practice. Hence, this paper is meant as a starting point for readers interested in these primitives, looking for new research and application opportunities. In other words, we think that both primitives deserve further attention.

I. INTRODUCTION

Traditional digital signatures become invalid as soon as a single bit of the protected message is changed [71]. However, there is a plethora of real-life use-cases where the alteration of a signed message by an external party is of paramount importance. In the best case, this is possible even without additional interaction with the original signer. Examples of such use-cases include privacy-preserving handling of patient-data, secure routing, work-flows, privacy-preserving document disclosure, privacy credentials, social networks, web-services, data, secure routing, work-flows, privacy-preserving document disclosure, privacy credentials, social networks, web-services, and blank signatures [5], [15], [33], [34], [35], [37], [42], [54], [55], [77], [107], [108], [115], [122], [124].

Sanitizable signature schemes (SSS) and redactable signature schemes (RSS) are primitives allowing for such controlled, and non-interactive, alterations. In a nutshell, RSSs allow to remove, i.e., redact, parts of a signed message, while SSSs allow a designated third party, named the sanitizer, to change, i.e., sanitize, signer-chosen parts of a signed message to different bitstrings. However, even though RSSs and SSSs have a lot of convincing application scenarios, they have not been deployed in practice yet. This is a pity, as in our opinion, both primitives have gained a lot of interest in the recent past from the scientific community, while existing prototype implementations show that both primitives are sufficiently efficient, even from a purely practical point of view. Thus, they can be considered “mature”, even though not all opportunities are exhausted yet.

A. Contribution

In this position paper, we review the state-of-the-art of RSSs and SSSs. We also compare both primitives to other related primitives. We point out additional real-life use-cases, but also where they can (and potentially have to) be extended to allow for more application scenarios. This may help to find additional research opportunities, and to increase the visibility of these, in certain contexts very useful, primitives. This may be a stepping stone to bring them into practice. Summarized, this paper is meant as an easy-to-read starting point into this interesting topic, proposing additional real-life use-cases, and new research opportunities.

For example, we ask if it is possible to combine RSSs and SSSs, if it possible to have an SSS which is unlinkable and (strongly) invisible simultaneously, or how to construct multi-redacter/multi-signer accountable RSSs, and if SSSs may help to tackle the DNS-enumeration problem [70].

B. Related Primitives

As already pointed out, both RSSs and SSSs, allow to partially delegate signing-rights to other parties. This section is devoted to give an overview over different, yet related, primitives. We only sketch their possibilities, as they have a slightly different focus. In particular, SSSs and RSSs are explicitly designed to alter, i.e., edit, signed data, while the primitives presented now are more meant to “derive” or create new signatures. The current state-of-the-art of RSSs and SSSs (as editable signatures) is discussed in their own sections, i.e., Section III and Section IV.

Due to space constraints, we can only give a very short overview over the related primitives. However, there are already comprehensive overviews, which cover the mentioned primitives in much greater detail [1], [30], [53], [67], [68].

1) Proxy Signature Schemes: Proxy signatures [24], [100] allow to “delegate” signing rights to another party. The second party can then sign messages on behalf of the first entity. However, in the current definition, proxy signatures require that the delegated signing key is generated using an interactive algorithm, while the message space from which the delegatee is allowed to sign messages from is fixed after that setup.

2) Policy-Based Signatures: An additional related primitive are policy-based signature schemes [17]. Here, a trusted authority issues the signing key for each participant, which can only issue signatures for the message space corresponding to the policy encoded into the signing key, while the signatures do not reveal the policy itself.

3) Functional Signatures: They are also signatures where a master signing key can be used to derive restricted so called secondary signing keys [28]. These secondary keys are parametrized by a function \( f \) and such a key can only sign
messages in the range of \( f \), i.e., given any \( m \), a key \( k \) for function \( f \) only allows to produce signatures for \( f(m) \).

4) Blank Digital Signature Schemes: In blank digital signature schemes [55, 77, 123], an originator can sign a template which allows the proxy (the party receiving the secret key) to sign a message based on this template. A template can be best interpreted as a document with single/multiple-choice fields.

