k-ANONYMITY IN CONTEXT OF DIGITALLY SIGNED CDA DOCUMENTS

Daniel Slamanig and Christian Stingl
Department of Medical Information Technology, Healthcare IT & Information Security Group
Carinthia University of Applied Sciences, Primoschgasse 10, 9020 Klagenfurt, Austria

Keywords: Clinical document architecture, Secondary use, *k*-anonymity, Digital signatures, Generalized redactable signatures, Privacy, Anonymization.

Abstract: If medical data are provided to third parties for secondary use, the protection of the patients privacy is an essential issue. In general this is accomplished by removing identifying and quasi-identifying information to provide *k*-anonymity for a given data set. This means, that one patient cannot be distinguished from at least $k-1$ other individuals. However, if the single records of the data set are digitally signed, the modification of the respective records destroys their integrity as well as their authenticity. Hence, digital signatures, which are an invaluable tool for verifying the integrity and authenticity of digital medical data, seem to be inadequate in this scenario. But, especially in context of secondary use, malicious manipulations and processing errors may lead to serious failures in a subsequent medical (treatment) process.

In this paper we propose a novel approach based on generalized redactable signatures that realizes *k*-anonymity for sets of digitally signed records. To the best of our knowledge this is the first work that combines these seemingly contradictory topics very efficiently. In particular, the proposed solution allows any party to verify the original digital signatures for medical data, although these data are modified during the process of achieving *k*-anonymity. The main advantage of this approach is that all parties involved in the aforementioned process are able to verify the integrity and authenticity based on the original digital signatures.

1 INTRODUCTION

One major drawback in context of digital signatures is, that the validity of a signature can solely be verified if the entire document is available. Usually, this is desirable and reasonable, but in some scenarios this aspect of digital signatures is counterproductive. A major aspect, that is in our opinion highly interesting, is that only a specific part of a medical document need to be given to other parties, e.g. de-identified medical data for a second opinion. Another example, that will be discussed below is a patient who removes results resp. diagnoses from a medical report to obtain an independent second opinion from another expert. But, in this scenario the receiving party is not able to verify the integrity and authenticity of these data anymore. Clearly, one possibility to overcome this problem is to contact the original signer to provide another digital signature for the modified medical report. However, in the majority of treatment processes this is absolutely impractical. Consequently, this means that the receiving party needs to absolutely trust the received information. For instance, errors that occur during the removal of parts or during the transmission or even result from malicious manipulation cannot be identified. This may lead to serious failures in the subsequent process. Thereby, it is essential that the patient can only remove information of a medical document that were specified by the original signer.

Subsequently, we will provide an overview of examples which can benefit from the method introduced in this paper. It should be noted, that this work focuses on Extensible Markup Language (XML) data and in particular the Clinical Document Architecture (CDA), which is in our opinion the most promising standard for the exchange of clinical documents. Furthermore, digital signatures are explicitly considered in this architecture.

- Verifiable disclosure of parts of a CDA document, e.g. for second opinions.
- Verification of anonymized CDA documents, e.g.

Slamanig D. and Stingl C. (2010).
k-ANONYMITY IN CONTEXT OF DIGITALLY SIGNED CDA DOCUMENTS.
In *Proceedings of the Third International Conference on Health Informatics*, pages 62-69
DOI: 10.5220/0002731700620069

for clinical studies.

- Signatures on partial information of CDA documents.
- Verification of medical data that is stored in an anonymized (pseudonymized) fashion in electronic or personal health record architectures (Huda et al., 2008; Riedl et al., 2008).
- Verifying signatures on CDA documents without having access to XML Schema files and/or Extensible Stylesheet Language (XSL) information for the layout of the content.

In these examples conventional digital signatures cannot be used to provide the integrity and authenticity of the modified parts. However, in our opinion these two aspects are essential when dealing with highly sensitive data. When considering a set of documents that is given to another party, e.g. for a clinical study, instead of a single document, the situation turns to be much more complex. Then, the entire set needs to be taken into consideration to protect the privacy of the individuals, i.e. to prevent unique identification of patients. This can be accomplished by means of techniques that are based on *k*-anonymity (Li et al., 2007; Machanavajjhala et al., 2007; Samarati, 2001; Sweeney, 2002). Roughly spoken, this means that relevant attributes of all documents are modified to such an extent that one patient cannot be distinguished from at least $k-1$ other individuals. It is seemingly a paradox, that a digital signature stays valid although the corresponding document is modified. However, in this paper we will show that the main methods that are applied to obtain *k*-anonymity do not negatively influence the original signature, when using generalized redactable signatures.

