|  |  |
| --- | --- |
| Họ tên: | Hoàng Kim Ngọc |
| MSSV: | K225480106053 |
| Lớp : | K58KTP |

**BÀI TẬP MÔN AN TOÀN VÀ BẢO MẬT THÔNG TIN**

***(Bài tập số 2)***

*Chủ đề : Chữ ký số trong file PDF*

*Giảng viên : Đỗ Duy Cốp*

*Thời điểm giao: 2025-10-24*

*Hạn nộp : 2025-10-31*

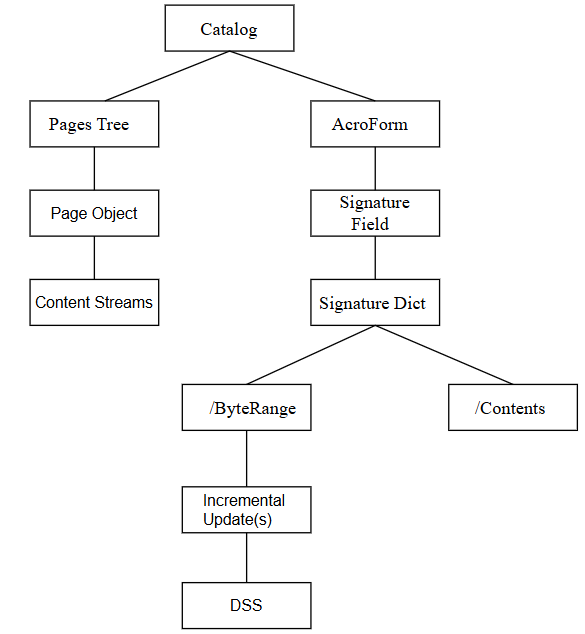
1. **Cấu trúc PDF:**

**-** PDF (*Portable Document Format*) được cấu trúc từ các **đối tượng (objects)** có thể tham chiếu lẫn nhau, gồm bốn phần chính: **Header, Body, Cross-reference table**, và **Trailer.**  
 + Trong đó, **Catalog** là đối tượng gốc liên kết đến **Pages tree** – nơi quản lý các **Page object** (mỗi trang của tài liệu). Mỗi trang chứa **Resources** (phông, hình, màu) và **Content streams** (nội dung hiển thị).

+ Các hình ảnh hoặc mẫu đồ họa chung được lưu trong **XObject**. Nếu tài liệu có biểu mẫu hoặc chữ ký số, phần **AcroForm** sẽ quản lý các trường, trong đó **Signature field (widget)** là vùng chứa chữ ký. Trường này liên kết đến **Signature dictionary (/Sig)** – nơi lưu thông tin người ký, thời gian, thuật toán và dữ liệu chữ ký.

+ Bên trong **/Sig**, hai mục quan trọng là **/ByteRange** (vùng dữ liệu được băm) và **/Contents** (chứa chữ ký mã hóa). PDF dùng cơ chế **incremental updates** để thêm chữ ký mà không sửa dữ liệu gốc. Chuẩn **PAdES** mở rộng thêm **DSS** để lưu chứng chỉ và thông tin xác thực lâu dài.

|  |  |
| --- | --- |
| **Objects** | **Vai trò** |
| **Catalog** | Đối tượng gốc, liên kết đến AcroForm chứa các trường chữ ký. |
| **AcroForm** | Quản lý các form field, bao gồm Signature field. |
| **Signature Field (Widget)** | Hiển thị vùng chữ ký, tham chiếu đến Signature dictionary. |
| **Signature Dictionary (/Sig)** | Lưu thông tin chữ ký (người ký, thời gian, thuật toán, hash). |
| **/ByteRange** | Xác định vùng dữ liệu được băm để đảm bảo tính toàn vẹn. |
| **/Contents** | Chứa dữ liệu chữ ký mã hóa (PKCS#7/CMS). |
| **Page Object** | Chứa vị trí hiển thị của vùng chữ ký trên trang. |
| **Incremental Updates** | Lưu phần cập nhật khi thêm chữ ký, không làm thay đổi dữ liệu gốc. |
| **DSS (Document Security Store)** | Lưu chứng chỉ, CRL/OCSP, timestamp để xác thực chữ ký lâu dài (PAdES). |