5) Attribute-Based Signature Schemes: Attribute-based signatures [99] allow users to sign messages which are bound to specific attributes. This type of signature scheme requires that the signer holds the corresponding attributes. More precisely, a signature verification only reveals that the signer has a signing key which fulfills the predicate on the claimed attributes.

6) Transitive Signature Schemes: Assume we have a graph \( G = (V, E) \), and each edge \( e_{i,j} \in E \) is signed. A transitive signature scheme [20], [102], [125] allows to derive a signature \( \sigma_{i,k} \), if there is a path from \( v_i \in V \) to \( v_k \in V \), even though only signatures for the edges are known. In other words, one can derive signatures on the transitive closure of the graph \( G \).

7) Homomorphic Signature Schemes: Homomorphic signature schemes [26], [27], [74], [84] also allow to derive new signatures from already existing ones. Examples include signing the mean value of grades, or the standard deviation. Related are \( \mathcal{P} \)-homomorphic signatures [1], where the message derived needs to follow some predicate \( \mathcal{P} \), and can be seen as a generalization of RSSs (See Section IV), even though they allow for way more possibilities. Stronger security (in particular privacy) definitions, and impossibility results, have also been presented [7], [8], [52].

8) Additional Related Signature Schemes: There are a lot more related primitives, such as group signature schemes [19], [21], [47], [59], append-only signature schemes [86], threshold signature schemes [60] (and specializations thereof [14], [25], [38]), traceable signature schemes [85], direct anonymous attestation [29], blind signature schemes [46], and generalizations of some of the ideas mentioned [9], [45]. We stress that this list is not exhaustive, but we think that this list provides a good overview, especially as a starting point for the interested reader.

II. Notation

The main security parameter is denoted by \( \lambda \in \mathbb{N} \). All algorithms implicitly take \( \lambda \) as an additional input. We write \( a \leftarrow A(x) \) if \( a \) is assigned to the output of algorithm \( A \) with input \( x \). An algorithm is efficient if it runs in probabilistic polynomial time (ppt) in the length of its input. For the remainder of this paper, all algorithms are ppt if not explicitly mentioned otherwise. Most algorithms may return a special error symbol \( \bot \notin \{0,1\}^* \), denoting an exception. If \( S \) is a set, we write \( a \leftarrow S \) to denote that \( a \) is chosen uniformly at random from \( S \). For a message \( m = (m[1], m[2], \ldots, m[\ell]) \), we call \( m[i] \) a block, while \( \ell \in \mathbb{N} \) denotes the number of blocks in a message \( m \). For a list we require that we have a unique, injective, and efficiently reversible, encoding, which maps the list to \( \{0,1\}^* \).

III. Sanitizable Signature Schemes

In this section, we discuss sanitizable signatures schemes (SSS). SSSs have been introduced by Ateniese et al. [5]. When signing, the signer determines which blocks \( m[i] \) of the message \( m = (m[1], m[2], \ldots, m[i], \ldots, m[\ell]) \) can be modified (i.e., are admissible). Any such admissible block can be altered to a new bitstring \( m[i]' = \{0,1\}^* \), where \( i \in \{1,2,\ldots,\ell\} \), by a semi-trusted sanitizer, also chosen by the signer. This sanitizer holds his own public key. The sanitization process requires the corresponding private key, but does not require the signer’s involvement. Sanitization of a message \( m \) then results in an altered message \( m' = (m[1]', m[2]', \ldots, m[i]', \ldots, m[\ell]') \), where \( m[i] = m[i]' \) for every non-admissible block, and also a signature \( \sigma' \), which verifies under the given public keys. Hence, authenticity of the message \( m' \) is still ensured. An example workflow is visualized in Figure 1, and Figure 2.

A. The Framework for Sanitizable Signature Schemes

Subsequently, we introduce the framework for SSSs. The definitions are essentially the ones given by Camenisch et al. [36], which are itself based on existing work [31]. However, we need to set some additional notation beforehand. The variable \( adm \) contains the set of indices of the modifiable blocks, as well as \( \ell \) denoting the total number of blocks in the message \( m \). For example, let \( adm = \{(1,2,4), \ell\} \). Then, \( m \) must contain four blocks, and all but the third are admissible. The variable \( MOD \) is a set containing pairs \((i, m[i]'\)} for those blocks that are modified, meaning that \( m[i] \) is replaced by \( m[i]' \).