2 CDA

The HL7 Clinical Document Architecture (CDA) (Dolin et al., 2001) is a document markup standard that specifies the structure and semantics of “clinical documents” for the purpose of exchange. CDA Release 2.0 became an ANSI-approved HL7 standard in May 2005. The content of CDA documents are derived from the HL7 Reference Information Model (RIM) and are encoded in XML.

A CDA document contains a header, which provides for instance information on the patient, the author of the document, the custodian, the recipient and the authenticator of the document. Additionally, it is possible to provide a participant who has legally authenticated the document, e.g. digitally signed the document. The body of a CDA document contains the clinical report and can either be unstructured (level 1), semi-structured (level 2) or highly structured (level 3). The structuring of a CDA document can be implemented by using so called CDA body entries, e.g. observation, procedure, encounter, substance administration, supply, observation media, etc. The basic principles of CDA are briefly discussed subsequently.

- **Persistence:** A clinical document continues to exist in an unaltered state for a time period, defined by local and regulatory requirements.
- **Stewardship:** A clinical document is maintained by an organization entrusted with its care.
- **Potential for Authentication:** A clinical document is an assemblage of information that is intended to be legally authenticated.
- **Context:** A clinical document establishes the default context for its contents.
- **Wholeness:** Authentication of a clinical document applies to the whole and does not apply to portions of the document without the full context of the document.
- **Human Readability:** A clinical document is human readable.

One important principle and a necessity regarding legal requirements is the (legal) authentication of documents. This aspect is one of the motivating factors of this work, i.e. the (legal) authenticator signs the document such that every receiving party is able to check the integrity and authenticity of the document. Furthermore, the originator is not able to repudiate that the document was created by him.

3 DE-IDENTIFYING HEALTH DATA

Person related health data are in general very sensitive information and consequently must be protected appropriately. Especially, when using these data for secondary use, which includes medical research, clinical studies and second opinions, the protection of the privacy of patients is obligatory. Hence, it is necessary to prepare data before passing it to another party for secondary use, such that the patients cannot be identified uniquely anymore. This process of preparation focuses primarily on attributes that directly identify the patients and secondarily on attributes that indirectly identify the patients. The latter class of attributes is often denoted as quasi-identifying information. CDA documents contain directly identifying attributes, e.g. the social security number, as well as several indirectly identifying attributes, e.g. name, gender, date of birth, geographic information, demographic data, etc. Furthermore, there are also document related attributes, e.g. unique document identifier, which can be used in combination with external sources to identify the respective patient. Subsequently, we will discuss two different approaches, which focus on a single CDA document and a set of CDA documents respectively.

3.1 Anonymization

In general, anonymization means to remove all patient related information from medical data. This includes all directly identifying attributes as well as a subset of the indirectly identifying attributes, e.g. the surname, but probably not the date of birth. For instance, there are heuristics like the Safe Harbor Rule, which is part of the Health Insurance Portability and Accountability Act (HIPPA) that provides among others a precise set of 18 specific categories of data that need to be removed (Emam, 2008).

3.2 *k*-Anonymity

When passing a set of data that contains person related health data to another party for secondary use, e.g. a clinical study, then it is necessary that the individuals contained in this set cannot be uniquely identified. However, the naive approach of removing solely the directly identifying attributes is unfortunately not sufficient. This is also reflected in a study (Sweeney, 2000), which estimates that 87% of the population of the United States can be uniquely identified based only on the seemingly harmless attributes gender, date of birth, and 5-digit zip code. Consequently, data that contains these attributes cannot be considered as anonymous. The concept of *k*-anonymity, introduced in (Samarati, 2001; Sweeney, 2002) provides methods to anonymize data sets by also taking into account indirectly identifying attributes and thus prevent unique identification of individuals. More precisely, a data set provides *k*-anonymity protection, if each person contained in the data set cannot be distinguished from at least $k-1$ individuals whose information also appear in this data set. Recent results showed, that *k*-anonymity may not be sufficient in specific scenarios and there have been proposed enhancements of this approach which are denoted as *l*-diversity (Machanavajjhala et al., 2007) and more recently *t*-closeness (Li et al., 2007). Thereby, the latter two concepts are based on *k*-anonymity.

