Preserving State Government Digital Information
Minnesota Historical Society

Authentication Methods

Summary

The importance of authentication of files is often discussed; however what is not generally discussed are the methods to do so. This piece summarizes common authentication methods in use today.

DISCLAIMER:
This is a topical overview and nowise intended to offer legal advice. Consult an attorney for assistance with specific concerns or for advice.

Any comments, corrections, or recommendations may be sent to the project team, care of:

Carol Kussmann
Collections Assistant, State Archives
Minnesota Historical Society
carol.kussmann@mnhs.org / 651.259.3262

Introduction

The purpose of authentication is to prove a person's identity or to verify that a document has not been modified since its creation. Logins and passwords are used by systems to authenticate user identities; but other methods are needed to authenticate digital content as it is created and shared.

Methods of authenticating digital content focus on the assurance of the transfer process, the assurance of the digital content itself, or both. Methods that focus on the transfer process are most concerned with the security during transfer and may use secure protocols or encrypted data. Other methods of authentication focus on the digital content by demonstrating that the content has not changed (over time or during transmission) or rely on a larger infrastructure of security and authentication methods.

Some of the primary means for authenticating documents are described below, providing a starting point for determining an authentication method/s that will work best for you.

Authentication Methods

HTTPS:

Most Internet transmissions are done over the Hypertext Transfer Protocol (HTTP), an insecure network. Hypertext Transfer Protocol Secure (HTTPS) is used to encrypt data during transmissions on the Internet, creating a secure channel over an insecure network. HTTPS is often used during online shopping to protect users when they send private and personal information such as their credit card number and address to online vendors.

HTTPS is a set of rules for transferring files on the Internet using HTTP and an additional protocol such as Secure Socket Layering (SSP) or Transport Layer Security (TLS) to ensure privacy between communicating applications. Using a combination of protocols, information is encrypted and decrypted during transmissions to protect against unauthorized attacks. HTTPS may also authorize the use of digital certificates (*see Digital Certificates*) to authenticate users/senders. Keep in mind that although information is 'safe' during transmission, there is nothing in place to authenticate the data itself over the long-term. Meaning, you can't prove that if the information was going to be reused, that it has not been changed since it was received.

Transport Layer Security (TLS):

"Transport Layer Security (TLS) is a protocol that ensures privacy between communicating applications and their users on the Internet. When a server and client communicate, TLS ensures that no third party may eavesdrop or tamper with any message. TLS is the successor to the Secure Sockets Layer (SSL). TLS is composed of two layers: the TLS Record Protocol and the TLS Handshake Protocol. The TLS Record Protocol provides connection security with some encryption method such as the Data Encryption Standard (DES). The TLS Record Protocol can also be used without encryption. The TLS Handshake Protocol allows the server and client to authenticate each other and to negotiate an encryption algorithm and cryptographic keys before data is exchanged." [1]

Other protocols such as Kerberos, Secure Shell (SSH), and Internet Protocol Security (IPsec) provide similar protections.

Checksums:

A checksum is a unique number assigned to a document (or collection of documents) based on a mathematical formula. The checksum provides a method for checking the authenticity of a document because a checksum is created based on the documents content at the bit level. Checksums are often used to verify that records have not changed over time (while sitting in storage or in use) or that they have not changed during network

[1] Nieminen, Mikko. *Transport Layer Security (TLS).* SearchSecurity.com. December 6, 2010. http://searchsecurity.techtarget.com/sDefinition/0,,sid14_gci557332,00.html

transmissions. Checksums can be run on a document at anytime, and if the numbers are identical then the document has not changed; if the same number was not generated then the document has changed. If differences are detected, this method does not tell you what has changed or when something was changed, just that the two documents are no longer identical.

For example, if you need to send a document to someone else and want to ensure that the document they received is exactly the same as what you sent, you could apply a hash function (the mathematical algorithm) to the document, resulting in the checksum value. The person who received the document could then run a check to see if the checksum value is still the same. If the checksums do not match, the document has been modified or corrupted.

There are many different checksum algorithms, with the most common belonging to the MD5 and SHA families. Checksums or similar hash values are often included as a portion of other authenticity methods.

Digital Signature:

A digital signature is a method for authenticating digital content as well as the signer's identity. To apply a digital signature to a document, an algorithm, or hash function, is run on the document which creates a unique number that can be used to prove authenticity. This hash value (or checksum) is then encoded using the signer's private key (see PKI), which in turn produces a digital signature (another unique number). The digital signature is attached to the original document and the document is then shared.

Anyone who receives a digitally signed document can then verify who signed the document and if the content received is the same as it was when the signature was applied. To do this, the receiver must have a public key (see PKI) that matches the signer's private key (a key pair); only then can the receiver decode the digital signature (a combination of the private key and hash value of the document) to verify the signer's identity.

If the signer's identity has been verified, the next step validates the hash value decoded from the digital signature and matches a newly run hash value on the content. If these two hash values match, the content has not been changed and can be considered authentic. If these tags do not match, or if the signer's identity could not be identified, the authentication process fails.

Digital signatures can be used alone or as part of a larger authentication process.

Public Key Infrastructure (PKI) / Public Key Cryptography:

Public Key Infrastructure (PKI) is a structure that uses public key cryptography, registration authorities, certificate authorities, and digital certificates to verify users and authenticate documents being transmitted over non-encrypted web services. The messages themselves may or may not be encrypted.

Public Key Cryptography

Public key cryptography uses two keys to verify and authenticate transactions being sent and received over the Internet. A key pair, a private key and a public key, are created per user. The private key is known only to the owner and is never transmitted over the Internet for security purposes. The public key can be shared with anyone that the owner of the private key wants to share information with.

