**BÁO CÁO BÀI TẬP VỀ NHÀ**

**MÔN: AN TOÀN VÀ BẢO MẬT THÔNG TIN**

**Chủ đề: Chữ ký số trong file PDF**

Sinh viên thực hiện: Triệu Trà My

MSSV:K225480106102

Lớp:K58KTP

Nội dung: Tệp PDF dùng để minh họa quy trình ký số theo 8 bước; báo cáo tóm tắt cấu trúc PDF liên quan chữ ký, nơi lưu thời gian ký, phân tích rủi ro và biện pháp giảm thiểu.

1. MỤC TIÊU BÀI TẬP

- Trình bày, giải thích và minh họa bằng ví dụ cách thức chữ ký số được lưu và bảo vệ tròng file PDF.

+ Vị trí và cấu trúc lưu chữ ký trong PDF (AcroForm, Signature Field, Signature Dictionary).

+ Cách lưu thời gian ký ( khác nhau giữa /M và timestamp RFC -3161).

+ Các rủi ro bảo mật phổ biến và biện pháp giảm thiểu.

**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)**

* Các objects chính trong PDF:

|  |  |
| --- | --- |
| * **1. Catalog Object** | Là “gốc” (Root) của tài liệu PDF. Từ đây có thể truy cập tới toàn bộ cây trang (Pages tree) và đến AcroForm – nơi chứa các trường form (Form fields). Trong trường hợp có chữ ký, Catalog có một key /AcroForm trỏ đến AcroForm dictionary. |

|  |  |
| --- | --- |
| **2. Pages Tree** | Là cấu trúc dạng cây mô tả toàn bộ các trang trong tài liệu. Mỗi Page Object nằm trong cây này. |

|  |  |
| --- | --- |
| **3. Page Object** | Đại diện cho từng trang cụ thể trong file PDF. Nó chứa các tham chiếu tới nội dung hiển thị (Content Stream), tài nguyên (Resources), và các annotation (bao gồm chữ ký). |

|  |  |
| --- | --- |
| **4. Resources** | Lưu thông tin tài nguyên dùng cho trang: phông chữ, hình ảnh, màu sắc, XObject (đối tượng nhúng),… |

|  |  |
| --- | --- |
| **5. Content Streams** | Là nơi chứa các lệnh vẽ (văn bản, hình khối, hình ảnh) – nói cách khác là “nội dung nhìn thấy được” của trang. |

|  |  |
| --- | --- |
| **6. XObject** | Là các đối tượng nhúng, ví dụ hình ảnh hoặc mẫu template. Không liên quan trực tiếp đến chữ ký, nhưng có thể tồn tại trong cùng trang. |

|  |  |
| --- | --- |
| **7. AcroForm Dictionary** | Là thành phần quản lý toàn bộ các **form fields** trong PDF (ví dụ: textbox, checkbox, signature field...). Khi có chữ ký, phần này sẽ chứa một danh sách các Signature Fields. |

|  |  |
| --- | --- |
| **8. Signature Field (Widget Annotation)** | Là trường chữ ký hiển thị trên trang. Nó có vị trí (tọa độ x, y), kích thước (width, height), và tham chiếu đến **Signature Dictionary** (object chứa dữ liệu chữ ký thật). |

|  |  |
| --- | --- |
| **9. Signature Dictionary (/Sig)** | Đây là thành phần cốt lõi của chữ ký PDF. Nó lưu dữ liệu PKCS#7 (chứa chữ ký số RSA hoặc ECDSA) cùng với thông tin thời gian, tên người ký, thuật toán băm, và phạm vi dữ liệu được ký. Các key quan trọng trong dictionary này gồm: |

|  |  |
| --- | --- |
|  | - /Filter : Tên bộ lọc xử lý chữ ký (thường là /Adobe.PPKLite). |

|  |  |
| --- | --- |
|  | - /SubFilter : Định dạng chữ ký (ví dụ /adbe.pkcs7.detached hoặc /ETSI.CAdES.detached). |

|  |  |
| --- | --- |
|  | - /ByteRange : Xác định vùng byte trong file được hash và ký (loại trừ vùng /Contents). |

|  |  |
| --- | --- |
|  | - /Contents : Chứa chữ ký PKCS#7/CMS dưới dạng chuỗi hex hoặc binary. |

