BÀI TẬP AN TOÀN VÀ BẢO MẬT THÔNG TIN

Nguyễn Đức Anh Tú – K225480106070

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)

- Mô tả ngắn gọn: Catalog, Pages tree, Page object, Resources, Content streams, XObject, AcroForm, Signature field (widget), Signature dictionary (/Sig), /ByteRange, /Contents, incremental updates, và DSS (theo PAdES).
- Liệt kê object refs quan trọng và giải thích vai trò của từng object trong lưu/truy xuất chữ ký.
- Đầu ra: 1 trang tóm tắt + sơ đồ object (ví dụ: Catalog → Pages → Page → /Contents ; Catalog → /AcroForm → SigField → SigDict).

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.

3) Các bước tạo và lưu chữ ký trong PDF (đã có private RSA)

- Viết script/code thực hiện tuần tự:

1. Chuẩn bị file PDF gốc.
2. Tạo Signature field (AcroForm), reserve vùng /Contents (8192 bytes).
3. Xác định /ByteRange (loại trừ vùng /Contents khỏi hash).
4. Tính hash (SHA-256/512) trên vùng ByteRange.
5. Tạo PKCS#7/CMS detached hoặc CAdES:
   - Include messageDigest, signingTime, contentType.
   - Include certificate chain.
   - (Tùy chọn) thêm RFC3161 timestamp token.
6. Chèn blob DER PKCS#7 vào /Contents (hex/binary) đúng offset.
7. Ghi incremental update.
8. (LTV) Cập nhật DSS với Certs, OCSPs, CRLs, VRI.
   - Phải nêu rõ: hash alg, RSA padding, key size, vị trí lưu trong PKCS#7.
   - Đầu ra: mã nguồn, file PDF gốc, file PDF đã ký.

4) Các bước xác thực chữ ký trên PDF đã ký:

- Các bước kiểm tra:

1. Đọc Signature dictionary: /Contents, /ByteRange.
2. Tách PKCS#7, kiểm tra định dạng.
3. Tính hash và so sánh messageDigest.
4. Verify signature bằng public key trong cert.
5. Kiểm tra chain → root trusted CA.
6. Kiểm tra OCSP/CRL.
7. Kiểm tra timestamp token.
8. Kiểm tra incremental update (phát hiện sửa đổi).

- Nộp kèm script verify + log kiểm thử.

BÀI LÀM

1. Cấu trúc PDF liên quan chữ ký:

❖ Mô tả ngắn gọn các object:

- Catalog**:** Đối tượng gốc của tài liệu, thường trỏ đến /Pages và /AcroForm.
- Pages tree / Page object**:** Tổ chức cấu trúc trang, chứa /Contents (dòng nội dung) và các tài nguyên liên quan.
- Resources / Content streams / XObject: Lưu trữ dữ liệu hiển thị thực tế trên trang.
- AcroForm**:** Từ điển cấp biểu mẫu, chứa danh sách Fields (trường biểu mẫu) và có thể bao gồm /SigFlags.
- Signature field **(**widget**):** Trường biểu mẫu kiểu chữ ký (Sig), đồng thời là chú thích hiển thị vùng chữ ký trên trang.
- Signature dictionary **(**/Sig**):** Đối tượng lưu thông tin chữ ký số, gồm /Type /Sig, /Filter, /SubFilter (loại chữ ký), /ByteRange, /Contents (dữ liệu chữ ký PKCS#7), /M (thời điểm ký), /Name, /Location.
- /ByteRange**:** Mảng bốn giá trị [start1 length1 start2 length2] xác định hai đoạn dữ liệu được băm, phần giữa là vùng chứa chữ ký.
- /Contents**:** Vùng chứa dữ liệu chữ ký PKCS#7 ở dạng DER, thường được cấp trước dung lượng cố định (ví dụ 8192 byte) trong bản cập nhật tăng dần.
- Incremental updates**:** Cho phép thêm các object và bảng xref mới mà không thay đổi nội dung cũ, giúp phát hiện mọi chỉnh sửa sau khi ký.
- DSS (Document Security Store): Lưu trữ thông tin xác thực lâu dài (LTV) như chứng thư, OCSP/CRL và token đóng dấu thời gian.

❖ Sơ đồ object:

❖ Object refs quan trọng & vai trò:

- /AcroForm (Root object): Chứa danh sách các trường biểu mẫu, giúp trình đọc xác định vị trí và cấu trúc của các trường chữ ký.
- SigField (Field object): Đại diện cho trường chữ ký, mô tả tên, vị trí hiển thị và liên kết với widget annotation trên trang.
- SigDict (Signature dictionary): Lưu trữ thông tin chữ ký, gồm metadata và dữ liệu PKCS#7 trong /Contents — đây là đối tượng trung tâm của chữ ký số.
- /ByteRange: Xác định phạm vi dữ liệu trong tệp được dùng để tính băm, bỏ qua phần /Contents.
- Incremental update (xref + trailer**):** Khi có thay đổi sau khi ký, tệp sẽ có xref/trailer mới khác với bản gốc, qua đó cho phép phát hiện chỉnh sửa.

2. Thời gian ký được lưu ở đâu?

❖ Vị trí có thể lưu thông tin thời gian:

- /M trong Signature dictionary: ghi dạng văn bản (ví dụ D:20251026...), chỉ mang tính mô tả, không có giá trị pháp lý.
- Token dấu thời gian (theo RFC 3161) được nhúng trong gói PKCS#7/CMS thông qua thuộc tính *timeStampToken* hoặc *id-aa-signatureTimeStampToken* trong CAdES. Token này do TSA ký trên giá trị băm để xác nhận thời điểm ký.
- Đối tượng timestamp ở cấp tài liệu (PAdES): là dạng timestamp riêng trong PDF, áp dụng cho toàn bộ tài liệu chứ không gắn với một trường chữ ký cụ thể.
- DSS (Document Security Store): nơi lưu token dấu thời gian (.tsr), chứng chỉ, phản hồi OCSP/CRL phục vụ xác thực dài hạn (LTV).

❖ Khác biệt chính /M vs RFC3161 timestamp:

- /M: chỉ là chuỗi định dạng ngày giờ do signer ghi vào dictionary; có thể bị giả mạo (không được ký độc lập). Không đủ cho chứng thực thời điểm.

- RFC3161 timestamp: do một Time Stamping Authority (TSA) ký trên digest của PKCS#7/CMS (hoặc của dữ liệu), do đó là bằng chứng thời điểm độc lập