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

Chủ đề: Chữ ký số trong file PDF

Giảng viên: Đỗ Duy Cốp

Hạn nộp: Trước 2025-10-31 23:59:59

Sinh viên thực hiện: Trần Thị Thu Hà

MSSV: K225480106009

Lớp: K58KTP

1. Cấu trúc PDF liên quan chữ ký.

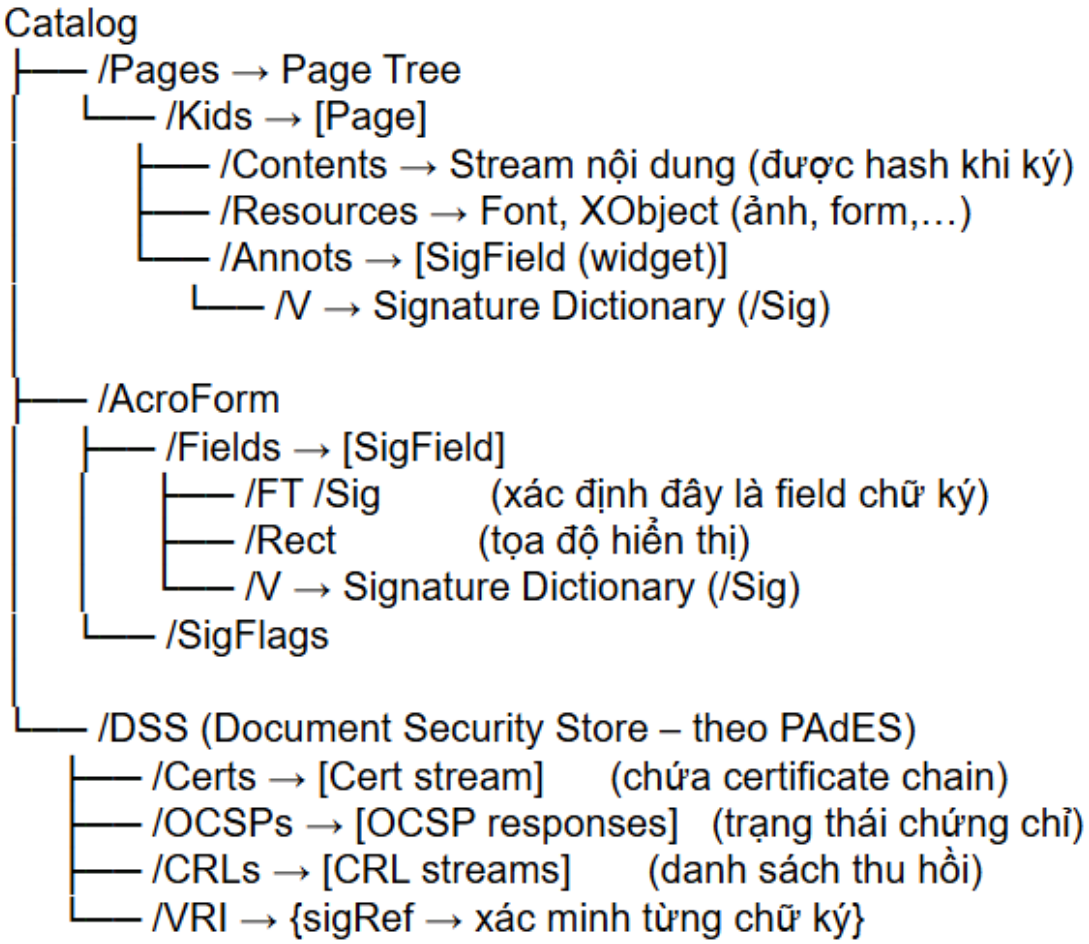
Mọi thành phần của file PDF được tổ chức dưới dạng object và liên kết với nhau bằng object reference. Chữ ký số trong PDF được lưu thông qua AcroForm, trong đó có Signature Field (Widget) để hiển thị chữ ký, Signature Dictionary(/Sig) để chứa thông tin chữ ký và dữ liệu mã hóa. Khi người dùng ký tài liệu, phần mềm PDF sẽ thực hiện incremental update là ghi thêm một lớp mới vào cuối file và giữ nguyên nội dung cũ để bảo đảm tính toàn vẹn, file còn có thêm DSS (Document Security Store) để lưu chứng thư phục vụ xác minh dài hạn.