3.3 Methods

In order to establish anonymization for a single document or *k*-anonymity for a set of documents, the subsequent methods have been proposed (Emam, 2008; Samarati, 2001; Samarati and Sweeney, 1998; Sweeney, 2002).

- **Record Suppression:** An entire record is removed from the data set.
- **Attribute Suppression:** One ore more attributes are removed from records of the data set.
- **Generalization:** This can be accomplished by attribute suppression, or by removing or replacing certain parts of attributes. For instance, the geographic information comprises city and region. The suppression of the city will result in a geographic area aggregation. An example for the second case based on the attribute `birthTime value="19970924"` (see figure 3) is the removal of day and month, which results in `birthTime value="1997****"`. Another example using the same attribute is a replacement of the last digit of the year, which results in `birthTime value="199 5 0924"`. The last example can be seen as grouping individuals after their year of birth in intervals of 5 years.
- **Sampling, Swapping Attributes and adding Noise:** These operations can be applied to the data set, while maintaining some overall statistical properties of the data set (Bakken et al., 2004; Ciriani et al., 2007).

The methods “record suppression”, “attribute suppression” and “generalization” are mainly used in context of anonymization and *k*-anonymity. Hence, in this paper we are focusing on these methods and propose techniques to apply them to sets of digitally signed medical documents.

4 METHODS

As mentioned in the introduction, the main target of our approach is that the integrity as well as the authenticity of the data need to be guaranteed in every single step of the process (see figure 1). More precisely, the initial digital signature created in the first step of the process need to be usable in every subsequent step to verify the integrity and determine the authenticity of every document. Consequently, this leads to the paradox situation that, although documents are modified during the preparation step, the original signatures need to stay valid. The process illustrated in figure 1 can be divided into three steps:

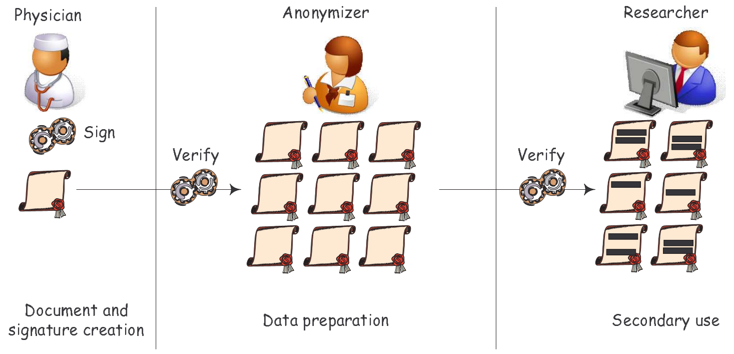

Figure 1: The process of achieving k-anonymity for a set of documents.

1. **Document and Signature Creation:** A CDA document is created and digitally signed by a (legal) authenticator.
2. **Data Preparation:** This is the process of anonymizing a single CDA document or to achieve k-anonymity for a set of CDA documents.
3. **Secondary Use:** Based on the modified documents a third party conducts the secondary use, e.g. a clinical study.

In the following, we describe properties which are used to evaluate three available variants of digital signatures discussed in section 4.2, with respect to their applicability in context of anonymization. Thereby, it should be noted that record suppression is only considered for the sake of completeness, since omitting entire documents is always possible during the preparation step. The first four properties are directly taken from section 3 and the property “controlled removal” means that the original signer is able to control which information can be removed during data preparation.

- Record suppression
- Attribute suppression
- Generalization by aggregation
- Generalization by removal of parts
- Generalization by replacement of parts
- Controlled removal
- Anonymization
- k-anonymity

Before we are going into details, we provide a brief introduction to digital signatures based on the hash-then-sign paradigm, with a special emphasis on the construction of the hash value of the message prior to signing.

