COMPUTERS & SECURITY 50 (2015) 16–32

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Integrity, authenticity, non-repudiation, and proof of existence for long-term archiving: A survey

Martín Vigil [a,*], *Johannes Buchmann* [a], *Daniel Cabarcas* [b], *Christian Weinert* [a], *Alexander Wiesmaier* [c]

[a] *Technische Universität Darmstadt, Hochschulstraße 10, 64289 Darmstadt, Germany*
[b] *Universidad Nacional de Colombia, Sede Medellín, Calle 59A No 63 - 20, Medellín, Colombia*
[c] *AGT Group (R&D) GmbH, Hilpertstraße 35, 64295 Darmstadt, Germany*

ARTICLE INFO

Article history:
Received 14 March 2014
Received in revised form
13 November 2014
Accepted 19 December 2014
Available online 27 December 2014

Keywords:
Long-term
Archiving
Authenticity
Non-repudiation
Integrity
Existence
Time-stamp
Notarization
Replication
Trust

ABSTRACT

The world increasingly depends on archives to store digital documents, such as land registers and medical records, for long periods of time. For stored documents to remain trustworthy, archives must provide proofs that a document existed on a certain date and has not been changed since. In addition, in many cases, the origin of the document must be verifiable and the originator must not be able to repudiate that she is the originator. In this paper, we survey the solutions that provide the above protection goals in the long term. We analyze and compare the solutions with respect to their functionalities (which protection goals do they achieve?), the trust assumptions they require, and their performance. From this analysis and comparison, we deduce deficiencies of the current solutions and important research problems that must be solved in order to come up with protection solutions that are even more satisfactory.

1. Introduction

Electronic archives are increasingly necessary to store information that needs to be available for many years or even indefinitely. For example, in 2014 the Estonian land register contained digital records of about one million Estonian properties and these records are kept indefinitely. Another example is the Irish Tax and Custom, which received 6.7 millions electronic tax returns only in 2013 (R. I. Tax, 2014) and stores all received tax returns indefinitely. In many countries, the health sector is also required by law to store patients' medical records for many years. For instance, in the United Kingdom (UK) retention may be mandatory for 20 years (DH/Digital Information Policy, 2009). Yet another important example are patent offices, which preserve records for long

* *Corresponding author.*
E-mail addresses: vigil@cdc.informatik.tu-darmstadt.de (M. Vigil), buchmann@cdc.informatik.tu-darmstadt.de (J. Buchmann), dcabarc@unal.edu.co (D. Cabarcas), christian.weinert@stud.tu-darmstadt.de (C. Weinert), awiesmaier@agtgermany.com (A. Wiesmaier).
http://dx.doi.org/10.1016/j.cose.2014.12.004
0167-4048/

periods, such as 27 years in the UK (Intellectual Property Office, 2012).

Long-term digital archives must guarantee *long-term protection* of their contents (i.e. documents) which means that the documents must be protected as long as they are in the archives. This time period may range from a few decades to many generations, for example in the case of land registers. To cover long-term protection requirements, techniques are needed that provide everlasting protection. We will in this paper therefore use the terms long-term and everlasting synonymously.

Important protection goals are *integrity*, *authenticity*, *proof of existence*, *non-repudiation*, and *confidentiality* of archived documents. Integrity means that a document has not been altered. Authenticity means that its origin can be identified. Proof of existence allows to identify a time reference when the document existed. Non-repudiation prevents an originator from repudiating that he is the origin of a document. These protection goals are closely related. Integrity and proof of existence are the basis. Stating that a document has not been changed includes a statement since when this is true. Thus, an additional proof of existence is desirable. Conversely, a proof of existence makes no sense without a proof of integrity. Also, authenticity and non-repudiation make no sense if there is no proof of integrity. Confidentiality refers to the protection of documents from unauthorized access.