Object refs quan trọng và vai trò:

Object Refs	**Vai trò**
Catalog	Là Root object chứa /AcroForm và /Pages. Quản lý toàn bộ cấu trúc file PDF, là điểm khởi đầu để truy suất đến SigField.
Pages tree	Cấu trúc cây trang, được tham chiếu từ Catalog. Tổ chức các Page object và Annots, trong đó có SigField widget.

Page object	Mô tả từng trang. Chứa /Contents (dòng lệnh vẽ nội dung trang) và /Resources (font, hình ảnh). Nếu có chữ ký hiển thị, vị trí của widget chữ ký được định nghĩa tại đây.
Resources	Dictionary chứa tài nguyên (font, XObject). Giúp render phần hiển thị chữ ký (ví dụ hình ảnh, text, logo của chữ ký).
Content streams	Stream chứa nội dung trang (text, vector, hình ảnh). Khi xác minh, phần này thuộc vùng /ByteRange để bảo vệ tính toàn vẹn nội dung.
XObject	External object (như hình ảnh). Thường được dùng để hiển thị hình ảnh chữ ký tay hoặc logo của người ký.
AcroForm	Form container nằm trong Catalog. Quản lý danh sách /Fields, trong đó có trường chữ ký (SigField).
Signature field (wifdget)	Annotation trên trang, biểu diễn khung chữ ký. Thuộc loại /Widget, có /FT /Sig. Xác định vị trí UI của chữ ký trên trang.
Signature dictionary	Lưu trữ dữ liệu chữ ký: /Contents (dạng PKCS#7), /ByteRange, /M (thời gian ký), /Name, /Reason, /Location. Là nơi thực hiện xác thực và validate hash.
ByteRange	Mảng [start1 length1 start2 length2]. Xác định phạm vi byte được ký (ngoại trừ vùng /Contents chứa chữ ký). Đảm bảo phát hiện thay đổi trái phép
Contents	Dữ liệu chữ ký số (PKCS#7/CMS) chứa hash, chứng chỉ, timestamp. Được giải mã để lấy thông tin người ký và xác minh tính toàn vẹn.
Incremental updates	Mỗi lần ký mới, PDF không bị ghi đè mà thêm phần "revision" mới ở cuối file. Giúp hỗ trợ ký nhiều lần (multi-signature) mà không phá vỡ chữ ký trước.
DSS	Document Security Store (chuẩn PAdES). Nằm trong Catalog, lưu trữ chứng chỉ (/Certs), phản hồi OCSP (/OCSPs), và danh sách CRL (/CRLs). Hỗ trợ xác minh chữ ký dài hạn (LTV)

Sơ đồ objects:



2. Nơi lưu thời gian ký.

Tất cả các vị trí có thể lưu thông tin thời gian:

Vị trí	**Miêu tả**	**Kiểu dữ liệu**	**Vai trò/Giá trị pháp lý**
/M (Modification time) trong Signature Dictionary	Thuộc tính /M nằm trong object /Sig	Chuỗi text định dạng pdf date	Là thời gian người ký khai báo tại thời điểm ký, không được bảo đảm bởi bên thứ ba nên không có giá trị pháp lý, chỉ mang tính thông tin

Timestamp Token (RFC 3161) trong PKCS#7	Là attribute timeStamp Token bên trong chữ ký PKCS#7 (CMS) trong trường /content	Mã hóa dạng binary DER (theo RFC 3161)	Là bằng chứng thời gian được cung cấp bởi Time Stamping Authority (TSA). Có giá trị pháp lý vì được ký bằng chứng thư của TSA.
Document TimeStamp Object (PAdES)	Được thêm như signature riêng biệt trong file PDF, dùng filter /ETSI.RFC3161	Một loại Signature Dictionary đặc biệt	Được dùng để đóng dấu thời gian cho toàn bộ tài liệu (document-level timestamp). Có giá trị pháp lý
DSS (Document Security Store)	Lưu chữ dữ liệu hỗ trợ xác minh lâu dài: chứng thư, OCSP, CRL và timestamp token	Dictionary (theo chuẩn PadES Part 4)	Không trực tiếp ký nhưng giữ lại timestamp token để xác thực trong tương lại khi TSA hoặc CA không còn hoạt động

Sự khác biệt giữa thông tin thời gian /M và timestap RFC3161:

