Long-Term Archive Service Requirements

Status of This Memo

Copyright Notice

Abstract

   There are many scenarios in which users must be able to prove the
   existence of data at a specific point in time and be able to
   demonstrate the integrity of data since that time, even when the
   duration from time of existence to time of demonstration spans a
   large period of time. Additionally, users must be able to verify
   signatures on digitally signed data many years after the generation
   of the signature. This document describes a class of long-term
   archive services to support such scenarios and the technical
   requirements for interacting with such services.

Table of Contents

1. Introduction

   Digital data durability is undermined by continual progress and
   change on a number of fronts. The useful lifetime of data may exceed
   the life span of formats and mechanisms used to store the data. The
   lifetime of digitally signed data may exceed the validity periods of
   public-key certificates used to verify signatures or the
   cryptanalysis period of the cryptographic algorithms used to generate
   the signatures, i.e., the time after which an algorithm no longer
   provides the intended security properties. Technical and operational
   means are required to mitigate these issues. A solution must address
   issues such as storage media lifetime, disaster planning, advances in
   cryptanalysis or computational capabilities, changes in software
   technology, and legal issues.

   A long-term archive service aids in the preservation of data over
   long periods of time through a regimen of technical and procedural
   mechanisms designed to support claims regarding a data object. For
   example, it might periodically perform activities to preserve data
   integrity and the non-repudiability of data existence by a particular
   point in time or take actions to ensure the availability of data.
   Examples of periodic activities include refreshing time stamps or
   transferring data to a new storage medium.

   A long-term archive service may be used to provide evidence that
   supports validation of the existence of documents or assertions of
   agreements that were originally asserted with digital signatures.
   Validation may occur at times in the future well beyond the validity
   period of the private key originally used to generate the signature,
   or even beyond the time when the algorithms available for digital
   signatures, message digesting, or data encryption cease to offer
   effective protection because of improvements in computing speeds and
   methods.

   A long-term archive service may be located within an enterprise
   network, communicating with local storage mechanisms and other
   applications, or a long-term archive service may be implemented as an
   external service accessible via the Internet. A long-term archive
   service may use functionality, e.g., time stamping, provided by
   independent service providers.

   A primary goal of a long-term archive service is to support the
   credible assertion of a claim that is currently asserted, at points
   well into the future. A long-term archive service may support a
   range of applications, including: wills, land records, medical data,
   criminal case files, personnel files, and contracts. A long-term
   archive service may be used by any type of entity, e.g.,

organizations, citizens, notaries. Examples of long-term archive
service usage by submitters include:

- A company stores contracts using a third party service.

- A hospital stores medical data using an internal service.

- An individual wants to generate evidence of data possession at a
  particular point in time, e.g., for intellectual property purposes
  or endorsement of a contract.

- A law enforcement officer wants to store criminal data such that
  integrity of the data can be demonstrated years later.

For each of the above examples, there is a corresponding example
involving retrievers, e.g., a company retrieves a contract in the
case of a dispute or a law enforcement officer prepares information
for a criminal trial.

This document addresses the technical requirements for a long-term
archive service.

2. Terminology

We define the following terms based on their usage in the archiving
community, in order to provide a vocabulary for describing
requirements and the standards around them.

Arbitrator: Principal for whom the validity of archived data
   characteristics, e.g., origin, integrity or time of existence,
   must be demonstrated.

Archival Period: The period during which an archived data object is
   preserved by a long-term archive service.

Archived Data Object: Data unit to be preserved by a long-term
   archive service.

Archive Package: Collection of information including archived data
   objects and associated Evidence Record.

Cryptographic Maintenance Policy: A set of rules that defines how
   to maintain the validity of digitally signed objects should one of
   the hash or asymmetric algorithms used to create a digital
   signature become weak, or one of the private keys used to create a
   digital signature be compromised or become weak.

   Evidence: Information that may be used to demonstrate the validity
      of an archived data object or related attestations.

   Evidence Record: Collection of evidence compiled for one or more
      archived data objects. An Evidence Record may include
      acknowledgements from a long-term archive service, time stamps and
      verification data, such as public-key certificates, revocation
      information, trust anchors, policy details and role information.