Definition 1 (Sanitizable Signatures): A sanitizable signature scheme SSS consists of the ppt algorithms (SSSPGen, KGen\textsubscript{sig}, KGen\textsubscript{san}, Sign, Sanit, Verify, Proof, Judge) such that:

SSSPGen. The algorithm SSSPGen, on input security parameter \( \lambda \), generates the public parameters:

\[ \text{pp}_{\text{sss}} \leftarrow \text{SSSPGen}(\lambda^{'}) \]

We assume that \( \text{pp}_{\text{sss}} \) is implicitly input to all other algorithms.

KGen\textsubscript{sig}. The algorithm KGen\textsubscript{sig} takes the public parameters \( \text{pp}_{\text{sss}} \), and returns the signer’s private key and the corresponding public key:

\[ (sk_{\text{sig}}, pk_{\text{sig}}) \leftarrow \text{KGen}_{\text{sig}}(\text{pp}_{\text{sss}}) \]

KGen\textsubscript{san}. The algorithm KGen\textsubscript{san} takes the public parameters \( \text{pp}_{\text{sss}} \), and returns the sanitizer’s private key as well as the corresponding public key:

\[ (sk_{\text{san}}, pk_{\text{san}}) \leftarrow \text{KGen}_{\text{san}}(\text{pp}_{\text{sss}}) \]

Sign. The algorithm Sign takes as input a message \( m \), \( sk_{\text{sig}}, pk_{\text{san}} \), as well as a description \( adm \) of the admissible blocks. If \( adm(m) = \text{false} \), this algorithm returns \( \bot \). It outputs a signature:

\[ \sigma \leftarrow \text{Sign}(m, sk_{\text{sig}}, pk_{\text{san}}, adm) \]
Sanitize. The algorithm Sanitize takes a message \( m \), modification instruction \( \text{MOD} \), a signature \( \sigma \), \( \text{pk}_{\text{sig}} \), and \( \text{sk}_{\text{san}} \). It outputs \( m' \leftarrow \text{MOD}(m) \) together with \( \sigma' \):

\[
(m', \sigma') \leftarrow \text{Sanitize}(m, \text{MOD}, \sigma, \text{pk}_{\text{sig}}, \text{sk}_{\text{san}})
\]

Here, \( m' \leftarrow \text{MOD}(m) \) means that the message \( m \) is modified to \( m' \) according to the modification instruction \( \text{MOD} \).

Verify. The algorithm Verify takes as input the signature \( \sigma \) for a message \( m \) w.r.t. the public keys \( \text{pk}_{\text{sig}} \) and \( \text{pk}_{\text{san}} \). It outputs a decision \( d \in \{\text{true, false}\} \):

\[
d \leftarrow \text{Verify}(m, \sigma, \text{pk}_{\text{sig}}, \text{pk}_{\text{san}})
\]

Proof. The algorithm Proof takes as input \( \text{sk}_{\text{sig}} \), a message \( m \), a signature \( \sigma \), a set of polynomially many additional message/signature pairs \( \{(m_i, \sigma_i)\} \), and \( \text{pk}_{\text{san}} \). It outputs a string \( \pi \in \{0, 1\}^* \) which can be used by the Judge to decide which party is accountable given a message/signature pair \( (m, \sigma) \):

\[
\pi \leftarrow \text{Proof}(\text{sk}_{\text{sig}}, m, \{ (m_i, \sigma_i) \mid i \in \mathbb{N} \}, \text{pk}_{\text{san}})
\]

Judge. The algorithm Judge takes as input a message \( m \), a signature \( \sigma \), \( \text{pk}_{\text{sig}} \), \( \text{pk}_{\text{san}} \), as well as a proof \( \pi \). Note, this means that once a proof \( \pi \) is generated, the accountable party can be derived by anyone for that message/signature pair \( (m, \sigma) \). It outputs a decision \( d \in \{\text{Sig, San}\} \), indicating whether the message/signature pair has been created by the signer, or the sanitizer:

\[
d \leftarrow \text{Judge}(m, \sigma, \text{pk}_{\text{sig}}, \text{pk}_{\text{san}}, \pi)
\]

B. The Past

The idea of sanitizable signature schemes was introduced by Ateniese et al. [5]. This basic scheme is based upon standard digital signature schemes and chameleon-hashes\(^1\) [88].

