Prototype for Authentication of Official Electronic Record and Pricing

-

Office of the Revisor of Statutes
Minnesota Legislature
https://www.revisor.mn.gov/beta/rules/

-

August 2012

Introduction

The Uniform Electronic Legal Material Act [1], section 5, contains requirements for the authentication of legal materials in an electronic record. This paper describes a software prototype, https://www.revisor.mn.gov/beta/rules/ , built to satisfy these requirements. The information technology (IT) components and the approximate cost of building the prototype are given.

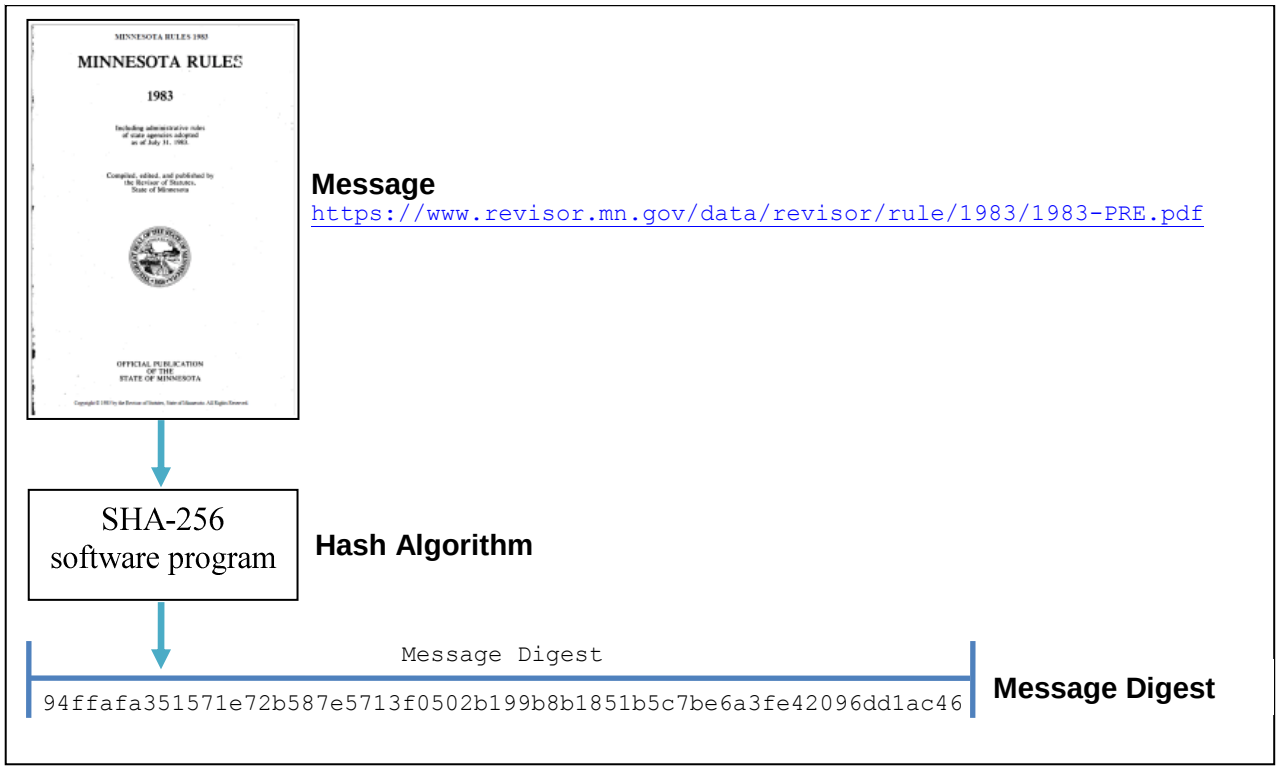
Description

In the prototype, the core technologies used for authentication are a hash algorithm, and secure communications across the Internet. The National Institute of Standards and Technology [2] (NIST) gives this definition of a hash algorithm:

> "A hash algorithm (alternatively, hash "function") takes binary data, called the message, and produces a condensed representation, called the message digest."

Figure 1 shows the message (a PDF computer file). The message is read by a hash algorithm (SHA-256 algorithm). The algorithm processes every bit in the message and then writes out the message digest. The message digest is unique to the message.

Figure 1. Hash algorithm usage



The second core technology is secure communications across the Internet. Secure communications are accomplished using a web server configured: a) to use the https protocol (instead of http); and b) a certificate signed by a trusted certificate authority. This technology eliminates third-party alteration of data transmitted between browser and web server. [3]

In the prototype, a message digest is computed for each PDF file published to the office's web server. The message digest and the PDF file are saved in a database. Additional metadata about the document is also saved, e.g., the document's official name.