   Long-Term Archive Policy: A set of rules that define operational
      characteristics of a long-term archive service.

   Long-Term Archive Service (LTA):   A service that is responsible for
      preserving data for long periods.

   Modifier: Principal who modifies attributes associated with an
      archived data object and/or Evidence Record held by a long-term
      archive service.

   Originator: Principal who produces, and possibly digitally signs,
      an archived data object. The Originator does not necessarily have
      any relationship with a long-term archive service or any awareness
      of an Evidence Record associated with the archived data object.

   Retriever: Principal who retrieves archived data objects and/or
      Evidence Records from a long-term archive service.

   Submitter: Principal who submits data objects for archiving.

   Time Stamp: An attestation generated by a Time Stamping Authority
      (TSA) that a data item existed at a certain time. For example,
      [RFC3161] specifies a structure for signed time stamp tokens as
      part of a protocol for communicating with a TSA.

   Time Stamping Authority (TSA):   A trusted service that provides
      attestations of existence of data at particular points in time.
      For example, [RFC3161] defines protocol elements for interacting
      with a TSA.

3. General Principles

   A long-term archive service may accept any type of data for
   preservation. The data might be in any format, whether textual data,
   images, documents, applications, or compound packages of multiple
   components. The data may be digitally signed, time stamped,
   encrypted, or not subject to any cryptographic processing.

A long-term archive service may preserve archived data objects as
opaque collections of bytes with the primary aim of data integrity.

A long-term archive service is not required to operate upon evidence
related to the content of archived data objects. Content-focused
operations, including data format migration or translation, may be
performed by another service. However, an LTA may incorporate
support for such services.

Different long-term archive services may establish policies and
procedures for archiving data objects over different lengths of time.
For example, an LTA may refuse to preserve archived data objects for
periods longer than 30 years. Similarly, LTAs may establish policies
that limit the types of data that will be accepted for deposit by a
particular LTA.

A long-term archive service provides evidence that may be used to
demonstrate the existence of an archived data object at a given time
and the integrity of the archived data object since that time.
Additionally, the evidence identifies the LTA(s) that have
participated in the preservation of the archived data object. If the
archived data object itself contains digitally signed data,
authentication of the signer is also possible.

A long-term archive service may be an adjunct component of a document
management system. In such cases, the Evidence Record generated and
maintained by the LTA is a property of data that is otherwise managed
by the document management system.

4. Technical Requirements

   This section describes the requirements for the protocol for
   accessing a long-term archive system and for the data formats
   associated with data preservation.

4.1. Enable Submission, Retrieval, and Deletion of Archived Data
      Objects

4.1.1. Functional Requirements

   A long-term archive service must permit clients to request the
   following basic operations:

   - submit data objects for archive

   - retrieve archived data objects

   - delete archived data objects

   Following submission, the service must provide an identifier that can
   be used to retrieve the archived data and/or associated evidence.
   For example, it may be possible to retrieve archive packages by using
   a hash value of an archived data object. Possession of this value is
   not necessarily an authorization to access the associated archived
   data object or evidence record.

   It must be possible to authenticate requests and responses, e.g., to
   enable LTAs to render an authorization decision. This may be
   accomplished by using transport security mechanisms. Requests, in
   particular retrieval or deletion requests, may be rejected if the
   requestor is not authorized. An authorization policy must be defined
   and observed by the long-term archive service. An LTA may disallow
   deletion as a matter of policy.

   The format for the acknowledgements must allow the identification of
   the archiving provider and the participating client.

   The LTA must provide an acknowledgement of the deposit that permits
   the submitter to confirm the correct data was accepted by the LTA.
   This proof need not be provided immediately.

4.1.2. Rationale

   Submission, retrieval, query state, and deletion of archived data
   objects are necessary basic functions of a long-term archive service.

   Deletion may be disallowed due to procedural difficulties in
   fulfilling the request. For example, an archived data object may be
   stored on write-once media, along with other records that are not
   subject to deletion.