4.1 Basic Principles of Digital Signatures

Digital Signatures are a widely used cryptographic method to provide authenticity, integrity and non-repudiation of electronic documents. In order to be able to sign documents, a user generates a secret signing key SK and a corresponding public verification key PK, which is certified together with the user’s identity by a so called certification authority (CA). To efficiently generate a digital signature for a document D, the user computes the hash value of the document $h = H(D)$ by means of a publicly known collision resistant cryptographic hash function, e.g. the SHA-2 family, and computes the signature σ by means of the signing algorithm S and his secret key SK, i.e. $\sigma = \mathsf{S}_{SK}(h)$. Every person who is in possession of (PK, D, σ) is able to use the verification algorithm V to check whether the signature σ represents a valid signature for the document D and was generated by the owner of PK, i.e. $\mathsf{V}_{PK}(H(D), \sigma) \in \{accept, reject\}$. An alternative to compute the hash value $h = H(D)$ is to build a Merkle-tree (hash tree) (Merkle, 1989) of the document D and finally sign the hash value of the root node of the Merkle-tree. This construction is very useful in context of redactable signature schemes, which are discussed in section 4.2.3, and also long-term archiving of documents.

4.2 Variants of Digital Signatures

In this section we compare three variants of digital signatures regarding their applicability in context of anonymization and k-anonymity.

4.2.1 XML Digital Signatures

The XML-DSig recommendation of the W3C (Eastlake et al., 2002) defines an XML syntax and processing rules for creating and representing digital signatures. It can be conveniently used to sign arbitrary XML or non-XML data using one of the following signing types:

- Enveloped: The XML signature is included in the XML document. It is contained within a child element of the XML document.
- Enveloping: The XML document is included in the XML signature. It is contained within a a child element of the XML signature.
- Detached: The XML signature is included in a separate document from the signed document. The location of the signed document is referenced in the XML signature. This type of signature is used for non-XML documents.

It is important to note, that when signing XML documents, one needs to serialize it prior to signing. Thereby, this process needs to guarantee that logically-identical documents produce exactly identical serialized representations, which do not depend on the actual encoding, white spaces, etc. This process is usually denoted as normalization (canonicalization) and one representative of canonicalization methods is $C14N$.

Since every modification of the document invalidates the digital signature, none of the above mentioned properties (except record suppression) can be achieved.

4.2.2 Partial Signatures based on XML-DSig

A partial signature is a signature on an arbitrary subdocument of an XML document. This means, that using this approach it is possible to append several independent signatures for subdocuments to the XML document. Although attribute suppression and generalization cannot be applied to partially signed documents without invalidating the original signature, the original signer is able to produce additional partial signatures for specific subdocuments that do not contain identifying attributes. This could be used in scenarios which require the anonymization medical documents. However, k-anonymity can practically not be achieved, since the parts that need to be removed from the document during the preparation depend on the actual set of CDA documents for a specific secondary use. Clearly, the attributes that need to be removed must have been known to the original signer at the time of creating partial signatures, which is usually not the case in practice.

One can conclude that the above two variants of digital signatures cannot be reasonably used to accomplish the before mentioned properties. The third variant, which represents a generalization of redactable signatures can surprisingly be used to realize nearly all of the above mentioned properties. But before we are going into details, we provide a brief introduction to redactable signatures.

4.2.3 Generalized Redactable Signatures

The concept of a redactable signature scheme was introduced in (Johnson et al., 2002) and allows any party to remove parts of a signed document D to obtain a redacted document D' such that a signature for D' can be derived from the signature of D without cooperation with the original signer. Consequently, it is possible to remove certain parts of a document and pass the remaining document to another party, who is able to verify the integrity and authenticity of the resulting document D'. It must be noted, that several variants of signature schemes realizing comparable ideas have been proposed (Ateniese et al., 2005; Miyazaki et al., 2006; Steinfeld et al., 2001).

Redactable signatures (Johnson et al., 2002) organize the content of a document as leafs of a complete binary tree. This is absolutely sufficient for unstructured documents. However, when using a structured document like an XML document, which itself represents a tree, then splitting up a document into blocks of fixed size and organizing these blocks in a binary tree is not desirable. The redactable signature of (Johnson et al., 2002) also works with variable block length, however, the rule for splitting up the document needs to be available to the redactor (anonymizer) as well as the verifier and consequently must be appended to the document. Thus, it is more natural to use the existing tree structure of a structured document and thereby use inner nodes as well as leaf nodes to hold parts of the document, instead of organizing the parts as leafs of a binary tree. This approach is denoted as generalized redactable signatures (Slamanig and Stingl, 2009). Subsequently, we will present additional transformation rules for the generalized redactable signature proposed in (Slamanig and Stingl, 2009).