When a user wants to authenticate a PDF file residing on the user's computer, a message digest of the file is computed and compared to the message digest saved at the time of publication.

IT Components

Figure 2 shows the authentication prototype's IT components. Table 1 lists the specific components used by the office.

Figure 2. Message digest comparison for document authentication

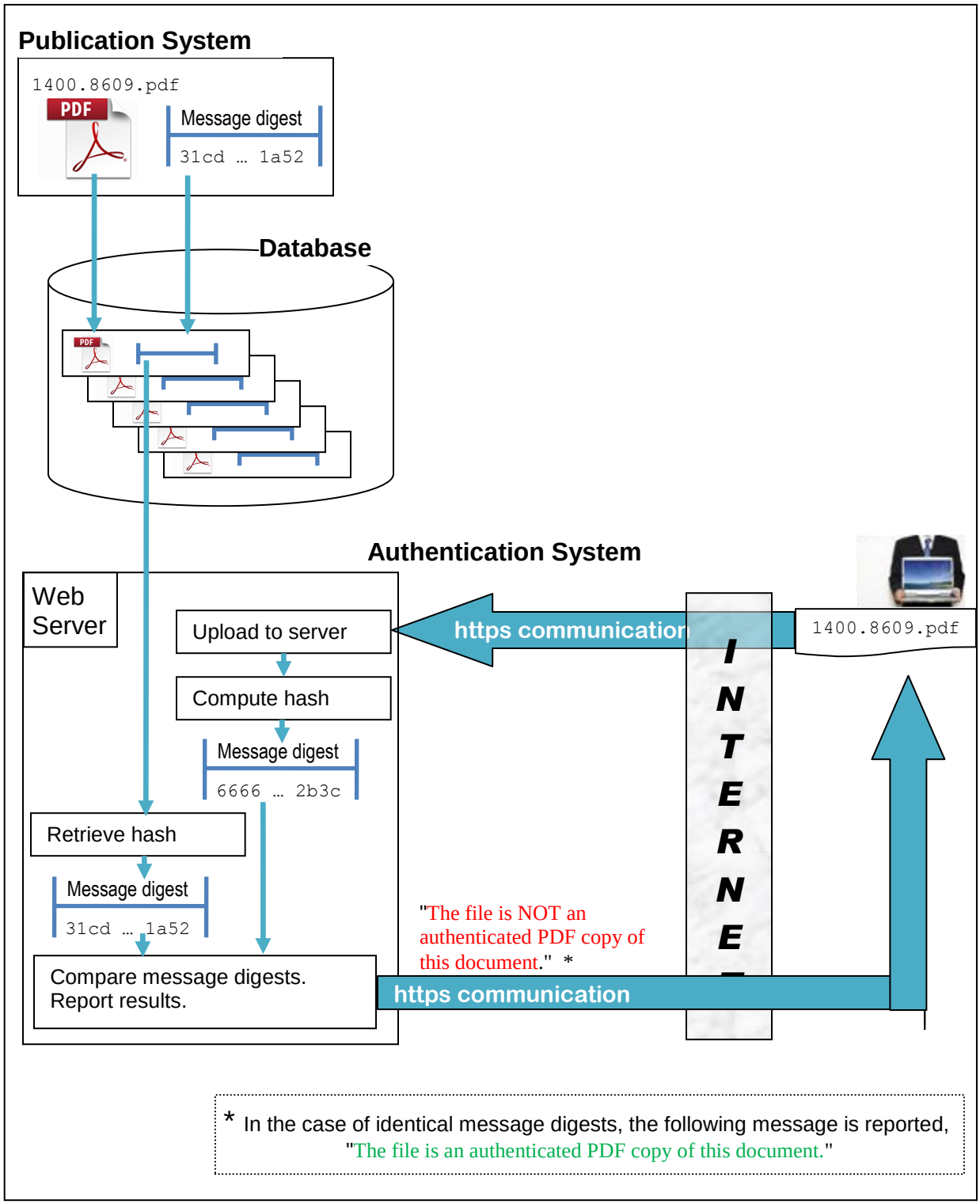


Table 1. Components used in prototype

Notes:

i. "New costs to office: $0" means that the office already possesses the necessary hardware, and/or commercial software.

ii. **Initial Cost** is the first-time cost incurred by an organization to acquire the item.

