I. GIỚI THIỆU CHUNG

- Báo cáo này trình bày quá trình phân tích, hiện thực và xác thực chữ ký số trong tài liệu PDF bằng thư viện iText7, tuân theo chuẩn PDF 2.0 (ISO 32000-2) và PAdES (ETSI EN 319 142)
Mục tiêu là giúp sinh viên hiểu rõ cấu trúc PDF khi có chữ ký số, quy trình ký và xác minh chữ ký, cũng như các vấn đề bảo mật liên quan.

- Chữ ký số trong PDF là cơ chế đảm bảo tính toàn vẹn, xác thực nguồn gốc và chống chối bỏ cho tài liệu điện tử. Khi người dùng ký một file PDF, hệ thống sẽ chèn dữ liệu chữ ký (PKCS#7/CMS) vào

vùng định sẵn trong cấu trúc file mà không làm thay đổi phần nội dung còn lại.

II. CẤU TRÚC PDF LIÊN QUAN ĐẾN CHỮ KÝ SỐ

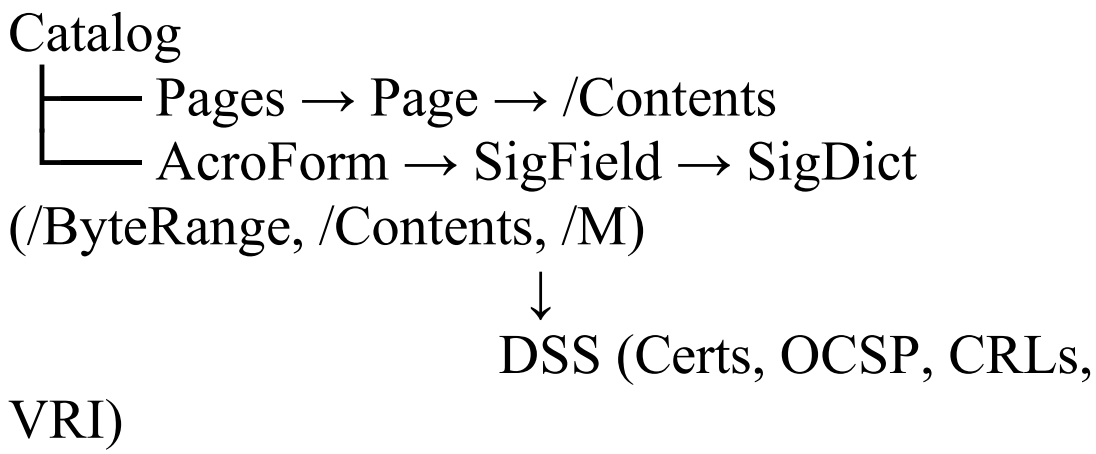
Trong tài liệu PDF, chữ ký được lưu thông qua một tập hợp các đối tượng (object) có quan hệ liên kết với nhau. Các thành phần quan trọng bao gồm:

- Catalog (Root Object): Đối tượng gốc của tài liệu, trỏ đến /Pages và /AcroForm.
- Pages Tree: Cấu trúc cây quản lý toàn bộ các trang của file PDF.

- Page Object: Mỗi trang là một object, chứa thông tin nội dung, tài nguyên và con trỏ đến /Contents.
- Resources / XObject: Tập hợp tài nguyên đồ họa (ảnh, font, form...).
- AcroForm: Đối tượng định nghĩa các trường biểu mẫu (form field), bao gồm trường chữ ký (Signature field).
- Signature Field (Widget): Vùng hiển thị chữ ký trên trang PDF.
- Signature Dictionary (/Sig): Chứa thông tin kỹ thuật của chữ ký
  như /Filter, /SubFilter, /ByteRange, /Contents, /M, /Reason, /Location...
- /ByteRange: Xác định vùng byte được dùng để tính hash (vùng chứa chữ ký bị loại trừ).

- /Contents: Vùng chứa dữ liệu chữ ký PKCS#7/CMS dưới dạng nhị phân (hex).
- Incremental Update: Mỗi lần ký thêm, PDF không ghi đè mà ghi phần mở rộng ở cuối file.
- DSS (Document Security Store): Lưu trữ dữ liệu phục vụ xác thực lâu dài (LTV): chứng chỉ, CRL, OCSP, timestamp.

*Sơ đồ cấu trúc:



III. VỊ TRÍ LƯU THỜI GIAN KÝ TRONG PDF

Khi ký tài liệu PDF, thông tin thời gian có thể được ghi ở nhiều nơi:

1. /M trong Signature Dictionary:
   Dạng text (ISO 8601), ví dụ D:20251024T114500Z. Đây chỉ là thông tin mô tả, không có giá trị pháp lý vì có thể thay đổi.
2. Timestamp Token (RFC 3161) trong PKCS#7:
   Là thuộc tính timeStampToken do TSA (Time Stamping Authority) cấp, có giá trị pháp lý, bảo đảm thời điểm ký là thực tế.

