

BÀI TẬP VỀ NHÀ – MÔN: AN TOÀN VÀ BẢO MẬT THÔNG TIN

Chủ đề: Chữ ký số trong file PDF

Giảng viên: Đỗ Duy Cốp

Sinh viên: Nguyễn Thị Thu Hiền – K225480106015

Thời điểm giao: 2025-10-24 11:45

Đối tượng áp dụng: Toàn bộ sv lớp học phần 58KTPM

Hạn nộp: Sv upload tất cả lên github trước 2025-10-31 23:59:59

I. MÔ TẢ CHUNG

Sinh viên thực hiện báo cáo và thực hành: phân tích và hiện thực việc nhúng, xác thực chữ ký số trong file PDF.

Phải nêu rõ chuẩn tham chiếu (PDF 1.7 / PDF 2.0, PAdES/ETSI) và sử dụng công cụ

thực thi (ví dụ iText7, OpenSSL, PyPDF, pdf-lib).

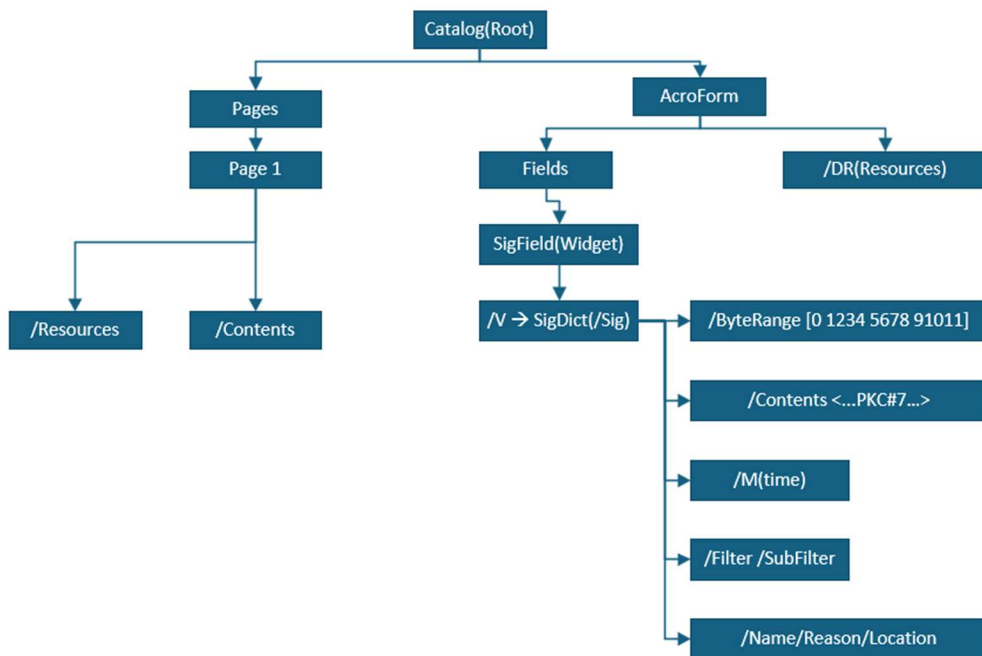
II. CÁC YÊU CẦU CỤ THỂ

1) Cấu trúc PDF liên quan chữ ký (Nghiên cứu)

| Thành phần | Vai trò |
|---------------------------------|--|
| Catalog (Root) | Là đối tượng gốc của PDF, liên kết đến toàn bộ cấu trúc (trang, form, v.v). |
| Pages Tree | Danh sách phân cấp chứa các Page Object (từng trang trong PDF). |
| Page Object | Mô tả nội dung của một trang (text, hình, font, resources,...). |
| Resources | Chứa tham chiếu đến các tài nguyên (font, ảnh, XObject, form fields). |
| Content Streams | Dòng lệnh vẽ (text, hình, vector) được render lên trang. |
| XObject | Đối tượng đồ họa có thể tái sử dụng (ảnh, form con, template). |
| AcroForm | Đối tượng mô tả các form fields , bao gồm cả Signature Field . |
| Signature Field (Widget) | Field hiển thị vùng chữ ký (hình ảnh, tên người ký, lý do,...). |

| | |
|--------------------------------------|--|
| Signature Dictionary (/Sig) | Chứa dữ liệu chữ ký số thật sự: cert, hash, thời gian,... |
| /ByteRange | Mảng 4 số chỉ định các vùng byte được hash (ngoại trừ vùng /Contents). |
| /Contents | Chứa dữ liệu chữ ký số PKCS#7 (đã mã hóa base16/hex). |
| Incremental Updates | PDF lưu chữ ký bằng cách thêm phần mới chứ không ghi đè, giúp giữ nguyên lịch sử. |
| DSS (Document Security Store) | (Theo chuẩn PAdES) – chứa chứng chỉ, OCSP, CRL phục vụ xác thực lâu dài (LTV). |