In general these two keys are used to make sure that only people that can match key pairs are able to read each other's messages. For example, a record creator can encrypt a document using his private key, and only people with his public key will be able to decode the message. This not only verifies who sent the message, but assists with the security of the message itself.

Digital signatures use Public Key Cryptography.

Registration Authorities and Certificate Authorities

Creating a larger authentication infrastructure, registration authorities and certificate authorities are used as trusted third parties to verify and certify identities.

A Registration Authority (RA) is responsible for verifying users' identities and notifying the Certificate Authority (CA) that a user is who they say they are. Registration Authorities are not always used; sometimes the CA performs this role themselves. Either way, it is important that both authorities are trusted.

A Certificate Authority is responsible for managing access to public keys by issuing digital certificates, while the private keys remain private. Digital certificates are only issued if the owner of the public key can be verified by the CA or RA as appropriate. Owners of public keys could be individuals, organizations, or other entities. Digital certificates issued by a CA are trusted because the CA is verifying identities as well as attaching their own signature to the document.

Digital Certificate

A digital certificate issued by a CA contains the user's public key, information about the certificate's owner, and one or more digital signatures (such as the CA's certifying the certificate). The digital signature is a way to validate the trustworthiness of the document and or sender. "The certificate is basically a public key with one or two forms of ID attached, plus a hearty stamp of approval from some other trusted individual."[2]

Third Party Vendors

Another option for complying with authenticity requirements is to work with third party vendors. These vendors use similar if not the same technologies as above but have often developed methods of applying these technologies that fulfill very specific roles or blend into daily routines. Depending on your goals, resources, and situation you may want to explore the options third party vendors have to offer. It is also important to know that different vendors have different ways of charging for their services. Examples of a few third party vendor products are provided below for informational purposes only.

- Adobe Live Cycle
  - Files saved as PDFs in this system can have a "blue ribbon" attached to them that certifies the document. Adobe Reader can detect and display this ribbon informing people who created the document and that it is authentic. The Government Printing Office (GPO) currently uses this method. For example, the blue ribbon on one document from their site says "Certified by Superintendent of Documents pkisupport@gpo.gov, United States Government Printing Office, certificate issued by GeoTrust CA for Adobe."[3] (In addition, the GPO uses digital signatures on their documents.)

- ProofSpace
  - ProofMark[4] applies a digital tamper-detection seal and trusted timestamp to electronic records. Using transient key technology, ProofMark creates a self-validating proof of time stamp with cryptographic tamper detection, to provide persistent authenticity of records throughout a record's lifecycle. Transient key technology assigns new key pairs to a record each time it is modified outside of a set time interval. Old private keys are destroyed, also increasing the security of the method.

[2] Network Associates, Inc. Introduction to Cryptography: *How PGP Works*. 1999. http://www.pgpi.org/doc/pgpintro/#p9
[3] Example certified document: http://www.gpo.gov/fdsys/pkg/GOVMAN-1997-05-30/pdf/GOVMAN-1997-05-30-Pg1.pdf
[4] Proofspace. *ProofMark*. http://www.proofspace.com/technology/index.php

- Surety
    - AbsoluteProof[5]: Using a web-based program that can be integrated into daily workflows, AbsoluteProof "seals" documents by applying digital timestamps that protect the integrity and authenticity of records. These timestamps can be verified at anytime by any third party.

- TruSeal
    - Uses digital timestamps and digital signatures to authenticate records in any electronic data format that can be validated to verify authenticity and detect tampering at anytime, anywhere, by anyone. TruSeal also verifies metadata integrity and is capable of keeping audit trails.[6] TruSeal can be run as a web application or hosted on your own servers.

NDIIPP Example Applications

- BagIt

  BagIt was created by the California Digital Library, Library of Congress, and Stanford University for the purpose of packaging digital content for transfer with automation of receipt, storage and retrieval. "Bags" are created and consist of the files, a txt file for their associated checksums, a txt file for information about the bag, a txt file for information about the version of BagIt being used, and a txt file containing checksums for the txt files. Hash values can be compared to make sure the files and bag have not changed during transmissions or over time.

- XML Wrapper

  The XML Wrapper was created by the Minnesota Office of the Revisor of Statues and Thomson Reuters for the purpose of creating an XML file that includes XML bill data, associated metadata, and all other file formats a legislature produces (PDF, Word). The prototype application for the wrapper not only wraps the files with metadata but computes hash values (md5 and sha-1) on the xmlwrapper file and creates a zip file containing the xmlwrapper file and both hash values. These hash values can then be used to verify that the content has not changed over time or during transmissions.

[5] Surety. *Absolute Proof.* 2010. http://www.surety.com/data-integrity-protection/absoluteproof.aspx
[6] Tru Data Integrity. *Home Page.* 2010. http://www.tru-dataintegrity.com/

Resources

Artic Soft Technologies Limited. *An Introduction to PKI (Public Key Infrastructure).* 2010. http://www.articsoft.com/public_key_infrastructure.htm

Introduces PKI, explains public and private keys used for digital signatures, certificates, storage methods for keys, certificate authorities, registration authorities, and certificate management techniques.

MBA Knowledge Base. *How Public Key Infrastructure (PKI) Works?* 2010. http://www.mbaknol.com/business-finance/how-public-key-infrastructure-pki-works/

Explains encryption, digital certificates, digital signatures, PKI, certificate authorities, and registration authorities.

Wikipedia. *File: Public-Key-Infrastructre.svg.* April 2008. http://en.wikipedia.org/wiki/File:Public-Key-Infrastructure.svg

Image representation of how key cryptology is used in relation with a certificate authority, registration authorities, and a verifying authority.