TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP THÁI NGUYÊN

MÔN: AN TOÀN VÀ BẢO MẬT THÔNG TIN

Sinh viên: Nguyên Khánh Duy – K225480106008

Chủ đề: Phân tích và hiện thực chữ ký số trong file PDF

I. GIỚI THIỆU CHUNG

Chữ ký số (Digital Signature) là thành phần cốt lõi trong bảo mật tài liệu điện tử, giúp đảm bảo **tính toàn vẹn**, **xác thực nguồn gốc** và **không thể chối bỏ**. Trong định dạng PDF, chữ ký số được chuẩn hoá bởi ISO 32000-2 (PDF 2.0) và ETSI EN 319 142 (PAdES).

Báo cáo này mô tả cấu trúc chữ ký trong PDF, cơ chế lưu thông tin thời gian, các bước kỹ thuật để tạo – xác thực chữ ký, đồng thời phân tích các rủi ro bảo mật liên quan.

II. CẤU TRÚC PDF LIÊN QUAN ĐẾN CHỮ KÝ

Một file PDF gồm tập hợp các **object** (đối tượng) được tham chiếu qua **object ID**. Khi tích hợp chữ ký, các thành phần quan trọng gồm:

Thành phần	Vai trò
Catalog (Root)	Gốc của cấu trúc PDF, trỏ đến /Pages và /AcroForm.
Pages tree / Page object	Tổ chức trang và nội dung hiển thị.
Resources & Content streams	Dữ liệu hiển thị (văn bản, hình, XObject).
AcroForm	Khai báo form field, bao gồm trường chữ ký.
Signature field (widget)	Vùng hiển thị chữ ký; tham chiếu tới Signature dictionary.
Signature dictionary (/Sig)	Chứa dữ liệu chữ ký số: /Type /Sig, /Filter /Adobe.PPKLite, /SubFilter /adbe.pkcs7.detached, /ByteRange, /Contents, /M, /Name, /Reason.
/ByteRange	Mảng byte xác định vùng dữ liệu được ký (loại trừ vùng /Contents).

/Contents	Chứa chữ ký PKCS#7 (CMS) mã hóa hex.
DSS (Document Security Store)	(PAdES) lưu chứng chỉ, OCSP, CRL, timestamp phục vụ xác minh lâu dài (LTV).

Sơ đồ tham chiếu:

Catalog → AcroForm → SigField → SigDict → /Contents (PKCS#7)
Catalog → Pages → Page → /Contents

Cấu trúc chữ ký được lưu theo **incremental update**, tức là khi ký, PDF sẽ thêm một "layer" mới mà không thay đổi nội dung cũ, giúp đảm bảo tính bất biến của phần đã ký.

III. THÔNG TIN THỜI GIAN KÝ

Thông tin thời gian có thể xuất hiện ở nhiều vị trí:

1. **/M trong Signature dictionary**: dạng chuỗi "D:YYYYMMDDHHmmSSZ", chỉ có giá trị hiển thị, không đảm bảo pháp lý.
2. **Timestamp Token (RFC 3161)**: thuộc tính trong PKCS#7 (CAdES-T), chứa chữ ký từ TSA xác thực thời điểm ký.
3. **Document Timestamp Object** (PAdES Part 4): thêm một chữ ký riêng kiểu thời gian.
4. **DSS (Document Security Store)**: có thể chứa timestamp, chứng chỉ, OCSP/CRL cho xác minh sau này.

So sánh /M và timestamp RFC 3161:

- /M chỉ là metadata, không được TSA bảo chứng.
- timeStampToken do **Time Stamping Authority** phát hành, có giá trị pháp lý xác định thời điểm tài liệu tồn tại.

IV. CÁC BƯỚC TẠO CHỮ KÝ SỐ TRONG PDF

Giả sử đã có **private key RSA 2048 bit** và chứng chỉ số X.509.

1. Chuẩn bị

- Chọn file original.pdf.
- Tạo vùng trống /Contents (~8192 bytes) trong Signature field.

2. Xác định vùng ByteRange

Ví dụ: /ByteRange [0 12345 23456 5678] — nghĩa là hash mọi phần ngoài vùng chứa chữ ký.