3. Document Timestamp (PAdES):
   Là một dạng chữ ký đặc biệt bao phủ toàn bộ tài liệu, giúp xác minh thời gian của toàn file.
4. DSS (Document Security Store):
   Có thể chứa timestamp cùng chứng chỉ xác thực để phục vụ xác minh dài hạn (LTV).

=> Khác biệt chính:
/M chỉ là metadata người ký tự ghi, còn RFC 3161 timestamp được ký bởi TSA – một bên thứ ba tin cậy, nên có giá trị pháp lý xác định thời điểm ký thực tế.

IV. QUY TRÌNH TẠO CHỮ KÝ SỐ TRONG PDF (với iText7)

Giả sử đã có private key RSA (được tạo bằng OpenSSL).

Các bước thực hiện:

1. Chuẩn bị file PDF gốc.
2. Tạo trường chữ ký (Signature field) trong AcroForm.
3. Dự trữ vùng /Contents (~8192 bytes) để chèn dữ liệu chữ ký.
4. Xác định /ByteRange – loại trừ vùng /Contents khỏi hash.
5. Tính hash (SHA-256 hoặc SHA-512) trên vùng ByteRange.
6. Tạo chữ ký PKCS#7 (CMS Detached):
    - Bao gồm các thuộc tính: messageDigest, signingTime, contentType.

    - Đính kèm certificate chain.
    - (Tùy chọn) thêm timestamp token (RFC3161).
7. Chèn blob DER PKCS#7 vào /Contents tại offset định sẵn.
8. Ghi incremental update để bảo toàn nội dung file gốc.
9. (Tùy chọn) Thêm DSS với Certs, OCSP, CRLs để hỗ trợ xác minh lâu dài (LTV).

Thuật toán dùng:

- Hash: SHA-256
- RSA Key: 2048-bit
- Padding: PKCS#1 v1.5 hoặc RSA-PSS

- Chuẩn
  ký: adbe.pkcs7.detached hoặc ETSI.CAdES.det
  ached

V. QUY TRÌNH XÁC THỰC CHỮ KÝ TRONG PDF

Khi người dùng mở file PDF đã ký, phần mềm (như Adobe Acrobat) hoặc script xác minh sẽ thực hiện:

1. Đọc Signature dictionary để
   lấy /Contents và /ByteRange.
2. Trích xuất PKCS#7/CMS blob, giải mã để lấy thông tin người ký và hash gốc.

3. Tính lại hash trên vùng ByteRange, so sánh với messageDigest trong PKCS#7.
4. Dùng public key trong certificate để verify chữ ký.
5. Kiểm tra chuỗi chứng chỉ (certificate chain) đến Root CA tin cậy.
6. Kiểm tra OCSP hoặc CRL để phát hiện chứng chỉ bị thu hồi.
7. Xác minh timestamp token (RFC 3161).
8. Phân tích incremental update để phát hiện file bị chỉnh sửa sau khi ký.

Kết quả xác minh có thể ở ba trạng thái:

- Chữ ký hợp lệ.
- Chữ ký hợp lệ nhưng chứng chỉ hết hạn/thu hồi.

- File bị chỉnh sửa sau khi ký hoặc hash không khớp.

VI. RỦI RO BẢO MẬT VÀ GIẢI PHÁP

1. Các rủi ro thường gặp:

- Lộ private key → giả mạo chữ ký.
- Padding oracle attack trên RSA-PKCS#1 v1.5.
- Replay attack: tái sử dụng chữ ký hợp lệ cho tài liệu khác.
- Chỉnh sửa file mà không cập nhật lại ByteRange.
- Lưu chữ ký không bảo mật (plain file).

2. Biện pháp an toàn:

- Dùng RSA-PSS và SHA-256 hoặc mạnh hơn.
- Sử dụng timestamp TSA và DSS (LTV).
- Bảo vệ private key trong HSM hoặc kho mã hóa an toàn.
- Thêm OCSP và CRL để phát hiện chứng chỉ bị thu hồi.
- Áp dụng chuẩn ETSI PAdES-B-LT hoặc PAdES-LTA để đảm bảo xác minh lâu dài.

VII. KẾT LUẬN

- Qua bài thực hành, sinh viên hiểu rõ cách thức nhúng, lưu trữ và xác thực chữ ký số trong file PDF bằng iText7.
  Chữ ký số giúp đảm bảo tính toàn vẹn, xác

thực và chống chối bỏ, đóng vai trò quan trọng trong các giao dịch điện tử và hành chính công.

- Việc triển khai đúng chuẩn (PDF 2.0 + PAdES) giúp tài liệu có giá trị pháp lý lâu dài và có thể xác minh độc lập mà không cần phần mềm riêng.