   Acknowledgements may not be provided immediately due to
   implementation of a grace period. A generic query state mechanism
   should be provided to address such situations. For example, a

submission response may indicate that a submission has been accepted
and a subsequent query state response may indicate a submission has
completed all necessary preservation steps.

4.2. Operate in accordance with a long-term archive policy

4.2.1. Functional Requirements

A long-term archive service must operate in accordance with a long-
term archive service policy that defines characteristics of the
implementation of the long-term archive service. A long-term archive
service policy contains several components, including:

- Archived data object maintenance policy

- Authorization policy

- Service policy

A long-term archive service policy must include specifications of the
preservation activities performed for archived data objects subject
to the policy. A maintenance policy should define rules for the
following operational aspects: preservation activity triggers,
default archival period, and default handling upon expiration of
archival period.

Maintenance policies should include mechanism-specific details
describing LTA operation. For example, where cryptographic
mechanisms are employed, a cryptographic maintenance policy ought to
be defined.

An authorization policy should define the entities permitted to
exercise services provided by the LTA, including who is permitted to
submit, retrieve, or manage specific archived data objects.

A service policy defines the types of services provided by an LTA,
including acceptable data types, description of requests that may be
accepted, and deletion procedures.

Policies must be unambiguously identified, e.g., by an object
identifier. Alternatively, an LTA may support a protocol that
permits clients to specify policy parameters explicitly instead of by
reference to a policy.

A long-term archive service must be able to provide information
identifying the policies relevant for a given archived data object.

4.2.2. Rationale

   Similar to a certificate policies [RFC3647], which are identified
   using object identifiers, a long-term archive policy provides a
   shorthand means of technically identifying a set of rules that govern
   the operation of a long-term archive service.

   Over the course of many years, the policies under which an LTA
   operates may undergo modification. Thus, an evidence record may
   feature multiple indications of policies active at various points
   during the life of an archived data object.

4.3. Enable Management of Archived Data Objects

4.3.1. Functional Requirements

   A long-term archive service must permit clients to request the
   following basic operations:

   - specify an archival period for submitted data objects

   - extend or shorten the archival period for an archived data object

   - specify metadata associated with an archived data object

   - specify an archive policy under which the submitted data should be
      handled

   It should be possible to express an archival period in terms of time,
   an event or a combination of time and event.

   Submitters should be able to specify metadata that, for example, can
   be used to enable retrievers to render the data correctly, to locate
   data in an archive or to place data in a particular context.
   Examples include, classification codes, type of format, contributors,
   title, author, and date. Alternatively, such information may be
   included in the content of an archived data object.

   If a long-term archive service does not support a requested policy,
   it must return an error indication. A service must provide an
   indication of the archive policy enforced by the service.

4.3.2. Rationale

   Submission, retrieval, and deletion of archived data objects are
   necessary basic functions of a long-term archive service.

Specification and management of the archival period is necessary to
avoid unnecessary preservation activities.

4.4. Provide Evidence Records that Support Demonstration of Data
      Integrity

4.4.1. Functional Requirements

A long-term archive service must be capable of providing evidence
that can be used to demonstrate the integrity of data for which it is
responsible, from the time it received the data until the expiration
of the archival period of the data.

This may be achieved by providing evidence records that support the
long-term non-repudiation of data existence at a point in time, e.g.,
in the case of legal disputes. The evidence record should contain
sufficient information to enable the validity of an archived data
object's characteristics to be demonstrated to an arbitrator. The
characteristics subject to verification will vary. For example,
authentication of an originator may not be possible in all cases,
e.g., where the object submitted to the archive is not signed or
where the object does not include the necessary information to
authenticate the object's signer.

Evidence records must be structured such that modifications to an
archived data object or its evidence record can be detected,
including modifications made by administrators of an LTA.

4.4.2. Rationale

Supporting non-repudiation of data existence, integrity, and origin
is a primary purpose of a long-term archive service. Evidence may be
generated, or otherwise obtained, by the service providing the
evidence to a retriever. A long-term archive service need not be
capable of providing all evidence necessary to produce a non-
repudiation proof, and in some cases, should not be trusted to
provide all necessary information. For example, trust anchors
[RFC3280] and algorithm security policies should be provided by other
services. An LTA that is trusted to provide trust anchors could
forge an evidence record verified by using those trust anchors.