When representing an XML document with its inherent tree structure, the resulting tree is in general neither binary nor complete. In the following we will define a unique transformation $\mathcal{T}$ which maps an arbitrary XML document uniquely to a N-ary tree. But, we want to emphasize that there exist other mappings which can also be used for this purpose. For the sake of simplicity of the presentation we are focusing on XML elements, attributes, attribute- and element-data and will present the transformation rules informally:

$\mathcal{R}_1$**:** Element `<TAG>VALUE</TAG>` : The label of the

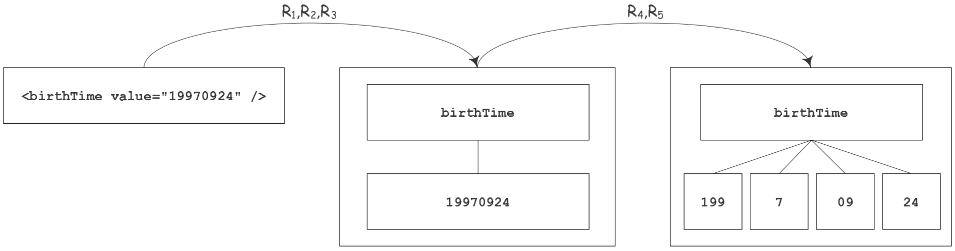

Figure 2: Applying the rules to a single XML element.

root node is TAG and has one child node with label VALUE .

$\mathcal{R}_2$: Element <TAG A $_1$=V$_1$,...,A $_n$=V$_n$></TAG> : Again, the label of the root node is TAG and for each attribute i a sub-tree with root labeled A_i and one child labeled V_i is constructed.

$\mathcal{R}_3$: Element <TAG><STAG $_1$></STAG $_1$>...<STAG $_n$> </STAG $_n$></TAG> : The label of the root node is TAG and for each subtag a sub-tree with root labeled STAG $_i$ and child node representing the value is constructed.

```
<recordTarget>
  <patientRole>
    <id extension="12345" root="2.16.840.1.113883.3.933" />
    <patientPatient>
      <name>
        <given>Henry</given>
        <family>Levin</family>
        <suffix>the 7th</suffix>
      </name>
      <administrativeGenderCode code="M"
                              codeSystem="2.16.840.1.113883.5.1" />
      <birthTime value="19970924" />
    </patientPatient>
  </patientRole>
</recordTarget>
```

Figure 3: A fragment of the header of a CDA document.

It should be noted, that in all above transformation rules the attributes are transferred prior to values. Since an arbitrary XML document cannot be usefully transformed into a binary tree, the standard Merkle-tree cannot be applied in this scenario. Considering this transformation, it must be noted that in addition to leaf nodes also the inner nodes contain valuable information, e.g. names of elements and attributes. Consequently, the Merkle-tree construction has to be adapted to fit these needs. Therefore, an adapted assignment ϕ for the label of the nodes is proposed in (Slamanig and Stingl, 2009), which is recursively defined in (1). Thereby, x is the value of the parent node p and $c_0,\ldots,c_k$ are the respective child nodes, where $0 \leq k < N$.

$$\phi(p) = \begin{cases} H(x||\phi(c_0)||\ldots||\phi(c_k)) & \text{if } p \text{ has } k+1 \text{ children,} \\ H(x) & \text{if } p \text{ is a leaf.} \end{cases} \tag{1}$$

Due to the construction of the Merkle-tree, it is possible to reconstruct certain (small) redacted parts of the document by brute-force attacks. Hence, it is necessary to randomize the Merkle-tree construction by applying a so called GGM-tree (Goldreich et al., 1986) resp. a modified GGM-tree (Slamanig and Stingl, 2009). The key issue of this construction is, that the original signer is able to determine the granularity of redactable information and any other party is able to remove these parts and can “adapt” the original signature such that the signature is valid for the redacted document. The details of the two above mentioned aspects are omitted here, since they do not influence the concept proposed in this paper and details can be found in (Johnson et al., 2002; Slamanig and Stingl, 2009).