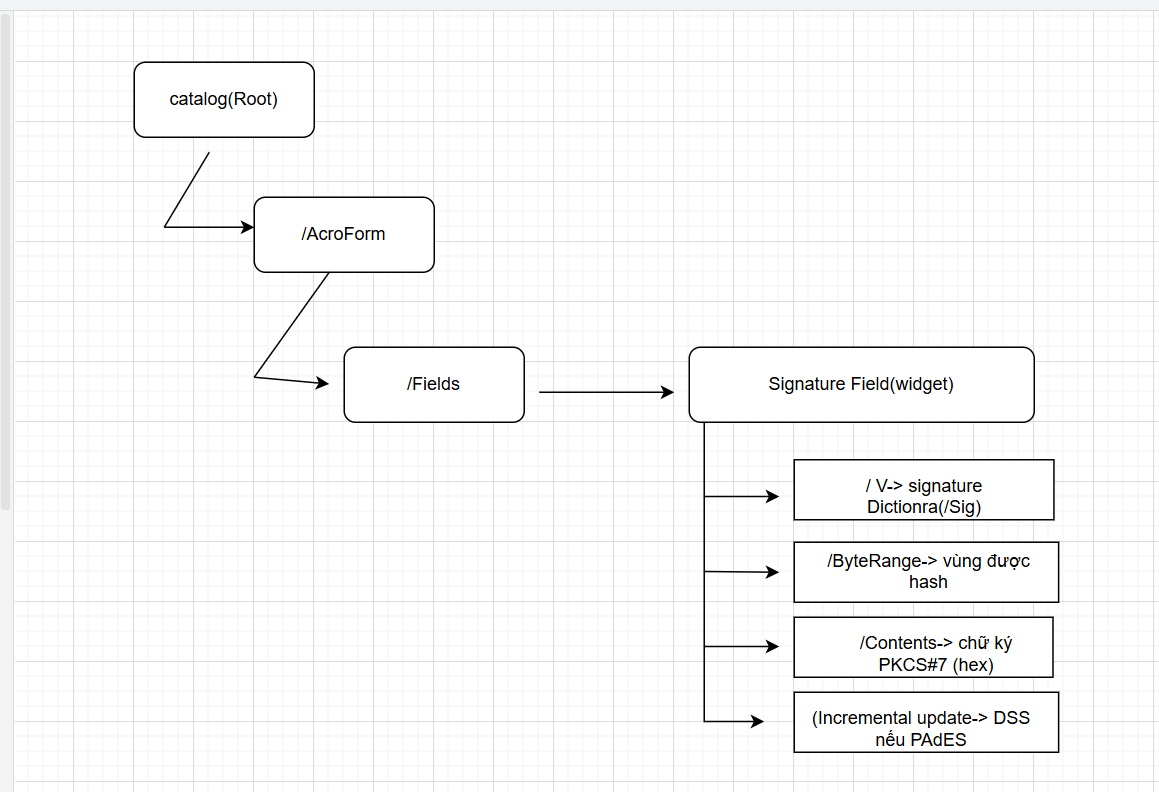
|  |  |
| --- | --- |
| **10. /ByteRange** | Là mảng gồm 4 giá trị (offset, length, offset, length). Nó cho biết phần nào của file được dùng để tính hash. Phần /Contents (chứa chữ ký) bị loại ra khỏi vùng này để tránh thay đổi dữ liệu sau khi ký. |

|  |  |
| --- | --- |
| **11. /Contents** | Là vùng chứa dữ liệu chữ ký số thực tế – thường là chuỗi PKCS#7 (chuẩn CMS) đã được mã hóa bằng khóa riêng của người ký. |

|  |  |
| --- | --- |
| **12. Incremental Update** | Khi ký, PDF không bị ghi đè mà chỉ thêm phần mới vào cuối file. Phần mới này chứa chữ ký và cập nhật lại bảng xref và trailer. Điều này cho phép xác minh toàn vẹn vì các phần cũ vẫn giữ nguyên. |

|  |  |
| --- | --- |
| **13. DSS (Document Security Store)** | Thành phần mở rộng theo chuẩn **PAdES (ETSI EN 319 142)**. Nó lưu trữ các dữ liệu cần thiết cho việc xác minh chữ ký lâu dài (Long-Term Validation – LTV): danh sách chứng chỉ, CRL, OCSP và timestamp. |
|  |  |

* Sơ đồ quan hệ object



**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 RFC**

Trong tài liệu PDF có chữ ký số, thông tin thời gian ký có thể được lưu tại nhiều vị trí khác nhau, tùy thuộc vào chuẩn ký và mức độ xác thực. Các vị trí lưu thời gian gồm:

1. **Giá trị /M trong Signature Dictionary**
   * Đây là thuộc tính nằm trong từ điển chữ ký (/Sig) của PDF.
   * Dữ liệu thời gian được lưu dưới dạng chuỗi văn bản theo định dạng:

 Thời gian này do **phần mềm ký** tự ghi vào, không có ràng buộc xác thực.

 Vì vậy **không có giá trị pháp lý**, do người ký hoặc thiết bị có thể thay đổi thời gian hệ thống máy tính trước khi ký gian hệ thống máy tính trước khi ký.

**2.** **Timestamp token (RFC 3161) bên trong chữ ký PKCS#7/CMS**

* Token thời gian được nhúng bên trong dữ liệu chữ ký /Contents.
* Token này được tạo bởi **Timestamp Authority (TSA)** — một tổ chức tin cậy có chứng thư số riêng.
* Timestamp chứa:
  + Hash của dữ liệu được ký
  + Thời gian ký
  + Chữ ký số của TSA xác nhận thời gian
* Đây là bằng chứng hợp pháp về thời điểm chữ ký được tạo ra.