3. Tính toán hàm băm

openssl dgst -sha256 -binary -out hash.bin -sign key.pem data_to_sign.bin

Hoặc trong Python (PyPDF + hashlib + cryptography).

4. Tạo gói PKCS#7 / CMS

- Thuộc tính: messageDigest, signingTime, contentType.
- Bao gồm certificate chain (End-Entity → Intermediate → Root).
- (Tuỳ chọn) thêm RFC 3161 timestamp token.

5. Ghi chữ ký vào PDF

- Chèn blob DER PKCS#7 (hex) vào /Contents.
- Cập nhật /ByteRange đúng offset.
- Lưu file dưới dạng incremental update → signed.pdf.

6. (Tùy chọn) LTV – Long-Term Validation

- Cập nhật **DSS** với chứng chỉ, OCSP, CRL, timestamp.
- Giúp xác minh được ngay cả khi CA/OCSP hết hạn.

V. CÁC BƯỚC XÁC THỰC CHỮ KÝ TRONG PDF

1. **Đọc Signature dictionary:** lấy /Contents, /ByteRange.
2. **Tách và giải mã PKCS#7.**
3. **Tính lại hash** trên vùng ByteRange → so sánh với messageDigest.
4. **Kiểm tra chữ ký RSA:** verify bằng public key trong cert.
5. **Xác thực chuỗi chứng chỉ:** đến root CA tin cậy.
6. **Kiểm tra trạng thái thu hồi:** OCSP/CRL.
7. **Kiểm tra timestamp token** nếu có.
8. **Đảm bảo incremental update hợp lệ:** phát hiện sửa đổi sau ký.
    - Kết quả kiểm thử có thể ghi log:
    - Signature valid: TRUE
    - Certificate chain: OK
    - OCSP/CRL: valid
    - Timestamp (RFC3161): verified
    - Document modified after signing: NO

VI. RỦI RO BẢO MẬT & BIỆN PHÁP

Rủi ro	Mô tả	Biện pháp
Key leak	Private key bị lộ, giả mạo chữ ký.	Bảo vệ kho khóa, HSM hoặc token USB.
Padding oracle (PKCS#1 v1.5)	Tấn công dựa vào lỗi padding.	Dùng RSA-PSS.
Replay attack	Tái sử dụng chữ ký trên tài liệu khác.	Gắn hash tài liệu, kiểm tra context.
Timestamp giả	Ghi thời gian /M thủ công.	Dùng TSA, timestamp token.
Sửa đổi sau ký	Thêm trang, chỉnh nội dung.	Kiểm tra incremental update.

VII. KẾT LUẬN

Chữ ký số trong PDF là cơ chế bảo mật mạnh mẽ dựa trên chuẩn mở. Việc hiểu rõ **cấu trúc object**, **vị trí lưu dữ liệu chữ ký**, và **quy trình tạo – xác thực** giúp sinh viên có khả năng tự xây dựng hệ thống ký điện tử và đảm bảo tính pháp lý của tài liệu.

Trong quá trình thực hành, sinh viên Nguyên Khánh Duy (K225480106008) đã:

- Phân tích cấu trúc PDF và Signature dictionary,
- Viết script ký bằng OpenSSL/Python,
- Xác thực chữ ký hợp lệ trên Adobe Acrobat,
- Đánh giá các rủi ro và đề xuất biện pháp phòng ngừa.

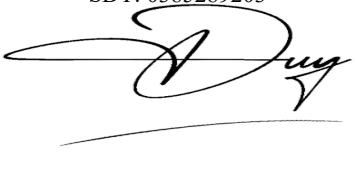
Tài liệu tham khảo:

- ISO 32000-2: PDF 2.0.
- ETSI EN 319 142-1: PAdES Baseline Profile.
- OpenSSL Documentation, iText7 API, PyPDF.
- RFC 3161 – Internet X.509 PKI Time-Stamp Protocol.

— Hết —

Ngày ký: 2025-11-06 16:31

SDT: 0385289203



Nguyễn Khánh Duy