\(^1\)Chameleon-hashes are a special type of collision-resistant hashes, where hashes are computed using a public key and knowledge of the corresponding secret key allows to find arbitrary collisions.

In a nutshell, each admissible block is hashed using such a chameleon-hash, while each non-admissible block is hashed using a standard cryptographic hash. Then, all hashes are concatenated and signed using the standard digital signature scheme. The main idea is now, that the sanitizer holds the secret key to the public hashing key and thus can find collisions in the chameleon-hash, which directly corresponds to sanitizing. We stress that the original construction is not secure, as it is neither unforgeable nor accountable in the security model by Brzuska et al. [31], which is in wide use.

1) Standard Security Properties: The first approach to formalize the ideas by Ateniese et al. [5] was given by Brzuska et al. [31]. In particular, they formalized the following properties, which we only sketch informally as it is sufficient for the purpose of this paper.

- **Unforgeability.** If one does not hold any secret keys, one should not be able to come up with a verifying signature \( \sigma^* \) for a message \( m^* \) which has neither been authorized by the signer nor the sanitizer.
- **Privacy.** An outsider not holding any private keys should not be able to derive any information about sanitized parts of a message.
- **Transparency.** Transparency is even stronger than privacy. Namely, an outsider not holding any private keys should not even be able to decide whether a signature has been created using the original signing algorithm or through the sanitization algorithm.
- **Immutability.** Immutability requires that only admissible parts of a message can be altered by the sanitizer.
- **Accountability.** If it comes to a dispute, the signer must be able to generate a proof which points to the accountable party. Accountability requires that neither the signer nor the sanitizer can generate “false” proofs pointing to the wrong party for a adversarially generated message/signature pair \( (m^*, \sigma^*) \).

C. The Present

1) Additional Security Properties: The above security properties were later extended to cover more use-cases. In
particular, the initial properties were extended for:

- **Unlinkability.** Unlinkability, as a very strong privacy notion, has been introduced by Brzuska et al. [33]. This property guarantees that a sanitized signature does not leak from which signature it was created. Thus, even if the original signature is known, it becomes hard to decide if two signatures are linked together (hence the name). Constructs (and stronger definitions) have gained recent attention [35], [63], [95].

- **Non-Interactive Public-Accountability.** From a privacy perspective, transparency seems to be a necessity. However, it turned out that certain use-cases, e.g., for legal compliance, transparency is actually counterproductive [79]. In particular, it may be necessary that it is visible if a signature was sanitized or not. Non-interactive public-accountability is exactly the property achieving this, and was introduced by Brzuska et al. [34], with a construction derived from [32]. Clearly, this property is mutually exclusive to transparency. It can also be facilitated to enforce a multiple-eyes principle [23].

- **Invisibility.** In the original paper, Ateniese et al. [5] introduced the notion of “strong transparency”. This notion guarantees that an outsider not holding any private keys cannot derive which blocks are admissible. However, as shown by Pöhl et al. [114], (standard) transparency is not related to this property. Then, Camenisch et al. [36] formalized this idea, and coined it “invisibility”, which has, quite recently, been strengthened by Beck et al. [13].

- **Contingency.** In 2009 the “dual of integrity” [116] termed contingency [16] has been introduced. Duality means here that something can be proven to be in only one of two states, in this case data either has “integrity” or it is “contingent”. Hence, contingency describes the verifiable state that the data’s integrity is intended to be unknown, which can be used to circumvent some legal challenges [116], [16], [107]. Pöhl proposed in [107] to technically build this using an SSS.

- **Group-Level Definitions.** All the above definitions are either defined on a per message/signature pair basis, or for each block [34]. However, to improve efficiency, Pöhl et al. proposed to group blocks [49], which also extends to some of the security properties mentioned earlier.