*Hình 1. Sơ đồ quan hệ các objects trong PDF có chữ ký số*

1. **Thời gian ký được lưu ở đâu?**

Thông tin thời gian trong PDF có chữ ký số được lưu ở hai khu vực chính, phân biệt bởi độ tin cậy và giá trị pháp lý:

- Vị trí cục bộ (Không đáng tin cậy): **/M (Modification Time)**: Là thời gian lấy từ đồng hồ máy tính cục bộ của người ký. Không có giá trị pháp lý vì dễ bị thay đổi.

- Vị trí kỹ thuật (Đáng tin cậy):

+ **Timestamp Token (RFC 3161)**: Là dấu thời gian được ký số và cấp bởi Máy chủ Dấu thời gian (TSA) độc lập. Đây là bằng chứng thời gian có giá trị pháp lý cao, được nhúng bên trong cấu trúc chữ ký.

+ **Document Timestamp (PAdES)**: Một loại chữ ký dựa trên TSA để chứng thực thời điểm tồn tại của toàn bộ tài liệu.

+ **DSS (Document Security Store)**: Lưu trữ các dấu thời gian liên quan đến việc thu thập dữ liệu xác minh, hỗ trợ Xác thực lâu dài (LTV) của chữ ký.

|  |  |  |
| --- | --- | --- |
| **Đặc điểm** | **/M (Modification Time)** | **Timestamp Token (RFC 3161)** |
| **Nguồn Gốc** | **Máy tính cục bộ** của người ký. | **Máy chủ Dấu thời gian (TSA)** độc lập, tin cậy. |
| **Giá trị Pháp lý** | **Không có giá trị pháp lý** như bằng chứng về thời gian ký. | **Có giá trị pháp lý cao** (Trusted Timestamp). |
| **Mục đích** | Ghi lại **thời gian bắt đầu quá trình ký** (trên máy tính của người ký). | Cung cấp bằng chứng không thể chối bỏ về **sự tồn tại của dữ liệu** (chữ ký và tài liệu) tại một thời điểm cụ thể. |
| **Bảo mật** | **Dễ dàng bị giả mạo** hoặc thay đổi bằng cách điều chỉnh đồng hồ hệ thống. | **Được bảo vệ bằng chữ ký số** của TSA, đảm bảo tính toàn vẹn và nguồn gốc thời gian. |
| **Vị trí Lưu trữ** | Một trường dữ liệu dạng văn bản **/M** nằm trong **Signature Dictionary** của tài liệu PDF. | Một **thuộc tính được mã hóa (attribute)** nằm bên trong cấu trúc kỹ thuật của chữ ký số (thường là PKCS#7/CAdES). |
| **Vai trò với LTV** | Không đóng góp vào **Xác thực lâu dài (LTV)**. | Là thành phần **then chốt** để duy trì giá trị pháp lý của chữ ký sau khi chứng thư số gốc hết hạn (vì nó chứng minh chữ ký được tạo ra khi chứng thư còn hiệu lực). |

1. **Các bước tạo và lưu chữ ký trong PDF (đã có private RSA)**

*Bước 1:Chuẩn bị file PDF gốc.*

- File original.pdf

*Bước 2: Tạo Signature field (AcroForm), reserve vùng /Contents (8192 bytes).*

-Thêm AcroForm Signature Field vào PDF (thường tên là /Sig1).

-Trong field, tạo mục /Contents — dành sẵn vùng trống (placeholder), thường là 8192 bytes để chứa dữ liệu chữ ký PKCS#7 sau này.

*Bước 3: Tính toán ByteRange*

- Xác định 4 giá trị ByteRange:  
***[0, offset\_start\_of\_Contents, offset\_end\_of\_Contents, remaining\_length]***

- Khi tính hash, bỏ qua vùng /Contents.

