BÀI TẬP SỐ 2

MÔN: AN TOÀN VÀ BẢO MẬT THÔNG TIN

Tên: Nguyễn Bá Trung

Msv: k215480106067

I. GIỚI THIỆU CHUNG

Chữ ký số trong file PDF là cơ chế cho phép xác thực **nguồn gốc**, **toàn vẹn**, và **thời gian ký** của tài liệu điện tử. Chuẩn PDF (ISO 32000-1, PDF 1.7) và mở rộng **PAdES (ETSI EN 319 142)** quy định chi tiết về cấu trúc và quy trình ký/xác thực chữ ký số trong tài liệu PDF.

Trong bài này, sinh viên nghiên cứu cách **nhúng**, **xác thực**, và **quản lý chữ ký số** theo các chuẩn sau:

- **PDF 1.7 / PDF 2.0** (ISO 32000-1 / ISO 32000-2).
- **PAdES (ETSI EN 319 142, 319 122, 319 102)** – chuẩn châu Âu cho chữ ký PDF dài hạn.
- Sử dụng công cụ thực thi: **OpenSSL**, **PyPDF**, hoặc **iText7**.

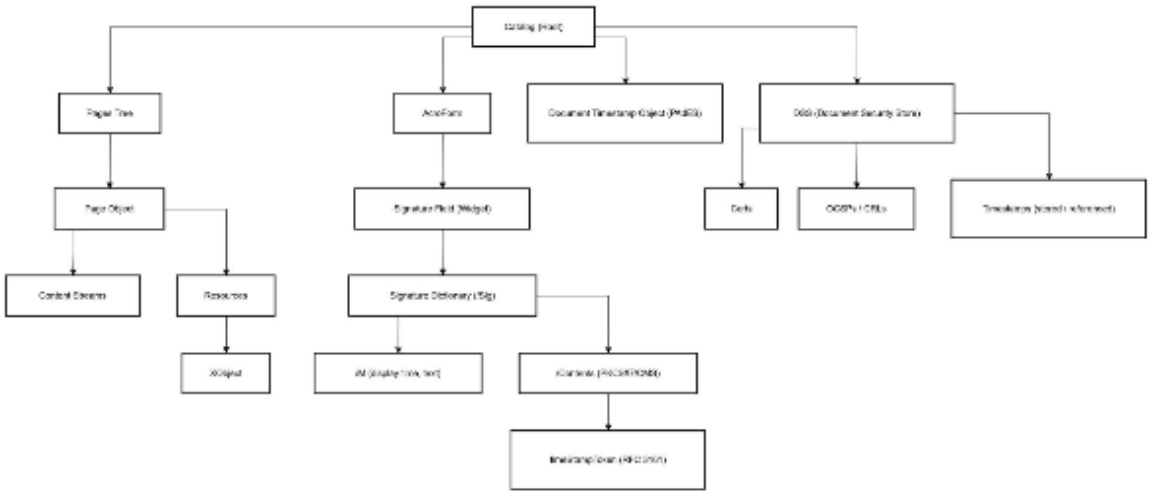
II. CẤU TRÚC FILE PDF LIÊN QUAN CHỮ KÝ SỐ

Một file PDF bao gồm tập hợp các **object** có liên kết chéo (cross-reference table). Khi chèn chữ ký, chỉ một phần của file được cập nhật theo cơ chế **incremental update**, giúp bảo toàn nội dung gốc.

1. Các thành phần chính

Thành phần	Vai trò
Catalog	Object gốc của tài liệu, trỏ đến Pages, AcroForm, và các thuộc tính toàn cục.
Pages Tree	Cấu trúc cây lưu danh sách các trang (Page Object).
Page Object	Đại diện cho từng trang, chứa tham chiếu đến nội dung và tài nguyên.
Resources	Chứa font, hình ảnh, XObject được sử dụng trên trang.
Content Streams	Mã lệnh mô tả nội dung hiển thị (text, hình, vẽ…).
XObject	Đối tượng đồ họa (hình ảnh, form XObject).
AcroForm	Biểu mẫu chứa các trường (field), trong đó có **Signature Field**.
Signature Field (Widget)	Trường biểu mẫu đại diện vùng hiển thị chữ ký.
Signature Dictionary (/Sig)	Lưu thông tin chữ ký số: /Filter, /SubFilter, /ByteRange, /Contents, /M, /Name, /Reason...
/ByteRange	Xác định vùng dữ liệu được băm để ký (loại trừ vùng /Contents).
/Contents	Chứa dữ liệu chữ ký số (PKCS#7/CMS blob).
Incremental Update	Cơ chế ghi phần chữ ký vào cuối file PDF mà không thay đổi nội dung cũ.
DSS (Document Security Store)	(Theo PAdES) – chứa chứng chỉ, OCSP, CRL, timestamp để xác minh dài hạn (LTV).