Coming up with solutions that provide everlasting protection is a challenging task. For example, there are solutions that use digital signatures to achieve the protection goals. However, digital signatures become insecure when their security properties are defeated by advances in computer power and cryptanalytic techniques. For example, today's attacks can defeat the security of 512-bit RSA signatures (Cavallar et al., 2000) which were considered secure in 1990. Therefore, single digital signatures cannot provide long-term protection.

In this paper, we survey, analyze, and compare the existing solutions that achieve proof of existence, integrity, authenticity, and non-repudiation of documents in archives. We do not cover confidentiality as the methods for achieving this protection goal are very different from the approaches to providing the other protection goals. An overview of long-term protection of confidentiality can be found in (Braun et al., 2014).

The first step of our analysis is to identify the common building blocks that the solutions use and to discuss their properties. This is done in Section 2. These building blocks are cryptographic, such as digital signatures, or non-cryptographic, for example widely visible media. As these building blocks are to be used in long-term protection schemes, we discuss their long-term security properties. Next, we present the solutions that achieve our protection goals in the long term. We distinguish between three types of solutions. First, we discuss time-stamping-based solutions in Section 3. They use a sequence of time-stamps (see Section 3.1), in which the first time-stamp provides a proof of existence and integrity of one or more documents and the subsequent time-stamps prove the validity of the previous time-stamp, thereby prolonging the validity of the first time-stamp. The surveyed schemes include *Advanced Electronic Signatures* (ETSI and 2010a; ETSI and 2010b), *Content Integrity Service* (Haber, 2006), and *Evidence Record Syntax* (Blazic et al., 2011; Gondrom et al., 2007). The second type of solutions use notarization. This is discussed in Section 4. They rely on notaries who issue attestations confirming the integrity, authenticity, non-repudiation, and proof of existence of documents or of previous attestations. There are only two such solutions, namely *Cumulative Notarizations* (Lekkas and Gritzalis, 2004) and *Attested Certificates* (Vigil et al., 2013). The third type, presented in Section 5, uses replication of documents and majority voting. To the best of our knowledge, this idea is only applied by the scheme *Lots of Copies Keep Stuff Safe* (Maniatis et al., 2005). All schemes are presented in some technical detail using a unified approach that allows to compare them.

An analysis and comparison of the schemes is presented in Section 6. We consider three important aspects. The first aspect is functionality: which scheme achieves which of our protection goals integrity, authenticity, proof of existence, and non-repudiation. We also discuss additional properties, for example, whether the solutions tolerate format changes of the archived documents. As a second aspect, we discuss the trust assumptions on which the schemes rely. For example, all schemes that use cryptography assume that there is no sudden break of cryptography: the used cryptographic algorithms remain secure until they are replaced. Likewise, notarization-based schemes assume that notaries are trustworthy. Knowledge of these trust assumptions is very important for the users of the solutions. They may not be convinced that the trust assumptions are justified and may therefore prefer to use another solution. The third comparison aspect is the performance of the various solutions. The paper provides an extensive experimental comparison of the schemes. This comparison refers to the times required for generating and verifying the evidence for the various protection goals, the size of this evidence, and the size of the data exchanged between an archivist and the required trusted third parties when the evidence is created.

As the task of protecting archived documents is so important there are several other schemes that provide such protection. However, they fail to guarantee long-term protection. In Section 7, we briefly present these schemes and explain why they do not provide long-term protection.

Finally, we discuss important open research problems in Section 8 and draw our conclusions in Section 9.

2. Building blocks

In this section, we describe the building blocks used by the schemes presented in the next sections. The basic components of the building blocks are hash functions and digital signatures, cryptographic primitives of limited lifetime. Therefore, we also present models to predict for how long such primitives can be used securely. Additionally, as the trustworthiness of the cryptographic proofs is further discussed, we present a model that approximates trustworthiness.

2.1. *Cryptographic hash functions*

Cryptographic hash functions are central building blocks in cryptography. For example, they are used to reduce integrity