*Bước 4: Tính hash*

- Dùng SHA-256 (hoặc SHA-512) để băm dữ liệu trong ByteRange.

- Kết quả là messageDigest.

*Bước 5: Tạo chữ ký PKCS#7 / CMS (detached)*

- Dạng chữ ký: CMS detached (PKCS#7) hoặc CAdES.

- Bao gồm:

+ signedAttrs:

* contentType = id-data
* messageDigest (giá trị hash ở bước 4)
* signingTime

+ digestAlgorithm = SHA-256

+ signatureAlgorithm = rsaEncryption (PKCS#1 v1.5)

+ certificates: toàn bộ chuỗi chứng thư.

+ (Tùy chọn): thêm timestamp (RFC3161) trong unsignedAttrs để tạo CAdES-T.

- Ký phần signedAttrs bằng private key RSA.

*Bước 6: Ghi chữ ký vào PDF*

- Chèn dữ liệu chữ ký PKCS#7 (dạng DER hex string) vào vùng /Contents.

- Nếu không đủ 8192 bytes, pad thêm 0x00 để khớp độ dài.

- Cập nhật lại /ByteRange.

*Bước 7: Ghi Incremental Update*

- Không ghi đè toàn bộ file — chỉ append phần ký mới để giữ nguyên lịch sử file PDF.

*Bước 8: Bổ sung LTV (Long-Term Validation)*

- Thêm mục /DSS vào PDF Catalog, gồm:

+ /Certs – chứng thư,

+ /OCSPs – phản hồi OCSP,

+ /CRLs – danh sách thu hồi,

+ /VRI – liên kết xác thực theo từng chữ ký.

- Nhằm đảm bảo file có thể xác minh hợp lệ về lâu dài (LTV-enabled).

1. **Các bước xác thực chữ ký trên PDF đã ký**

*Bước 1: Đọc Signature dictionary*

- Signature trong PDF được lưu trong Signature dictionary, chứa các trường quan trọng:

+ /Contents: dữ liệu chữ ký (PKCS#7)

+ /ByteRange: vùng dữ liệu PDF được ký

- Mục tiêu: xác định phần dữ liệu nào được ký và tách chữ ký ra.

*Bước 2: Tách PKCS#7, kiểm tra định dạng*

- Chữ ký PDF thường là PKCS#7/CMS.

- Kiểm tra:

+ Chữ ký hợp lệ (ASN.1 DER format)

+ Có chứa chứng chỉ của signer.

*Bước 3: Tính hash và so sánh messageDigest*

- Dựa vào /ByteRange, tính hash của dữ liệu PDF đã ký.

- So sánh với messageDigest trong PKCS#7.

- Nếu trùng → dữ liệu chưa bị chỉnh sửa kể từ lúc ký.

*Bước 4: Verify signature bằng public key*

- Dùng public key trong certificate của signer để xác thực chữ ký:

+ Xác minh chữ ký khớp với messageDigest.

+ Nếu khớp → chữ ký xác thực người ký.

*Bước 5: Kiểm tra chain chứng chỉ*

- Kiểm tra xem certificate signer → root CA có hợp lệ không:

+ Root CA phải nằm trong trusted CA store.

+ Kiểm tra các intermediate CA.

*Bước 6: Kiểm tra trạng thái chứng chỉ (OCSP/CRL)*

- Đảm bảo certificate chưa bị thu hồi.

- Sử dụng:

+ OCSP (Online Certificate Status Protocol)

+ CRL (Certificate Revocation List)

*Bước 7: Kiểm tra timestamp token*

- Nếu chữ ký có timestamp, verify timestamp token để xác nhận thời điểm ký.

- Timestamp giúp chứng minh chữ ký hợp lệ ngay cả khi certificate sau này hết hạn.

*Bước 8: Kiểm tra incremental update*

- PDF có thể bị incrementally updated.

- Kiểm tra xem dữ liệu sau chữ ký có bị sửa đổi hay không.

- Nếu bị chỉnh sửa → chữ ký không còn hợp lệ.