- Mối quan hệ giữa các Object



2) Thời gian ký được lưu ở đâu?

- Nêu tất cả vị trí có thể lưu thông tin thời gian:

- + /M trong Signature dictionary (dạng text, không có giá trị pháp lý).
 - + Timestamp token (RFC 3161) trong PKCS#7 (attribute timeStampToken).
 - + Document timestamp object (PAdES).
 - + DSS (Document Security Store) nếu có lưu timestamp và dữ liệu xác minh.
- Giải thích khác biệt giữa thông tin thời gian /M và timestamp RFC3161.

- Các vị trí có thể lưu thông tin thời gian

| Vị trí lưu | Mô tả ngắn gọn | Đặc điểm |
|---|--|---|
| /M trong Signature Dictionary (/Sig) | Thuộc tính /M (ModificationDate) lưu chuỗi thời gian khi ký , dạng D:YYYYMMDDHHmmss+TZ. | Do phần mềm ký tự ghi lại – chỉ có giá trị tham khảo, không có giá trị pháp lý vì có thể chỉnh sửa. |
| Timestamp Token (RFC 3161) | Nằm bên trong chữ ký PKCS#7/CMS , ở phần signedAttributes → timeStampToken . Do Time Stamping Authority (TSA) cấp. | Có chữ ký của TSA , đảm bảo thời điểm ký được chứng thực – có giá trị pháp lý . |
| Document Timestamp Object (PAdES) | Đối tượng riêng trong PDF (loại /Sig có /Type /DocTimeStamp). | Dùng để “đóng dấu thời gian toàn tài liệu”, kể cả khi không có người ký. Chuẩn của ETSI EN 319 142 (PAdES) . |
| DSS (Document Security Store) | Chứa dữ liệu xác minh lâu dài: OCSP, CRL, Timestamp, Cert chain . | Dùng để lưu giữ timestamp & xác minh lâu dài (LTV) . Không phải vị trí gốc lưu, mà là nơi bảo tồn thông tin thời gian phục vụ kiểm chứng sau này. |

- So sánh /M và timestamp RFC3161

| Tiêu chí | /M (ModificationDate) | Timestamp (RFC 3161) |
|------------------|--|---|
| Nguồn tạo | Phần mềm ký PDF tự thêm (local system time). | Cấp bởi TSA (Time Stamping Authority) – tổ chức tin cậy. |
| Định dạng | Văn bản dạng D:YYYYMMDDHHmmss+TZ. | Một đối tượng PKCS#7 có chữ ký điện tử riêng . |

| | | |
|------------------------|---|---|
| Giá trị pháp lý | Không có — có thể bị chỉnh sửa, phụ thuộc máy người ký. | Có — được TSA ký và xác minh được thời điểm thật. |
| Chức năng chính | Chỉ hiển thị “thời điểm ký theo máy người dùng”. | Chứng minh tài liệu đã tồn tại tại thời điểm cụ thể được TSA xác nhận. |
| Mức độ tin cậy | Thấp | Cao |

- Rủi ro bảo mật

1. Rò rỉ hoặc đánh cắp Private Key

- Nguy cơ: Nếu khoá bí mật (private key) bị lộ — ví dụ do lưu trên máy tính cá nhân, USB không mã hoá, hoặc bị malware lấy cắp — kẻ tấn công có thể ký giả mạo tài liệu hợp lệ.
- Hậu quả: Mất hoàn toàn tính pháp lý của chữ ký, vì người khác có thể “ký thay” chủ sở hữu.
- Giải pháp:
 - Lưu khoá trong HSM hoặc USB Token bảo mật.
 - Dùng password mạnh và mã hoá khóa riêng khi lưu file .pem.
 - Không dùng khoá cá nhân cho nhiều mục đích (ký + test...).

