1) Cấu trúc PDF liên quan chữ ký số (Nghiên cứu)

I. Mô tả ngắn gọn các thành phần chính

- Catalog (Root object): Là đối tượng gốc của file PDF. Catalog tham chiếu đến các phần quan trọng khác như cây trang (/Pages) và biểu mẫu (/AcroForm). Đây là điểm khởi đầu khi phần mềm đọc cấu trúc PDF, và cũng là nơi tham chiếu đến các trường chữ ký (signature fields) nếu có.

- Pages tree (/Pages): Là cấu trúc phân cấp quản lý toàn bộ các trang trong tài liệu PDF. Mỗi node có thể là một tập hợp con (/Pages) hoặc một trang đơn lẻ (/Page).

- Page object: Đại diện cho một trang cụ thể trong PDF. Page object chứa nội dung hiển thị của trang, liên kết đến /Resources (tài nguyên như phông chữ, ảnh, mẫu vẽ) và /Contents (luồng dữ liệu mô tả nội dung trang). Nó cũng có thể chứa danh sách chú thích (/Annots), bao gồm vùng hiển thị chữ ký.

- Resources: Gồm phông chữ, ảnh, màu, mẫu vẽ, và các XObject mà trang sử dụng. Đây là "tài nguyên" để vẽ nội dung trong PDF.

- Content streams (/Contents): Là phần dữ liệu dạng lệnh PDF (PostScript-like), mô tả nội dung hiển thị thực tế: văn bản, hình ảnh, đường vẽ, hoặc hình khối.

- XObject: Là đối tượng có thể tái sử dụng trong nhiều trang khác nhau, ví dụ ảnh bitmap, hình vector, hoặc Form XObject. Khi chèn ảnh chữ ký tay vào PDF, hình ảnh đó thường được lưu như một XObject.

- AcroForm: Là phần biểu mẫu của PDF, chứa các trường dữ liệu người dùng có thể điền, ví dụ: text field, checkbox, radio button, hoặc signature field. Đây là nơi chứa danh sách tất cả các trường chữ ký của tài liệu.

- Signature field (Widget Annotation): Là một trường (field)

đặc biệt trong AcroForm, xác định vị trí hiển thị của chữ kýtrên trang. Signature field liên kết giữa vùng hiển thị chữ ký (ô trên trang PDF) và dữ liệu chữ ký số thực tế.

- Signature dictionary (/Sig): Là phần lưu trữ dữ liệu chữ ký thật. Bên trong có các khóa như:

+ /Filter: định nghĩa phần mềm ký (ví dụ "Adobe.PPKLite").
+ /SubFilter: định dạng chữ ký (ví dụ "adbe.pkcs7.detached").
+ /ByteRange: xác định vùng dữ liệu được ký.
+ /Contents: chứa dữ liệu chữ ký PKCS#7/CMS (dạng nhị phân hoặc hex).
+ /M: thời gian ký.
+ /Name, /Reason, /Location: thông tin bổ sung.

- /ByteRange: Là mảng 4 giá trị cho biết phạm vi byte của file PDF được dùng để tính toán hash. Phần /Contents (chứa chữ ký) bị bỏ qua khi tính hash.

- /Contents: Là vùng chứa chữ ký số thực tế (dạng PKCS#7 hoặc CMS). Dữ liệu này là kết quả của quá trình mã hóa bằng khóa riêng tư của người ký.

- Incremental updates: PDF hỗ trợ cập nhật theo kiểu gia tăng. Khi có thêm chữ ký, phần mới được nối vào cuối file mà không thay đổi nội dung trước đó. Mỗi lần ký tạo ra một revision mới.

- DSS (Document Security Store): Là phần mở rộng theo chuẩn PAdES (PDF Advanced Electronic Signatures). DSS chứa dữ liệu phục vụ xác minh lâu dài, như OCSP, CRL và timestamp token. Điều này cho phép xác minh chữ ký ngay cả khi chứng chỉ gốc đã hết hạn.

II. Sơ đồ liên kết giữa các đối tượng (object)

- Cấu trúc khái quát của file PDF có chữ ký như sau:

Catalog

├── /Pages → Pages Tree → Page → /Contents → Stream data

│

└── /AcroForm → /Fields → Signature Field (Widget)

└── /V → Signature Dictionary (/Sig)

├── /Filter → "Adobe.PPKLite"

├── /SubFilter → "adbe.pkcs7.detached"

├── /ByteRange → [0, 1234, 2345, 6789]

├── /Contents → <PKCS#7 data>

├── /M → D:YYYYMMDDHHmmSSZ

└── /Name, /Reason, /Location

Nếu tài liệu theo chuẩn PAdES có thêm phần:

└── /DSS → (CRLs, OCSPs, Certificates, TimeStamps)

III. Vai trò của các đối tượng quan trọng

- Catalog: Đối tượng gốc, liên kết tới /Pages (cây trang) và /AcroForm (biểu mẫu).

- Pages tree: Tổ chức và quản lý danh sách các trang trong PDF.

- Page object: Chứa nội dung hiển thị và annotation (bao gồm vùng chữ ký).

- AcroForm: Quản lý toàn bộ các trường biểu mẫu, trong đó có các signature field.

- Signature field (Widget): Xác định vị trí và hiển thị chữ ký.

- Signature dictionary (/Sig): Lưu trữ dữ liệu chữ ký, thông tin người ký, lý do ký, và thời gian ký.

- /ByteRange: Cho biết vùng dữ liệu được ký (để tính hash).

- /Contents: Lưu giá trị chữ ký số (PKCS#7 hoặc CMS).

- DSS: Chứa dữ liệu xác minh lâu dài, như OCSP, CRL, chứng chỉ, và timestamp.

- **Incremental update**: Cho phép thêm chữ ký mới mà không làm thay đổi nội dung đã ký trước đó.

2) Thời gian ký được lưu ở đâu?

I. Các vị trí có thể lưu thông tin thời gian

1. /M trong Signature dictionary

- Dạng: /M (D:YYYYMMDDHHmmSSZ)
- Là thời điểm mà phần mềm ký ghi lại khi tạo chữ ký.
- Thông tin này chỉ mang tính hiển thị, không có giá trị pháp lý vì được lấy từ đồng hồ hệ thống của người ký.

2. Timestamp token (RFC 3161) trong chữ ký PKCS#7

- Nằm trong thuộc tính timeStampToken của đối tượng CMS.
- Do máy chủ Timestamp Authority (TSA) cung cấp.
- Có giá trị pháp lý nếu TSA được tin cậy, vì được ký số riêng và có thể xác minh.

3. Document timestamp (PAdES)

- Là một chữ ký RFC 3161 được áp dụng ở cấp tài liệu, không gắn với người ký cụ thể.
- Dùng để chứng minh rằng tài liệu tồn tại tại một thời điểm xác định, ngay cả khi chưa có người ký.

4. DSS (Document Security Store)

- Phần mở rộng trong PDF theo chuẩn PAdES, có thể lưu timestamp, OCSP, CRL phục vụ xác minh lâu dài (LTV).
- Cho phép xác thực chữ ký ngay cả khi chứng chỉ hoặc TSA hết hạn.