- **/M (Metadata time):** nơi lưu là trong Signature Dictionary (/Sig), dạng dữ liệu là chuỗi text (PDF date string) do phần mềm ký chèn vào, cách tạo là người ký hoặc phần mềm tự ghi, không đảm bảo và có thể giả mạo, không có giá trị pháp lý, công dụng chính là hiển thị thời gian ký.
- **Timestamp RFC3161:** nơi lưu là trong PKCS#7 (CMS) → SignedAttributes → timeStampToken, dạng dữ liệu là binary DER-encoded (ASN.1 structure) do TSA (time stamping authority) phát hành, cách tạo là gửi hash đến TSA và TSA ký rồi trả về token, được ký bằng khóa riêng của TSA nên được xác thực, có giá trị pháp

lý, chống giả mạo (bằng chứng thời gian độc lập), công dụng chính là chứng minh tài liệu tồn tại trước một thời điểm cụ thể.

3. Các rủi ro bảo mật trong chữ ký số PDF

- **Padding Oracle Attack (tấn công đệm RSA):** Do lỗ hổng trong cơ chế RSA PKCS#1 v1.5, kẻ tấn công có thể khai thác phản hồi lỗi để đoán khóa, gây ra hậu quả có thể giả mạo chữ ký hoặc suy ra khóa riêng. Biện pháp dùng RSA-PSS, cập nhật thư viện crypto, không hiển thị lỗi chi tiết.
- **Lộ private key (key leak):** do khóa riêng bị sao chép hoặc đánh cắp (file .pem không mã hóa, malware,…), gây ra hậu quả kẻ tấn công ký giả mạo mọi tài liệu, phòng chống bằng cách lưu khóa trong HSM/USB token, mã hóa bằng passpphare và giới hạn quyền truy cập.
- **Replay Attack:** Tái sử dụng blob#7 hợp lệ trên tài liệu khác nếu không kiểm tra ByteRange gây ra hậu quả tài liệu vẫn có chữ ký hợp lệ. Phòng chống: kiểm tra /ByteRange và hash, dùng **RFC3161 timestamp**, ghi log.
- **Signature Wrapping:** Kẻ tấn công có thể di chuyển hoặc nhân bản object chữ ký sang vị trí khác trong PDF. Gây ra hậu quả nội dung hiển thị khác với phần thực sự được ký. Phòng chống bằng cách kiểm tra vị trí SigDict, xác minh ByteRange, khóa nội dung sau ký.
- **Incremetal Update Injection:** Kẻ tấn công có thể thêm update mới sau chữ ký (append) để thay đổi nội dung hiển thị, gây ra hậu quả là hai tài liệu khác nhau có cùng hash → chữ ký "hợp lệ" cho cả hai. Phòng chống bằng cách dùng **SHA-256 / SHA-512**, từ chối SHA-1/MD5.
- **Chứng thư kiểm tra không đầy đủ (CRL/OCPS thiếu):** Không kiểm tra trạng thái thu hồi hoặc EKU của certificate. Hậu quả là chấp nhận cert đã bị thu hồi/hết hạn. Phòng chống bằng cách kiểm tra **OCSP/CRL**, xác minh EKU, lưu kết quả trong **DSS (PAdES-LTV)**.

- **Timestamp giải hoặc thiếu:** Chỉ có /M (local time) hoặc không có RFC3161 từ TSA, hậu quả là không xác minh được thời điểm ký và dễ lách hạn cert. Phòng chống bằng cách sử dụng RFC3161 timestamp token từ TSA tin cậy và lưu token trong PKCS#7/DSS.

4. Kết luận:

Qua bài tập này em đã nắm được cấu trúc của một file pdf có chữ ký số gồm các thành phần chính *Catalog, AcroForm, Signature Field và Signature Dictionary (/Sig)*. Việc ký chữ ký số vào file pdf giúp đảm bảo tính toàn vẹn và xác thực của tài liệu.

Nắm được quy trình tạo và xác thực chữ ký số theo chuẩn RSA – SHA256/PAdES, cách lưu thời gian ký qua /M và timestamp RFC3161 và vai trò của DSS trong việc xác minh lâu dài giúp đảm bảo tính toàn vẹn và xác thực của tài liệu.

Nhận biết được các rủi ro bảo mật phổ biến như lộ khóa riêng, tấn công padding oracle, replay attack hay chỉnh sửa incremental update. Và thông qua đó, em hiểu rõ hơn tầm quan trọng của việc bảo vệ khóa ký, sử dụng chuẩn an toàn (RSA-PSS, SHA-256), và áp dụng timestamp, OCSP/CRL để tăng tính pháp lý.