BÁO CÁO BÀI TẬP VỀ NHÀ
MÔN: AN TOÀN VÀ BẢO MẬT THÔNG TIN
Chủ đề: Chữ ký số trong file PDF

Sinh viên thực hiện: Nguyễn Như Khiêm

MSSV: K225480106030

Lớp: K58KTP

Nội dung:

Tệp PDF này dùng để thử nghiệm quy trình tạo chữ ký số (8 bước) theo yêu cầu của đề bài môn An toàn và Bảo mật thông tin. Báo cáo mô tả cấu trúc PDF liên quan chữ ký, cách lưu thời gian ký, và các rủi ro bảo mật, dựa trên ISO 32000-1 và PAdES. Minh họa qua file original.pdf (gốc), signed.pdf (đã ký), tampered.pdf (bị chỉnh sửa).

1. Cấu trúc PDF liên quan chữ ký số

PDF là định dạng dựa trên object (indirect objects), với chữ ký số nhúng qua AcroForm và Signature dictionary (SigDict). Chữ ký sử dụng PKCS#7/CMS để mã hóa dữ liệu (/Contents), /ByteRange chỉ vùng hash không thay đổi. Incremental updates cho phép ký nhiều lần, DSS (PAdES) lưu dữ liệu xác minh dài hạn.

Object refs quan trọng và vai trò

Object Ref	Vai trò
Catalog (ref 0)	Root, chứa /AcroForm và /Pages. Quản lý toàn bộ file, truy xuất SigField.
Pages tree (/Pages ref N)	Cây trang từ Catalog đến Page. Tổ chức /Annots (SigField widget).
Page object (ref M)	Mô tả trang, chứa /Contents và /Resources. Hiển thị vị trí chữ ký.
Resources (ref P)	Dictionary font/XObject. Hỗ trợ render hình chữ ký tay (như "Khiem" trong signed.pdf).
Content streams (/Contents ref Q)	Stream vẽ nội dung. Được hash trong /ByteRange; tamper phá chữ ký.
XObject (ref R)	External object (hình ảnh). Lưu chữ ký tay minh họa.
AcroForm (ref S)	Từ Catalog, quản lý /Fields (SigField).
Signature field (widget ref T)	Annotation trong Page, trỏ /Sig. Vị trí UI chữ ký.
Signature dictionary (/Sig ref U)	Chứa /Contents (PKCS#7), /ByteRange, /M. Lưu dữ liệu ký, validate hash.
/ByteRange	[start1, length1, start2, length2]. Vùng hash (trừ /Contents).

Object Ref	Vai trò
/Contents	Stream PKCS#7 (cert + hash). Giải mã để lấy cert "Khiem".
Incremental updates	Append revision mới. Hỗ trợ ký lặp.
DSS (/DSS ref V)	Trong Catalog, lưu OCSP/CRLs. Xác minh offline.

Ví dụ minh họa: Trong signed.pdf, SigField1 (widget) trỏ SigDict với /ByteRange [0, 267828, 274120, 821], hash nội dung gốc. Tamper thêm text phá ByteRange, dẫn đến invalid.

Sơ đồ object (ASCII):

Catalog
├── /Pages ──> Page ──> /Contents (hash)
│ ├── /Resources ──> /XObject (hình "Khiem")
│ └── /Annots ──> SigField ──> /Sig (/Contents, /ByteRange)
└── /AcroForm ──> /Fields ──> SigField ──> SigDict
└── /DSS ──> OCSP/CRLs

2. Thời gian ký được lưu ở đâu?

Thời gian ký lưu đa vị trí để chống chối bỏ, theo ISO 32000-1 và PAdES.

❖ **Vị trí lưu**

+ **/M trong /Sig:** Thời gian từ máy ký (text "D:20251027165803+00'00'"). Dễ truy xuất nhưng không signed.
+ **Timestamp token (RFC 3161):** Trong /Contents (PKCS#7), attribute timeStampToken từ TSA. Signed cryptographically.
+ **Document timestamp (PAdES-DTS):** SigDict riêng, timestamp toàn file sau ký.
+ **DSS:** Trong Catalog, lưu timestamps + OCSP/CRLs cho validation dài hạn.

❖ **Khác biệt /M và RFC3161**

+ /M: Nội bộ, từ hệ thống (dễ giả mạo, không pháp lý cao). Ví dụ signed.pdf: /M = "D:20251027165803+00'00'" (UTC, khớp Adobe 23:58 VN).
+ RFC3161: External từ TSA, signed hash document + thời điểm (chính xác, pháp lý, chống chối bỏ). PAdES yêu cầu cho advanced signature; verify kiểm tra token TSA.

Ví dụ: signed.pdf dùng /M (không RFC3161), verify pyHanko báo thời gian signer_reported_dt = 2025-10-27 23:58:03+07:00.

3. Rủi ro bảo mật chữ ký số trong PDF

Chữ ký PDF an toàn nếu tuân thủ PAdES, nhưng dễ bị tấn công nếu tamper hoặc cert yếu.

❖ **Rủi ro chính**

+ Tamper nội dung (/Contents hoặc ByteRange): Thay đổi text/hình phá hash, dẫn đến invalid (như tampered.pdf thêm "Xin chào", mod_level = FORM_FILLING). Rủi ro: Giả mạo tài liệu, phát hiện qua verify (integrity check).
+ Replay attack: Ký lại SigDict với timestamp cũ, dùng incremental updates. Rủi ro: Chối bỏ thời gian; giảm bằng RFC3161/DSS.
+ Cert revocation (CRL/OCSP): Cert "Khiem" hết hạn hoặc thu hồi không kiểm tra. Rủi ro: Ký giả mạo; PAdES-DSS lưu CRL để validate offline.
+ Side-channel attack: Leak private key từ signer_key.pem. Rủi ro: Forge chữ ký; bảo vệ bằng HSM (Hardware Security Module).
+ Incremental updates lạm dụng: Ký nhiều lớp, lớp sau che lớp trước. Rủi ro: Ẩn thay đổi; kiểm tra qua pyHanko's modification_level.

Ví dụ minh họa: signed.pdf hợp lệ (bottom_line = VALID). Tamper thêm text → invalid (bottom_line = False, "File bị chỉnh sửa"). Giảm rủi ro: Dùng PAdES-LTA (long-term validation) với DSS.

Biện pháp: Validate bằng công cụ (pyHanko/Adobe), kiểm tra /ByteRange, dùng TSA cho timestamp.

Kết luận :

Chữ ký số PDF đảm bảo toàn vẹn qua cấu trúc AcroForm/SigDict và ByteRange, thời gian ký qua /M/RFC3161, nhưng rủi ro tamper/cert cần DSS/PAdES để giảm. Thử nghiệm với original.pdf → signed.pdf → tampered.pdf chứng minh quy trình 8 bước hiệu quả.

Tài liệu tham khảo :

- ISO 32000-1: PDF Standard.
- ETSI EN 319 142-1: PAdES.
- pyHanko docs: https://pyhanko.readthedocs.io.