WebBai2Security

Trương Văn Quyến - K225480106083

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

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)

- 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.

III. YÊU CẦU NỘP BÀI

1. Báo cáo PDF ≤ 6 trang: mô tả cấu trúc, thời gian ký, rủi ro bảo mật.

2. Code + README (Git repo hoặc zip).

3. Demo files: original.pdf, signed.pdf, tampered.pdf.

4. (Tuỳ chọn) Video 3–5 phút demo kết quả.

IV. TIÊU CHÍ CHẤM

- Lý thuyết & cấu trúc PDF/chữ ký: 25%
- Quy trình tạo chữ ký đúng kỹ thuật: 30%
- Xác thực đầy đủ (chain, OCSP, timestamp): 25%
- Code & demo rõ ràng: 15%
- Sáng tạo mở rộng (LTV, PAdES): 5%

V. GHI CHÚ AN TOÀN

- Vẫn lưu private key (sinh random) trong repo. Tránh dùng private key thương mại.
- Dùng RSA ≥ 2048-bit và SHA-256 hoặc mạnh hơn.
- Có thể dùng RSA-PSS thay cho PKCS#1 v1.5.
- Khuyến khích giải thích rủi ro: padding oracle, replay, key leak.

VI. GỢI Ý CÔNG CỤ

- OpenSSL, iText7/BouncyCastle, pypdf/PyPDF2.
- Tham khảo chuẩn PDF: ISO 32000-2 (PDF 2.0) và ETSI EN 319 142 (PAdES). ---

---

Bài tập

Cấu trúc PDF liên quan chữ ký số

Khi tích hợp chữ ký số, tệp PDF không ghi đè lên nội dung gốc mà thêm một bản cập nhật gia tăng (incremental update). Phần này chứa các đối tượng (object) mới phục vụ việc lưu và hiển thị chữ ký.

1. Thành phần chính trong cấu trúc PDF có chữ ký

Thành phần Vai trò:

Catalog: Nút gốc của tài liệu, trỏ tới cây trang (/Pages) và biểu mẫu (/AcroForm).

Pages Tree / Page Object: Quản lý các trang và nội dung hiển thị.

AcroForm: Định nghĩa các trường biểu mẫu, trong đó có trường chữ ký.

Signature: Field (Widget) Vùng hiển thị chữ ký, gắn vào vị trí cụ thể trên trang PDF.

Signature: Dictionary (/Sig) Lưu thông tin về người ký, thời gian ký và dữ liệu chữ ký (PKCS#7/CAdES).

ByteRange: Xác định vùng byte được ký (loại trừ vùng chứa chữ ký).

Contents Chứa dữ liệu chữ ký số (chuỗi PKCS#7).

DSS: (Document Security Store) Lưu dữ liệu xác minh dài hạn (certificates, OCSP, CRL, timestamp).

3. Cấu trúc Signature Dictionary

Đây là đối tượng trung tâm của quá trình ký:

/Type /Sig

/Filter /Adobe.PPKLite

/SubFilter /adbe.pkcs7.detached

/Name (Người ký)

/M (D:20251030T123456+07'00')

/Reason (Ký báo cáo học phần)

/Location (Thai Nguyen)

/ByteRange [0 12345 56789 99999]

/Contents <3082A5...> % dữ liệu PKCS#7

/Cert (certificate chain)

Giải thích:

- /Filter và /SubFilter: xác định định dạng chữ ký (Adobe, PKCS#7, CAdES, RFC3161…).
- /ByteRange: vùng dữ liệu được hash trước khi ký.
- /Contents: khối dữ liệu PKCS#7/CMS, chứa hash, chứng chỉ và timestamp.
- /M: thời gian ký (dạng text, không có giá trị pháp lý).
- /Reason, /Location, /Name: thông tin mô tả của người ký.

4. Incremental Update và DSS

- Incremental Update: mọi thay đổi (chữ ký, chỉnh sửa) được ghi nối tiếp vào cuối file; phần mềm xác minh có thể phát hiện mọi thay đổi sau khi ký.
- DSS (Document Security Store): lưu trữ thêm chứng chỉ, OCSP, CRL và timestamp để xác minh lâu dài (LTV).

5. Tóm tắt

File PDF có chữ ký số là sự mở rộng của cấu trúc PDF chuẩn, bổ sung các đối tượng:

- AcroForm để chứa trường ký,
- Signature Dictionary để lưu dữ liệu chữ ký,
- DSS để đảm bảo xác minh lâu dài.

Nhờ cơ chế incremental update, PDF có thể lưu nhiều chữ ký mà vẫn bảo toàn nội dung gốc.