- **Stronger Security Definitions.** The definitions by Brzuska et al. [31] only take the message and signature account. However, as shown by Gong et al. [72], their security model still allows to “fake” adm. Gong et al. thus argue that protecting adm may be necessary, and present a new security framework [72] achieving this. Recently, Krenn et al. [89] further strengthened the security model by also taking the signature into account, much like the difference between unforgeable [71] and strongly unforgeable standard signatures [4], [80]. Then, de Meer et al. showed how to transform SSSs into RSSs, if certain privacy definitions are altered in a non-standard way [50].

2) **Multiple Signers and Sanitizers:** As already clear from the given framework, SSSs are normally defined in a way that only one signer and one sanitizer are considered. This may not be enough for certain use-cases. Thus, the idea of trapdoor SSSs has been introduced by Canard et al. [41]. In this primitive, the signer can, after signature generation, grant additional sanitizers the possibility to sanitize a signature. Later, this primitive has been extended to allow for more efficient instantiations, and accountability [32], [94], [126]. This has later been extended to also cover multiple signers and unlinkability [40].

3) **Limiting the Sanitizer:** In the above schemes, the sanitizer is allowed to alter the admissible blocks to arbitrary bitstrings. Clearly, this is not always wanted. To tackle this situation, new schemes have been proposed which allow to limit the sanitizer to signer-chosen values for certain blocks [39], [58], [87].

4) **Sanitization of Encrypted Data:** Recently, the idea of sanitizing encrypted data was introduced [11], [48], [62], [65] and also a patent in this area exists [66]. In this primitive, the sanitizer does not learn which data is sanitized, i.e., they add an additional layer of confidentiality.

5) **Implementations:** To prove that SSSs are practical, they were implemented on a wide variety of devices, ranging from smart cards [110] to potent desktop PCs [13], [49], even in very different contexts such as XML-signatures [114]. All implementations presented so far show that SSSs can be considered efficient.

D. The Future

1) **Legal Implications and Standardization:** On the one hand, from a purely practical perspective, it is necessary to know which legal consequences arise, if SSSs are deployed in real-world application. As it turned out, non-interactive public-accountability seems to be necessary [109], [79], [?] to achieve a level of trust which meets legal requirements, as (qualified) digital signatures are equivalent to handwritten ones in court. However, as SSSs have not yet been deployed, there is still a lot of discussion missing here.

On the other hand, there are a lot of competing security definitions with subtleties that are of importance, as this section has already shown. We think that standardizing SSSs may help here to establish a minimum baseline for a common understanding especially for legal arguments.

2) **New Application Scenarios:**

- **Receipts.** Assume that you buy two items at a store, e.g., a TV and a toaster. However, at home it turns out that the wrong TV was bought. Normally, one returns to the store, and exchanges the TV for a different one. Then, however, the receipt needs to be changed to prevent fraud. SSSs can help here: if an item is exchanged, the receipt can simply be altered for the new item, without the need to give the customer two different receipts. This kind of idea may be useful in other contexts as well, e.g., supply-chain management.

- **Data Loss Prevention.** The main challenges for data loss prevention are clearly confidentiality and availability of data. However, one form of data loss is the so called data leakage that can not only affect data at rest (e.g. theft of database entries), but also data in motion (e.g. data sent by mail, ftp, or peer-to-peer) [97]. And in many cases the

---

2However, Brzuska et al. explicitly require that adm is always correctly recoverable [31].
integrity and authenticity of sent messages are crucial for the receiving side of communication. Assume that some data is classified to be for internal use only and hence never should leave your network. The first and most obvious approach would be to never send the data. This would require all parts of a system (e.g., applications and/or users) to be aware of the data classification and the according rules (which might simply not be possible in many systems). A second approach would be to implement a central entity that checks the outgoing communication and alters or removes all classified data. In case of standard signatures, two problems arise for the integrity and authenticity of the sent messages: either the central entity invalidates existing signatures (i.e., by altering messages signed by the original data source), or the central entity has to sign the message again with its private key after a change was performed. In the latter case, the trust relation between receiver and data source cannot be verified anymore (i.e., the altered version of the message was never signed by the original data source). SSSs can help here: by making the central entity an authorized sanitizer, changes on outgoing messages can be performed without invalidating the signature, while the trust relation between receiver (verifier) and original data source (signer) can be preserved. Notice that RSSs could be used in a similar manner, however this will limit the options for the central entity (i.e., only removal of message parts is possible). A related approach [22] was presented for web-services in combination with a proxy server (sanitizer).