When applying the transformation $\mathcal{T}$ and subsequently signing the document using the generalized redactable signature scheme, the properties “attribute suppression”, “generalization by aggregation” can be realized completely and “controlled removal” partially. However, “generalization by removal of parts” cannot be achieved at all. Subsequently, we will propose an extended transformation $\mathcal{T}'$. This transformation uses two additional rules, which are introduced below.

$\mathcal{R}_4$: Element <TAG>VALUE</TAG> : The label of the root node is TAG and has a child nodes with label VALUE $_1,\ldots,$VALUE $_n$ where these new values are a unique and complete representation of the original VALUE .

$\mathcal{R}_5$: Element <TAG A $_1$=V$_1$,...,A $_n$=V$_n$></TAG> : The label of the root node is TAG and for each attribute i a sub-tree with root labeled A_i and a child nodes labeled $V_{i_1},\ldots,V_{i_n}$ is constructed as in $\mathcal{R}_4$.

When applying transformation $\mathcal{T}$ or $\mathcal{T}'$, the result does not represent a valid XML-document, which is however not necessary for the generalized redactable

signature scheme. It should be noted, that the transformation could easily be adapted by adding additional tags such that the result represents a valid XML document.
Based on an example, we will now demonstrate the potential as well as the limitations of the rules $\mathcal{R}_4$ and $\mathcal{R}_5$ (see figure 2). Obviously, it is possible to remove certain parts of the `birthTime` , e.g. the day and/or the month, which realizes exactly the "generalization by removal of parts". However, the property "generalization by replacement of parts" cannot be achieved, since this concept somewhat contradicts the idea behind generalized redactable signatures.

Efficiency Analysis. The costs of a single generalized redactable signature consists of exactly n calls to the hash function H, the computation of the (modified) GGM tree and one digital signature, where n is the number of elements, attributes and values or the corresponding split values of the transformed XML document.

The results of this section are summarized in table 1. As a consequence, in the subsequent section we solely consider generalized redactable signatures to achieve k-anonymity, since it is the only approach that is applicable.

5 WORKFLOW TO ACHIEVE k-ANONYMITY

In this section we will discuss the workflow as well as the parties introduced in section 4 more detailed. Recall, these parties are the creator and original signer, the anonymizer and the party conducting the secondary use. For the sake of simplicity, we are omitting details on the normalization (canonicalization) and the encoding of the XML document prior to signing. This is clearly essential, but does not influence the concept.

5.1 Creator and Original Signer

The creator and original signer proceed as follows:

1. Definition of the granularity of the redaction and splitting of values according to a common agreed policy, which defines the values that are allowed to split and the rules to split.
2. Applying transformation $\mathcal{T}'$ to the CDA document.
3. Computation of the hash value of the root node by means of the adapted Merkle-tree.
4. Signing of the hash value using the private key of a conventional digital signature scheme, e.g. RSA-PSS, ECDSA, etc.

The resulting signature can be verified by any party who is in possession of the document and the signers public key.

5.2 Anonymizer

The anonymizer proceeds as follows, when given a set of signed CDA documents.

1. According to the specification of the secondary use, the anonymizer identifies all documents that are relevant for the specific clinical study.
2. Removal of all directly identifying attributes.
3. Identification of the set of indirectly identifying attributes (quasi-identifiers).
4. Achieving k-anonymity by applying the methods of section 3.3 to the indirectly identifying attributes, by taking into account the common agreed policy.
5. Adapting the digital signature of each redacted document

5.3 Party Conducting the Secondary Use

1. After receiving the set of redacted documents, the integrity and authenticity is verified by means of the (adapted) digital signatures. Thereby, the public keys of the original signers are used.
2. If the signature verification succeeds, the set of redacted documents can be used for the intended clinical study.

6 CONCLUSIONS & FUTURE ASPECTS

In this paper we have introduced a novel approach that covers the concept of k-anonymity in context of digitally signed medical documents. To the best of our knowledge this is the first work that combines these seemingly contradictory topics. As we have shown, generalized redactable signatures provide an efficient and practical solution to realize nearly all methods to achieve k-anonymity. With respect to the properties illustrated in table 1, generalized redactable signatures solely fail to support "generalization by replacement of parts", whereas there is no way to realize this property solely by redacting information.