2. Tấn công sửa đổi nội dung (Tampering)

- Nguy cơ: Nếu phần mềm ký không xác định chính xác ByteRange, có thể chừa lại vùng chưa được băm → cho phép chèn nội dung mới mà vẫn xem là hợp lệ.
- Ví dụ: “Incremental update attack” – thêm trang mới sau khi ký, hoặc sửa chú thích mà không phá chữ ký.
- Giải pháp:
 - Kiểm tra ByteRange chặt chẽ trong quá trình xác thực.
 - Không tin chữ ký chỉ vì viewer báo “hợp lệ”, hãy so sánh hash thực tế.

- Sử dụng PDF 2.0 + PAdES-LTV (có cơ chế bảo vệ incremental).
-

3. Giả mạo hiển thị chữ ký (UI attack)

- Nguy cơ: Một số trình xem PDF (PDF viewer) hiển thị vùng chữ ký, tên, lý do... nhưng không hiển thị chính xác nội dung đã được bấm → kẻ xấu có thể “tráo vùng hiển thị”.
 - Hậu quả: Người dùng tin rằng nội dung được ký, trong khi vùng thực sự ký không chứa dữ liệu đó.
 - Giải pháp:
 - Dùng viewer đáng tin cậy (Adobe, Foxit, pyHanko verify).
 - Khi xác minh, xem toàn bộ hash range, không chỉ “ô chữ ký”.
-

4. Thuật toán băm hoặc mã hóa yếu

- Nguy cơ: Dùng SHA-1, MD5, hoặc RSA < 2048-bit → dễ bị va chạm hash hoặc phá mã.
 - Giải pháp:
 - Chỉ dùng SHA-256 hoặc SHA-512, RSA \geq 2048-bit.
 - Ưu tiên chuẩn mới như RSA-PSS, ECDSA nếu được.
-

5. Tấn công Replay / Resigning

- Nguy cơ: Dữ liệu chữ ký (PKCS#7 blob) có thể bị sao chép sang tài liệu khác có cấu trúc tương tự để “tái sử dụng chữ ký”.
- Giải pháp:
 - Nhúng unique document ID vào vùng hash.
 - Kiểm tra các trường /DocMDP (document permissions) trong PDF.

- Dùng timestamp token (RFC 3161) để đảm bảo tính duy nhất theo thời gian.

6. Tấn công vào Timestamp hoặc TSA

- Nguy cơ: Nếu TSA (Time Stamping Authority) bị giả mạo hoặc compromise, có thể tạo timestamp sai lệch thời gian ký.
- Giải pháp:
 - Dùng TSA tin cậy có chứng thực CA quốc gia.
 - Khi verify, kiểm tra chain của TSA cert.
 - Lưu timestamp token trong DSS để xác thực về sau (LTV).

7. Sai sót trong xác minh chuỗi chứng chỉ (Chain validation)

- Nguy cơ: Nếu trình xác minh không kiểm tra OCSP/CRL, chứng chỉ hết hạn hoặc bị thu hồi vẫn coi là hợp lệ.
- Giải pháp:
 - Bắt buộc xác thực OCSP/CRL mỗi lần verify.
 - Lưu kết quả trong DSS (Document Security Store) theo chuẩn PAdES-LTV.

8. Rủi ro phần mềm ký không chuẩn

- Nguy cơ: Dùng thư viện PDF hoặc công cụ ký (open-source) nhưng không tuân chuẩn PDF 1.7 / PAdES → dễ sinh ra chữ ký không hợp lệ hoặc không tương thích Adobe.
- Giải pháp:
 - Ưu tiên thư viện có chuẩn rõ: iText7, BouncyCastle, pyhanko, Adobe Acrobat.

- Sau khi ký, luôn verify bằng ≥ 2 công cụ khác nhau để đối chiếu.

9. Không bảo vệ long-term (LTV)

- Nguy cơ: Khi chứng chỉ hoặc OCSP/CRL hết hạn, chữ ký không còn xác minh được.
- Giải pháp:
 - Áp dụng PAdES-LTV để lưu chứng chỉ + OCSP + CRL + timestamp vào DSS.
 - Khi xác minh sau nhiều năm vẫn có thể khôi phục trạng thái hợp lệ.