- Preventing DNS-Zone Enumeration. The DNS-Zone enumeration-problem essentially says that certain DNS-extensions, such as DNSSEC, allow to enumerate DNS-zones [70] by provoking (negative) responses from DNS-servers, eventually allowing to enumerate all zones. As also proven by Goldberg et al. [70], there is no simpler solution than public-key cryptography to this problem. Thus, SSSs may offer a direct solution here: each response is pre-computed for each domain by the corresponding zone’s secret key. For non-existing zones, however, the DNS-server queried sanitizes a (pre-computed and signed by the root’s DNS-server) message \(m = \text{domain} \text{ "BLANK";}\) does not exist to \(m' = \text{domain} \text{ "QUERY";}\) does not exist, where QUERY is the domain to be resolved, and only \(m[2]\) is admissible. To prevent “false” responses by corrupt servers, one may restrict \(m[2]\) to be an non-existing domain, which can be set up by the administrator. This can, e.g., be achieved using the “restricting the sanitizer”-extensions discussed earlier.

3) Future Research Opportunities and Questions:

- Seals. The current definition of transparency prohibits that a legitimate sanitization makes the verifier able to decide whether that action took place. However, it may also be useful to have an extension which allows that, if certain signer-chosen admissible blocks are sanitized, transparency is lost. This may, e.g., be the case in a break-glass scenario, where certain changes to a signed message are necessary, as extraordinary circumstances force the sanitizer to do this, e.g., if a name of a list is pseudonymized to account for data protection laws. A natural extension is to also hide which changes would break that “seal” – only if that seal is broken, the verifier then learns where the seal was. An even stronger definition would only reveal that the seal was broken, but not where.

- Invisible and Unlinkable SSSs. There are already invisible [36] and unlinkable [33] SSSs. However, there are no constructions which achieve both properties simultaneously. Thus, it is clearly a question if both security properties can be combined. It is also an open question, if strongly unlinkable [35], yet transparent SSSs can be constructed.

- Black-Box Accountability. SSSs can also be used to make RSSs accountable [112]. It is therefore a natural question, if they can be used to make other (non-accountable) primitives accountable as well, maybe even using the same underlying construction idea, and for multiple sanitizers.

- Levels of Accountability. For SSS the non-interactive public form of accountability has been introduced [34]. It prohibits the property of transparency that has been widely discussed. Still, it remains an open question if there are useful applications for notions between these two extremes, and if such a notion can be extended to one of a SSS’s core property: privacy, i.e., whether there exists an application where an outsider must know whether a block/message was altered (and not the signature, as for non-interactive public-accountability), but cannot derive any additional useful information.

- Time-Locked Sanitization. It may also be interesting if it is possible to construct SSSs which allow that only at some later point in time the sanitizer can sanitize. This can even be extended in way such that at some later point in time sanitization is no longer possible.

IV. REDACTABLE SIGNATURE SCHEMES

As already clarified, RSSs allow to remove (and only remove) certain parts of a message \(m\), without interaction with the original signer. This redaction can be performed by any party, i.e., no additional key pair is required. As for SSSs, the resulting signatures still verify. An example workflow for RSSs is given in Figure 3.

A. The Framework of Redactable Signature Schemes

The following definitions for RSSs are compiled from [30], [118], but extended to support parameter generation to match the definitions of SSSs.