Table 1: Overview of properties provided by the discussed variants of digital signatures (≈ means partial support).

Property	XML-DSig	Partial Sig	Generalized Redactable Sig
Record suppression	✓	✓	✓
Attribute suppression	×	×	✓
Generalization by aggregation	×	×	✓
Generalization by removal of parts	×	×	✓
Generalization by replacement of parts	×	×	×
Controlled removal	×	×	✓
Anonymization	×	≈	✓
k-anonymity	×	×	✓

Nevertheless, one future research direction is to use so called *Bloom* filters (Bloom, 1970) to efficiently "store" the set of possible replacements to realize the open method "generalization by replacement of parts".

REFERENCES

Ateniese, G., Chou, D., de Medeiros, B., and Tsudik, G. (2005). Sanitizable Signatures. In *ESORICS 2005*, volume 3679 of *LNCS*, pages 159–177. Springer.

Bakken, D. E., Parameswaran, R., Blough, D. M., Franz, A. A., and Palmer, T. J. (2004). Data Obfuscation: Anonymity and Desensitization of Usable Data Sets. *IEEE Security and Privacy*, 2(6):34–41.

Bloom, B. H. (1970). Space/Time Trade-offs in Hash Coding with Allowable Errors. *Commun. ACM*, 13(7):422–426.

Ciriani, V., di Vimercati, S. D. C., Foresti, S., and Samarati, P. (2007). *k*-Anonymity. In *Secure Data Management in Decentralized Systems*, pages 323–353. Springer.

Dolin, R., Alschuler, L., and et al., C. B. (2001). The HL7 Clinical Document Architecture. *J. Am. Med. Inform. Assoc*, 6:552–569.

Eastlake, D., Reagle, J., and Solo, D. (2002). XML-Signature syntax and processing. http://www.w3.org/TR/xmldsig-core/.

Emam, K. E. (2008). Heuristics for De-identifying Health Data. *IEEE Security & Privacy*, 6(4):58–61.

Goldreich, O., Goldwasser, S., and Micali, S. (1986). How to Construct Random Functions. *J. ACM*, 33(4):792–807.

Huda, N., Sonehara, N., and Yamada, S. (2008). A Privacy Management Architecture for Patient-Controlled Personal Health Record System. In *NetApps 2008*. IEEE Computer Society.

Johnson, R., Molnar, D., Song, D., and Wagner, D. (2002). Homomorphic Signature Schemes. In *CT-RSA '02*, volume 2271 of *LNCS*, pages 244–262. Springer.

Li, N., Li, T., and Venkatasubramanian, S. (2007). *t*-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *ICDE 2007*, pages 106–115. IEEE Computer Society.

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). *l*-Diversity: Privacy beyond *k*-Anonymity. *ACM Transactions on Knowledge Discovery from Data*, 1(1).

Merkle, R. (1989). A Certified Digital Signature. In *CRYPTO '89*, volume 435 of *LNCS*, pages 218–238. Springer.

Miyazaki, K., Hanaoka, G., and Imai, H. (2006). Digitally Signed Document Sanitizing Scheme Based on Bilinear Maps. In *ASIACCS 2006*, pages 343–354. ACM.

Riedl, B., Grascher, V., and Neubauer, T. (2008). A Secure e-Health Architecture based on the Appliance of Pseudonymization. *Journal of Software*, 3(2):23–32.

Samarati, P. (2001). Protecting Respondents' Identities in Microdata Release. *IEEE Trans. Knowl. Data Eng.*, 13(6):1010–1027.

Samarati, P. and Sweeney, L. (1998). Generalizing Data to Provide Anonymity when Disclosing Information (Abstract). In *PODS' 98*, page 188. ACM Press.

Slamanig, D. and Stingl, C. (2009). Disclosing Verifiable Partial Information of Signed CDA Documents using Generalized Redactable Signatures. In *IEEE Healthcom 2009*. IEEE Communications Society.

Steinfeld, R., Bull, L., and Zheng, Y. (2001). Content Extraction Signatures. In *ICISC 2001*, volume 2288 of *LNCS*, pages 285–304. Springer.

Sweeney, L. (2000). Uniqueness of simple demographics in the u.s. population. Technical report, Carnegie Mellon University.

Sweeney, L. (2002). *k*-Anonymity: a Model for Protecting Privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570.