2. Sơ đồ quan hệ object



III. THỜI GIAN KÝ TRONG FILE PDF

Trong PDF, thông tin về **thời gian ký** có thể xuất hiện tại nhiều vị trí khác nhau:

Vị trí	Mô tả	Giá trị pháp lý
/M (trong /Sig)	Thời gian ký do phần mềm ghi lại, dạng D:YYYYMMDDHHmmSSZ.	❌ Không có giá trị pháp lý.
RFC 3161 Timestamp Token	Thẻ thời gian được **TSA** cấp trong PKCS#7 (timeStampToken).	✅ Có giá trị pháp lý, chứng minh thời điểm ký.
Document Timestamp (PAdES)	Một chữ ký đặc biệt gắn thời gian cho toàn bộ tài liệu.	✅ Dùng trong PAdES-LTV.
DSS	Lưu timestamp và dữ liệu xác minh (Certs, OCSP, CRL).	✅ Hỗ trợ xác thực dài hạn.

Khác biệt giữa /M và RFC 3161 Timestamp

Tiêu chí	/M	Timestamp (RFC 3161)
Nguồn cấp	Phần mềm ký	Tổ chức TSA
Dữ liệu	Text	Token nhị phân có chữ ký TSA
Giá trị pháp lý	Không có	Có
Mục đích	Hiển thị thời gian ký	Chứng minh thời điểm ký thực tế

IV. RỦI RO BẢO MẬT TRONG CHỮ KÝ PDF

1. Rủi ro kỹ thuật

- **Giả mạo /M**: ứng dụng có thể ghi sai thời gian.
- **Thay đổi nội dung sau khi ký**: nếu không bảo vệ đúng ByteRange hoặc bị ghi đè update.
- **Chữ ký không có timestamp**: mất giá trị khi chứng chỉ hết hạn.
- **Thiếu LTV (DSS)**: không thể xác minh khi CA/OCSP bị thu hồi.
- **Tấn công "incremental update forgery"**: thêm nội dung giả phía sau vùng được ký.

2. Rủi ro xác thực

- Chữ ký hợp lệ về mặt kỹ thuật nhưng chứng chỉ không còn tin cậy (revoked/expired).
- Timestamp được cấp sau thời điểm chứng chỉ hết hạn.
- Phần mềm đọc PDF không kiểm tra đúng PAdES dẫn đến "false valid".

3. Biện pháp khắc phục

- Áp dụng chuẩn **PAdES-BES / PAdES-LTV** để lưu đầy đủ dữ liệu xác minh.
- Sử dụng **timestamp RFC 3161** được cấp bởi TSA.
- Kiểm tra toàn bộ **OCSP/CRL chain** trong DSS khi xác minh.
- Duy trì **bản sao gốc** và **bản đã ký** tách biệt.
- Sử dụng công cụ tin cậy như **Adobe Acrobat**, **DigiDoc4**, hoặc **PyPDF + OpenSSL**.

V. KẾT LUẬN

Chữ ký số trong PDF là nền tảng quan trọng cho **tài liệu điện tử an toàn**. Việc hiểu rõ **cấu trúc PDF**, **các vùng chữ ký**, và **cơ chế timestamp** là điều kiện bắt buộc để hiện thực và xác thực đúng chuẩn **PAdES**.

Khi triển khai thực tế:

- **PDF 1.7/PAdES** nên được chọn làm chuẩn tham chiếu.
- **Công cụ khuyến nghị**: iText7 (Java/.NET), PyPDF + OpenSSL (Python).
- Kết hợp **RFC 3161 timestamp + DSS** để đạt mức xác thực lâu dài (LTV).

Phụ lục: Chuẩn tham chiếu

Chuẩn	Tên đầy đủ	Cơ quan
ISO 32000-1:2008	PDF 1.7 Specification	ISO
ISO 32000-2:2020	PDF 2.0 Specification	ISO
ETSI EN 319 142-1	PAdES – Baseline Profile	ETSI
RFC 3161	Internet X.509 Public Key Infrastructure Time-Stamp Protocol	IETF