- Definition 2 (Redactable Signature Scheme (RSS)): A redactable signature scheme RSS consists of five algorithms, i.e., (RSSPGen, KGen\(_{\text{sig}}\), Sign, Redact, Verify), such that:

  \[
  \text{RSSPGen. The algorithm RSSPGen generates the public parameters:}
  \]

  \[
  \text{pp}\_{\text{rss}} \leftarrow \text{RSSPGen}(1^\lambda)
  \]

  We assume that \(\text{pp}\_{\text{rss}}\) is implicitly input to all other algorithms.

  \[
  \text{KGen}_{\text{sig}}. The signer generates a key pair, based on the security parameter \(\lambda\):
  \]

  \[
  (\text{pk}_{\text{sig}}, \text{sk}_{\text{sig}}) \leftarrow \text{KGen}_{\text{sig}}(\text{pp}_{\text{rss}})
  \]
Signer. The Sign algorithm takes as input a message \( m \), adm, and \( sk_{\text{sig}} \). It outputs a signature \( \sigma \), and some redaction information \( \text{red} \):
\[
(\sigma, \text{red}) \leftarrow \text{Sign}(m, sk_{\text{sig}}, \text{adm})
\]

Redact. Algorithm Redact takes a message \( m \), MOD, redaction information \( \text{red} \), a signature \( \sigma \), and \( pk_{\text{sig}} \). It modifies the message \( m \) according to the modification MOD. Redact outputs \( m' \) for each \( m, \sigma \), and \( \text{red} \):
\[
(m', \sigma', \text{red}') \leftarrow \text{Redact}(m, \text{MOD}, \sigma, \text{red}, pk_{\text{sig}})
\]

Verify. Algorithm Verify outputs a decision \( d \in \{\text{true, false}\} \) verifying the correctness of a signature \( \sigma \) for a message \( m \) w.r.t. a public key \( pk_{\text{sig}} \):
\[
d \leftarrow \text{Verify}(m, \sigma, pk_{\text{sig}})
\]

B. The Past

The idea of redactable signature schemes was introduced by Steinfeld et al. [121], and, with a slightly different focus, by Johnson et al. [84], even though some prior work on authenticated data-structures was already present [73].

1) Standard Security Properties: The first complete approach to formalize certain security properties was given by Brzuska et al. [30], which, however, focus on trees instead of lists. In particular, Brzuska et al. formalized the following properties, which we only sketch informally as it is sufficient for the purpose of this paper.

- **Unforgeability.** If one does not hold any secret keys, one should only be able to derive signatures for messages which are explicitly derivable by redaction, and no other ones.
- **Privacy.** An outsider not holding any private keys should not be able to derive any information about redacted parts of a message.
- **Transparency.** Transparency is even stronger than privacy. Namely, an outsider not holding any private keys should not be able to decide whether a signature has been created using the original signing algorithm or through the redaction algorithm. There are also some weaker definitions, which allow transparency for a certain amount of redacted blocks [76].

Clearly, these properties are akin to the definitions for SSSs. However, as RSSs allow for public redactions, i.e., there is no other entity holding any secret keys, there is no need to define security against another insider, simplifying the security model significantly.

C. The Present

As for SSSs, RSSs have then been extended for additional possibilities, and security. A generalized framework has later been given by Derler et al. [57], which covers most of the following additional properties from existing constructions, e.g., [43], [83], [10], [81], [82]

1) Additional Security Properties: The above security properties were later extended to cover more use-cases. In particular, the initial properties were extended for:

- **Unlinkability.** As for SSSs, unlinkability of RSSs requires that a signature does not leak from which signature it was derived [37]. This is clearly a very strong privacy notion, and not many constructions achieve this notion [1], [7], [8], especially if their target is slightly different.
- **Updatable and Mergeable Redactable Signatures.** RSSs, as their very name suggests, only allow to remove blocks. This, however, was extended to allow updating signature by adding new elements [96], but also to merge signatures derived from the same origin [111]. This is essentially a mixture between RSSs and append-only signature schemes [86], borrowing ideas from “incremental cryptography” [18].
- **Adding Zero-Knowledge.** Some of the ideas have also been extended into the zero-knowledge realm, allowing for order-queries, closest-neighbor searches, yet also zero-knowledge sets [67], [69], [68], [101], and generalizations of the underlying primitives [44].
- **Disclosure Control.** The first definitions allowed that every block of a message is redactable. However, there are application scenarios where either the signer, or even some other intermediate party, wants to prohibit that certain blocks can be redacted [75], [98], [105], [104], [103], [117].
- **Accountable RSS.** Even though RSSs are very useful primitives, their original definition does not consider accountability. This situation has been tackled by Pöhs and Samelin which show how to use SSSs to make RSSs accountable [112]. A related direction are signer-anonymous RSSs, where a verifier cannot decide which signing key was used [56], but can be traced by a trusted by party.
- **Complex Data-Structures and Generalizations.** The idea of RSSs has also been extended to cover more complex data-structures than sets [106] and lists, such as trees, graphs, and forests [12], [61], [78], [92], [93], [90], [91], [114], [113], [51], and related ideas, e.g., dependencies between blocks [119].
2) Implementations: As for SSAs, a lot of work was put into proving that RSSs are practical. Namely, implementations have been presented for smart cards [110], small sensor-nodes [64], desktop PCs [91] (and some later follow-ups by Kundu et al.) to powerful servers used for hospitals [124]. These implementations show that RSSs are practical.