Demonstration that data has not been altered while in the care of a
long-term archive service is a first step towards supporting non-
repudiation of data. Certification services support cases in which
data must be modified, e.g., translation or format migration. An LTA
may provide certification services.

4.5. Support Data Confidentiality

4.5.1. Functional Requirements

   A long-term archive service must provide means to ensure
   confidentiality of archived data objects, including confidentiality
   between the submitter and the long-term archive service. An LTA must
   provide a means for accepting encrypted data such that future
   preservation activities apply to the original, unencrypted data.
   Encryption, or other methods of providing confidentiality, must not
   pose a risk to the associated evidence record.

   A long-term archive service should maintain contact information for
   the parties responsible for each archived data object so warning
   messages can be sent when encryption algorithms require maintenance.

4.5.2. Rationale

   Individuals may wish to use the services of a commercial long-term
   service without disclosing data to the commercial service. However,
   access to the original data may be necessary to perform some
   preservation activities.

4.6. Provide Means to Transfer Data and Evidence from One Service to
      Another

4.6.1. Functional Requirements

   It must be possible to submit data along with previously generated
   evidence, i.e., to support transfer of data from one archive to
   another. A long-term archive service must support the transfer of
   archived data objects, evidence and evidence records from one service
   to another. It must be possible for evidence records to span
   multiple providers over the course of time, without losing value as
   evidence.

4.6.2. Rationale

   Before the end of an archived data object's archival period, a long-
   term archive service may cease operation. In such cases, it must be
   possible for the archived data object (and any associated evidence)
   to be transferred to another service that will continue preservation
   of the data until the end of the archival period.

   Submitters may change service providers before the end of an archived
   data object's archival period. In such cases, it must be possible
   for the submitter to transfer an archived data object and all
   associated evidence from the original LTA to a new LTA.

4.7. Support Operations on Groups of Data Objects

4.7.1. Functional Requirements

   An LTA should support submission of groups of data objects.
   Submitters should be able to indicate which data objects belong
   together, i.e. comprise a group, and retrievers should be able to
   retrieve one, some or all members of a group of data objects.

   It should be possible to provide evidence for groups of archived data
   objects. For example, it should be possible to archive a document
   file and a signature file together such that they are covered by the
   same evidence record.

   Where an LTA operates upon groups of data objects, non-repudiation
   proof must still be available for each archived data object
   separately.

4.7.2. Rationale

   In many cases data objects belong together. Examples include:

   - a document file and an associated signature file, which are two
      separate objects

   - TIF-files representing pages of a document

   - a document file and an evidence file (possibly generated by
      another LTA)

   - a document and its translation to another format or language

   In these cases, it is to the best advantage to handle these data
   objects as a group.

5. Operational Considerations

   A long-term archive service must be able to work efficiently even for
   large amounts of archived data objects. In order to limit expenses
   and to achieve high performance, it may be desirable to minimize the
   use of trusted third parties, e.g., LTA operations should be designed
   to limit the number of time stamps required to provide the desired
   level of service.

   Necessity to access archived data objects should be minimized. It
   may only be necessary to access the archived data objects if the
   archived data objects are requested by users, or if hash algorithms
   used for indexing, or evidence record generation become insecure.

An LTA must be capable of operating in accordance with any applicable
legal regime. For example, an LTA may be required to reject a
deletion request from an authorized requestor if the target of the
request has been subpoenaed by law enforcement authorities.

Some applications may require processing of a chain of archive
policies present in an evidence record, e.g., to ensure that
compatible policies were used throughout the lifetime of the archived
data objects.

6. Security Considerations

Data is the principal asset protected by a long-term archive service.
The principle threat that must be addressed by a long-term archive
service is an undetected loss of data integrity.

In cases where signature verification relies on a PKI, certificate
revocation could retroactively invalidate previously verified
signatures. An LTA may implement measures to support such claims by
an alleged signer, e.g., collection of revocation information after a
grace period during which compromise can be reported or preservation
of subsequent revocation information.