Thời gian ký được lưu ở đâu trong PDF

Khi một file PDF được ký, thời điểm ký có thể được ghi lại ở nhiều vị trí khác nhau. Mỗi loại lưu trữ có mục đích và giá trị pháp lý khác nhau.

2. Mô tả chi tiết

a. /M trong Signature Dictionary

- Dạng: /M (D:YYYYMMDDHHmmSS+07'00')
- Lấy từ đồng hồ hệ thống người ký.

- Không được ký bảo vệ → có thể chỉnh sửa → không hợp pháp.

b. signingTime trong PKCS#7

- Là thuộc tính trong phần signedAttributes của CMS (PKCS#7).
- Được bảo vệ bởi chữ ký, nhưng vẫn phụ thuộc thời gian máy người ký → chỉ có giá trị tham khảo.

c. timeStampToken (RFC 3161)

- Do TSA (Time Stamping Authority) cấp và ký, xác nhận dữ liệu tồn tại tại thời điểm cụ thể.
- Có chữ ký của TSA → chứng cứ pháp lý về thời gian ký.

d. Document Timestamp Object (PAdES)

- Là object riêng trong PDF, thường có /SubFilter /ETSI.RFC3161.
- Được TSA ký, áp dụng cho toàn bộ tài liệu (document-level timestamp).

e. DSS (Document Security Store)

- Lưu trữ timestamp, OCSP, CRL giúp xác minh lâu dài (LTV).
- Hỗ trợ xác thực ngay cả khi TSA/CA gốc hết hạn.

3. So sánh giữa /M và RFC 3161 timestamp

4. Kết luận

Chỉ tem thời gian RFC 3161 hoặc document timestamp (PAdES) mới có giá trị chứng thực pháp lý.

Các trường như /M chỉ mang tính hiển thị và không thể dùng làm bằng chứng thời gian ký.

III. Rủi ro bảo mật khi ký và xác thực PDF

Mặc dù chữ ký số trong PDF dựa trên nền tảng mật mã học an toàn, song trong thực tế vẫn tồn tại nhiều rủi ro kỹ thuật và khai thác nếu không tuân thủ chuẩn.

1. Rủi ro thường gặp

a. Thay đổi nội dung sau khi ký (Incremental Update Attack)

- PDF cho phép lưu thêm nội dung ở phần cuối mà không xóa phần cũ.

- Kẻ tấn công có thể chèn nội dung “vô hại” vào incremental update để đánh lừa người đọc, trong khi chữ ký vẫn hiện “hợp lệ”.

b. Sửa vùng /ByteRange

- Nếu công cụ xác minh xử lý sai, có thể bỏ qua một phần dữ liệu nằm ngoài /ByteRange.
- Điều này cho phép chèn nội dung giả mạo mà vẫn không phát hiện thay đổi hash.

c. Tái sử dụng chữ ký (Replay Attack)

- Một chữ ký hợp lệ từ tài liệu khác có thể bị sao chép sang file mới nếu không có cơ chế ràng buộc nội dung (content binding).

d. Lộ khóa riêng (Private Key Leak)

- Người ký lưu private key không an toàn hoặc gửi cùng mã nguồn → có thể bị giả mạo toàn bộ chữ ký.
- Đề xuất: sử dụng RSA ≥ 2048 bit, lưu khóa trong HSM hoặc token bảo mật.

e. Tấn công vào định dạng PKCS#1 và padding

- Một số thư viện cũ dùng RSA/PKCS#1 v1.5 dễ bị “padding oracle”.
- Khuyến nghị dùng RSA-PSS hoặc ECDSA/SHA-256.

f. Lỗi xác thực chuỗi chứng chỉ hoặc timestamp

- Nếu CA hết hạn, TSA ngừng hoạt động, hoặc dữ liệu xác minh không được lưu trong DSS → file sẽ không còn xác thực được.
- Giải pháp: sử dụng PAdES-LTV để nhúng toàn bộ dữ liệu xác minh vào file.

3. Kết luận

- Chữ ký số trong PDF giúp bảo đảm tính toàn vẹn, xác thực và chống chối bỏ, nhưng chỉ khi được thực hiện đúng chuẩn (PAdES, RFC 3161).
- Mọi hệ thống ký phải bảo vệ khóa riêng, lưu timestamp hợp lệ và kiểm tra đầy đủ các vùng dữ liệu được ký.
- Đảm bảo điều này sẽ giúp tài liệu PDF có giá trị pháp lý và an toàn lâu dài.