Publication System New costs to office: $0 **Note**: This table describes the components used to calculate the message digest and save data in a database. It does not include a description of the commercial software used to create PDF files.			
Component	**Used in Prototype**	**Initial Cost**	**Ongoing Cost**
Custom software.	- Java - SQL - Eclipse (development environment)	- $0. Java, SQL, and Eclipse are free. - 3-5 days of programmer's time.	- Perpetual maintenance of custom code.

Database New costs to office: $0			
Component	**Used in Prototype**	**Initial Cost**	**Ongoing Cost**
Relational database management system (RDBMS)	- Oracle Database	- $ xx,000 (depends on configuration)	- $ xx,000 (depends on configuration)
Table design and creation	- SQL commands	- 3-5 days of database administrator's time.	- Perpetual database administration.

Authentication System New costs to office: $0			
Component	**Used in Prototype**	**Initial Cost**	**Ongoing Cost**
Web server hardware	- HP DL360 server - Red Hat Linux operating system	- $5,000 (depends on configuration)	- $5,000 every 4 years for server replacement.
Web server software application	- Apache HTTP Server	- $0 (free)	- $0
SSL certificate	- DigiCert.com wildcard SSL certificate	- $475 per year	- $475 per year
Custom software, web pages	- HTML - PHP - SQL	- 10 days of programmer's time.	- Perpetual maintenance of custom code.

Advantages

- No/Low initial cost. Prototype was built using existing office resources.
  - $0 for new developers. Existing programmers and database administrator built the prototype.
  - $0 for new commercial hardware or software.
  - $0 for training in new languages or commercial applications.
- Low and stable ongoing costs.
- No reliance on external companies.
  - No risk that the company: a) closes; b) discontinues their product; or c) increases the price of their product.
  - No license imposed limits on the number of documents that can be processed per year.
- Public users are not required to install and learn third-party applications.
- System can expand to authenticate additional file formats e.g., XML, scanned image files, audio files, etc.
- Every PDF document is saved in the database. As future hash algorithms are developed the new message digest for each PDF can be programmatically computed and updated in the database.
- System design supports long-term document preservation. When documents are moved to new hardware and database applications, the message digests can be used to confirm that documents are unchanged.

Disadvantages

- Custom software need to be developed
- Perpetual maintenance of custom code.
- Perpetual maintenance of the database.

References

[1] National Conference of Commissioners on Uniform State Laws (2011). **UNIFORM ELECTRONIC LEGAL MATERIAL ACT**. http://www.uniformlaws.org/Shared/Docs/AM2011_Prestyle%20Finals/UELMA_PreStyleFinal_Jul11.pdf

[2] National Institute of Standards and Technology, Computer Security Resource Center.

A. . **Cryptographic Hash Project**. http://csrc.nist.gov/groups/ST/hash/index.html
B. **Drivers**. http://csrc.nist.gov/drivers/index.html .
C. March 2012. **FIPS PUB 180-4 "Secure Hash Standard (SHS)".** http://csrc.nist.gov/publications/fips/fips180-4/fips-180-4.pdf .
D. **Example Algorithms**. http://csrc.nist.gov/groups/ST/toolkit/examples.html .

[3]Biztech (July 2007). **HTTP vs. HTTPS**.
http://www.biztechmagazine.com/article/2007/07/http-vs-https

[4] Office of Legislative Counsel (2011). **Authentication of Primary Legal Materials and Pricing Options**.
http://www.mnhs.org/preserve/records/legislativerecords/docs_pdfs/CA_Authentication_WhitePaper_Dec2011.pdf

[5] Minnesota State Archives. **Preserving state government digital information**.
http://www.mnhs.org/preserve/records/legislativerecords/authentication.htm

Appendix A. Database Schema.

Relevant columns in DB table

Name	Type
DOC_KEY	NUMBER
DATE_INSERT	DATE
DATE_MODIFY	DATE
DATE_EXPIRE	DATE
CHAPTER_NUMBER	VARCHAR2(16)
PART_NUMBER	VARCHAR2(16)
DOCUMENT_NAME	VARCHAR2(25)
HTML_FILE	VARCHAR2(50)
HTML_SIZE	NUMBER(8)
HTML_HASH	VARCHAR2(64)
PDF_FILE	VARCHAR2(50)
PDF_SIZE	NUMBER(8)
PDF_HASH	VARCHAR2(64)
XML_FILE	VARCHAR2(50)
XML_SIZE	NUMBER(8)
XML_HASH	VARCHAR2(64)
HASH_ALGORITHM	VARCHAR2(12)
HASH_DATE	DATE