D. The Future

1) Legal Implications and Standardization: As for SSAs, RSSs clearly help protecting data, i.e., may help to fulfill existing data-protection laws [120]. However, also RSSs have not been standardized yet. Therefore, the same arguments and wishes as for SSAs apply: Standardization would provide a common baseline for practitioners and cryptographers.

2) Social Implications and Human Computer Interaction (HCI): Even more for RSSs than for SSAs their actions are often explained with a physical metaphor of redacting using black ink. Clearly this seems helpful to convey the meaning, however as we have seen digital redaction going wrong a powerful tools of RSS needs to be solidly understood. First research in this direction is under way [3], [2] and the results shall be carefully considered when moving ahead to not loose the link to the human users.

3) New Application Scenarios:

- Redactable Commercials for Ad-Blocking in Authentic Content. There is already the concept of redactable blockchains [6]. Following their ideas, it may be reasonable to use RSSs to redact certain commercials from a website at a client, e.g., an employee at a company which disallows certain commercials of questionable nature. This may even be paired with a special redaction key, which is only given to the company, if it pays for it. This would allow to block advertisements by redacting them from authenticity protected content.

- Content Distribution Networks. Let \( S \) be a set. Assume that this set is replicated across multiple servers, e.g., in the cloud. Further assume that these servers form a CDN. Mergeable RSSs [111] allow a client to request different authenticated subsets \( S_i \subset S \) from different servers to increase download speed but verify all individual downloads as authentic, and then later merge them before further distributing a larger blob of authenticated data.

4) Future Research Opportunities and Questions:

- Combining RSS and SSAs. As already argued, SSAs allow to alter (not redact) admissible blocks, while RSSs allow only complete removal of blocks. It is therefore a natural question to ask, if both primitives can be combined, i.e., if complete removal and sanitization are possible at the same time, maybe even with some of the extensions discussed earlier.

- Multi-Signer/Sanitizer RSSs. The notion of accountable RSSs currently only supports one signer and one sanitizer [112]. We therefore ask, if it is possible to combine the ideas given in [40], [56], [112] to allow a scheme which has multiple signers, and multiple sanitizers, in an accountable way.

- Functional RSSs. Assume that a message \( m \) consists of three parts. Further assume that the signer wants that either \( m^{[3]} \) (or \( m^{[1]} \)) can be redacted, while if \( m^{[2]} \) is removed, also \( m^{[3]} \) must be redacted. This may, e.g., be useful if patient data is redacted. This idea is not new per se [119], but it remains the question whether there exists transparent (or even unlinkable) constructions, and if more expressive and useful “functions” exists.

V. Conclusion

In this position paper, we gave an overview of sanitizable signature schemes and redactable signature schemes. We kept it at an introductory-level, still providing a clear separation between the two concepts and achieving a broad coverage of the current body of literature. We compared those primitives to other related schemes, and discussed their current state-of-the-art. Based on arising application scenarios, we identified additional research opportunities, but also stated what, in our opinion, is missing to bring both primitives into practice.

VI. Acknowledgments

H. C. Pöhls was supported by EU H2020 project PRISMA CLOUD: This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement n\textsuperscript{\underline{6}}44962. K. Samelin was supported by the EU ERC PERCY, grant agreement n\textsuperscript{\underline{3}}2131.

We also thank Daniel Slamanig for some interesting discussions regarding this paper.

REFERENCES