When selecting access control mechanisms associated with data stored
by a LTA, the lifespan of the archived data object should be
considered. For example, the credentials of an entity that submitted
data to an archive may not be available or valid when the data needs
to be retrieved.

During the lifespan of an archived data object, formats may cease to
be supported. Software components to process data, including content
or signatures, may no longer be available. This could be a problem
particularly if non-standard formats are used or proprietary
processing is employed. The submitter should take care to avoid such
problems. For example, the submitter (or other authorized entity)
could periodically retrieve data, convert the data, and re-submit it
in a new format. Additional mechanisms, applications, or tools may
be needed to preserve the value of evidence records associated with
the original archived data object.

A long-term archive system may require correlation of different
identities that represent the same entity at different points in
time. For example, an individual's identity may be represented by
different employers at different points in time.

A long-term archive system must perform maintenance activities on a
schedule that considers factors such as the strength of relevant
cryptographic algorithms, lifespan of relevant certification

authorities, and revocation status of relevant entities, e.g.,
timestamp authorities. Standards for use of cryptographic algorithms
are expected to be established by organization or governmental
bodies, not by individual LTAs.

7. Acknowledgements

Thanks to members of the LTANS mailing list for review of earlier
drafts and many suggestions. In particular, thanks to Larry
Masinter, Denis Pinkas, and Peter Sylvester for review and
suggestions.

8. Informative References

   [RFC3161]   Adams, C., Cain, P., Pinkas, D., and R. Zuccherato,
               "Internet X.509 Public Key Infrastructure Time-Stamp
               Protocol (TSP)", RFC 3161, August 2001.

   [RFC3280]   Housley, R., Polk, W., Ford, W., and D. Solo, "Internet
               X.509 Public Key Infrastructure Certificate and
               Certificate Revocation List (CRL) Profile", RFC 3280,
               April 2002.

   [RFC3647]   Chokhani, S., Ford, W., Sabett, R., Merrill, C., and S.
               Wu, "Internet X.509 Public Key Infrastructure Certificate
               Policy and Certification Practices Framework", RFC 3647,
               November 2003.

Appendix A. Application Scenarios

   Below are several example application scenarios demonstrating one or
   more of the basic service features mentioned above.

A.1. Archive Service Supporting Long-Term Non-Repudiation

   A long-term archive service may store data objects, such as signed or
   unsigned documents, for authenticated users. It may generate time
   stamps for these data objects and obtain verification data during the
   archival period or until a deletion request is received from an
   authorized entity.

A.2. Pure Long-Term Non-Repudiation Service

   A long-term archive service may only guarantee non-repudiation of
   existence of data by periodically generating time stamps and
   obtaining verification data. It stores data objects (e.g., documents
   and signatures) locally only for the purpose of non-repudiation and
   does not function as a document archive for users. It does not
   support retrieval and deletion of data objects.

A.3. Long-Term Archive Service as Part of an Internal Network

   A long-term archive service may be part of an enterprise network.
   The network provider and archive service may be part of the same
   institution. In this case, the service should obtain non-repudiation
   evidence from a third party. An internally generated acknowledgement
   may be viewed worthless.

A.4. Long-Term Archive External Service

   A long-term archive service may be provided over the Internet for
   enterprises or consumers. In this case, archiving and providing
   evidence (via time stamps or other means) may be adduced by one
   organization and its own technical infrastructure, without using
   external services.

Authors' Addresses

    Carl Wallace
    Cygnacom Solutions
    Suite 5200
    7925 Jones Branch Drive
    McLean, VA   22102

    Fax: +1(703)848-0960
    EMail: cwallace@cygnacom.com

    Ulrich Pordesch
    Fraunhofer Gesellschaft
    Rheinstrasse 75
    Darmstadt, Germany   D-64295

    EMail: ulrich.pordesch@zv.fraunhofer.de

    Ralf Brandner
    InterComponentWare AG
    Otto-Hahn-Strabe 3
    Walldorf, Germany   69190

    EMail: ralf.brandner@intercomponentware.com

Full Copyright Statement

Intellectual Property

Acknowledgement