**3.** **Document Timestamp (theo chuẩn PAdES)**

* PDF có thể chứa một chữ ký đặc biệt dựa trên timestamp, không gắn với người ký mà gắn với tài liệu.
* Được dùng để đóng dấu thời gian cho cả tài liệu sau khi chữ ký được tạo.
* Không yêu cầu private key của người ký, chỉ cần timestamp của TSA.

**4.** **DSS – Document Security Store**

* Lưu ở phần incremental update cuối cùng của PDF.
* DSS chứa các dữ liệu hỗ trợ xác minh chữ ký lâu dài:
  + Certificate chain
  + CRL và OCSP
  + Timestamp token (nếu có)
* Hỗ trợ chế độ **LTV (Long Term Validation)**, đảm bảo tài liệu vẫn xác thực được dù chứng thư số hết hạn.

**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

**5. Rủi ro chính và biện pháp giảm thiểu**

|  |  |  |
| --- | --- | --- |
| **Rủi ro** | Mô tả | Cách phát hiện & Biện pháp xử lý |
| **Sửa đổi nội dung sau khi ký (Tampering)** | Kẻ tấn công có thể sửa nội dung file PDF, thay đổi dữ liệu bên ngoài vùng ký hoặc chỉnh lại giá trị /ByteRange để vô hiệu hoá liên kết giữa nội dung và chữ ký. Kết quả: chữ ký không còn khớp với dữ liệu ban đầu. | **Phát hiện:** Trong bước xác minh chữ ký, trình verify tính lại hash trên vùng /ByteRange và so sánh với messageDigest trong PKCS#7/CMS. Nếu không trùng → chữ ký **không hợp lệ**.  **Biện pháp:** luôn thực hiện ký theo chuẩn incremental update của PDF; trình xác minh phải kiểm tra ByteRange, modification level và phát hiện thay đổi sau khi ký. |
| **Lợi dụng cơ chế Incremental Update / Replay Attack** | PDF cho phép append dữ liệu (incremental update). Kẻ tấn công có thể chèn thêm Signature Dictionary giả hoặc ghi đè phiên bản mới lên nội dung cũ khiến người dùng tưởng tài liệu vẫn hợp lệ. | Phát hiện: phân tích lịch sử incremental, kiểm tra xref/trailer và thứ tự signature.  Biện pháp: mỗi lần ký mới phải kèm timestamp từ TSA; lưu metadata, timestamp, OCSP/CRL trong DSS (Document Security Store) để bảo đảm truy vết lịch sử tài liệu. |
| **Không kiểm tra thu hồi chứng chỉ (OCSP / CRL)** | Nếu chứng chỉ của người ký đã bị thu hồi nhưng trình xác minh không kiểm tra OCSP/CRL thì chữ ký vẫn có thể bị chấp nhận dù không hợp lệ về mặt pháp lý. | Phát hiện: trình verify kiểm tra trạng thái chứng chỉ qua OCSP/CRL.  Biện pháp: nhúng OCSP/CRL vào DSS trong lúc ký → hỗ trợ LTV (Long-Term Validation) để tài liệu vẫn xác minh được dù CA/OCSP hết hạn trong tương lai. |
| **Lộ khóa riêng (Private Key Exposure) / quản trị khóa yếu** | Nếu khoá riêng bị đánh cắp hoặc người dùng lưu file chứa khóa (.pfx/.pem) không an toàn, kẻ tấn công có thể ký thay chủ sở hữu mà không bị phát hiện. | Phát hiện: khó phát hiện từ phía hệ thống xác minh (vì chữ ký vẫn hợp lệ).  Biện pháp: lưu trữ khóa trong HSM / Smartcard / eToken, bật password bảo vệ khóa, quy trình quản lý khóa an toàn, kiểm soát truy cập nghiêm ngặt. |

**6. Khuyến nghị kỹ thuật**

- Dùng SHA-256 hoặc mạnh hơn cho message digest.

- Dùng RSA 2048+ hoặc RSA-PSS(khuyến nghị) cho chữ ký và server TSA đáng tin cậy cho timestamp RFC -3161.

- Thực hiện LTV (PAdES-LTV) bằng cách nhúng chứng thư, OCSP/CRL và timestamp token và DSS.

- Kiểm tra modification level và đảm bảo trình verify báo rõ ràng khi có incremental updates.

**7. Minh họa File đính kèm**

Trong bài nộp kèm các file mẫu :

- original.pdf -file gốc.

- signed.pdf- file sau khi đã ký ( chứa/Contents PKCS#7 và ByteRange hợp lệ).

- tampered.pdf - phiên bản đã bị chỉnh sửa ngoài vùng được ký ( dùng để minh chứng verify thất bại).

**8.Kết luận**

Bài tập này giúp hiểu rõ cơ chế lưu và xác minh chữ ký trong PDF thông qua các thành phần /ByteRange ,/Contents và incremental update. Trường /M chỉ lưu thời gian hiện thị không, không có giá trị pháp lý, trong khi timestamp RFC-3161 trong PKCS#7 mới chứng minh được thời điểm ký thực tế. Để đảm bảo tính pháp lý và xác minh lâu dài (LTV), cần kết hợp PKCS#7 + timestamp từ TSA và nhúng dữ liệu OCSP/CRL vào DSS theo chuẩn PAdES.