1) Cấu trúc PDF liên quan đến Chữ ký số

Tóm tắt các đối tượng (Object)

Một file PDF là một tập hợp các "đối tượng" (objects) được liên kết với nhau. Chữ ký số không phải là một "hình ảnh" đơn giản, mà là một tập hợp các đối tượng cấu trúc được chèn vào file để bảo mật.

Catalog (Root Object): Là "mục lục" gốc của toàn bộ file PDF. Mọi thứ trong file đều được truy cập bắt đầu từ đây.

Pages Tree (/Pages): Đối tượng mà Catalog trỏ tới, chứa một danh sách (cấu trúc cây) tất cả các đối tượng trang (/Page) trong tài liệu.

Page Object (/Page): Đại diện cho một trang riêng lẻ. Nó định nghĩa kích thước trang và trỏ đến nội dung của trang đó.

Content Streams: Đối tượng "luồng" chứa các lệnh vẽ thực tế (văn bản, đường kẻ, hình ảnh) để hiển thị nội dung của một trang.

AcroForm: Một đối tượng cấp cao, được trỏ đến từ Catalog. Nó chứa định nghĩa cho tất cả các trường tương tác (interactive fields) trong tài liệu (như text box, check box, và quan trọng nhất là chữ ký).

Signature Field (Trường Chữ ký): Một loại trường cụ thể trong AcroForm. Về mặt kỹ thuật, đây là một "Widget Annotation" (Chú thích Widget) vừa xác định vị trí và hình ảnh của chữ ký trên trang, vừa trỏ đến dữ liệu chữ ký.

Signature Dictionary (/Sig): Đây là "trái tim" của chữ ký số. Nó là một đối tượng riêng biệt chứa toàn bộ thông tin mật mã.

/Contents: Một mục (entry) bên trong Signature Dictionary. Nó chứa chính chữ ký mật mã, thường là một gói tin PKCS#7/CMS (dưới dạng nhị phân/hex).

/ByteRange: Một mục (entry) quan trọng bên trong Signature Dictionary. Đây là một mảng (array) các cặp số nguyên (offset, length) chỉ định chính xác những phần nào của file PDF đã được băm (hash) để tạo chữ ký. Nó băm toàn bộ file, ngoại trừ chính vùng /Contents (vì vùng này chưa tồn tại khi băm).

Incremental Updates (Cập nhật tăng dần): Đây là phương pháp để lưu chữ ký. Thay vì ghi đè file gốc (sẽ làm hỏng tất cả các offset), một chữ ký hợp lệ luôn luôn được thêm vào cuối file dưới dạng một "cập nhật tăng dần". File PDF đã ký sẽ có ít nhất hai bảng xref (bảng chỉ mục). Việc này đảm bảo nội dung gốc không bị thay đổi.

DSS (Document Security Store): Một đối tượng PAdES (PDF Advanced Electronic Signatures) được thêm vào (thường trong một incremental update sau chữ ký) để hỗ 'trợ xác thực lâu dài (LTV). Nó chứa "bằng chứng" để xác minh chữ ký (như chuỗi chứng thư, phản hồi OCSP, CRLs).

Sơ đồ Luồng truy xuất Chữ ký

Đây là cách một trình đọc PDF tìm thấy và xác thực chữ ký:

1. Luồng tìm trường Chữ ký (Từ Gốc đến Chữ ký):

Catalog (Root Object)

→ /AcroForm (Danh sách các trường tương tác)

→ [/Fields] (Mảng các trường, bao gồm...)

→ Signature Field (Trường Chữ ký / Widget)

→ /V (Giá trị của trường này...)

→ Signature Dictionary (/Sig) (Đối tượng chứa dữ liệu ký)\

2. Luồng xác thực (Từ Chữ ký đến Dữ liệu):

Signature Dictionary (/Sig)

→ /ByteRange (Đọc mảng này để biết cần băm (hash) phần nào của file)

→ /Contents (Đọc khối dữ liệu PKCS#7 để lấy public key và chữ ký đã mã hóa)

3. Luồng LTV (PAdES):

Catalog (Root Object)

→ /DSS (Truy cập vào kho lưu trữ bằng chứng xác thực)

→ [/Certs, /OCSPs, /CRLs] (Lấy các chứng thư, OCSP, CRL cần thiết để kiểm tra)

2) Thông tin Thời gian ký được lưu ở đâu?

Thời gian là một yếu tố then chốt trong chữ ký số, nhưng không phải mọi thông tin thời gian đều có giá trị pháp lý như nhau. Có nhiều nơi lưu trữ thông tin này, với các mức độ tin cậy khác nhau.

Các vị trí lưu trữ thông tin thời gian

a. Trong Signature Dictionary (/M):

- Vị trí: Signature Dictionary -> /M (D:YYYYMMDDHHMMSS+ZZ'OO')

  Mô tả: Đây là một chuỗi văn bản (string) đơn giản, lưu lại thời gian theo giờ địa phương của máy tính người ký.

- Giá trị pháp lý: Không có. Nó chỉ mang tính tham khảo.

b. Trong Gói tin PKCS#7 (signingTime):

- Vị trí: Signature Dictionary -> /Contents -> PKCS#7 -> SignerInfo -> AuthenticatedAttributes -> signingTime
- Mô tả: Đây là một thuộc tính đã được xác thực (authenticated attribute), nghĩa là nó được ký bởi private key của người ký. Nó chứng minh rằng người ký đã "khai" thời điểm đó.
- Giá trị pháp lý: Tốt hơn /M, nhưng vẫn yếu. Nó vẫn dựa trên đồng hồ hệ thống của người ký (vốn có thể bị thay đổi).

c. Dấu thời gian (Timestamp Token RFC 3161):

- Vị trí: Signature Dictionary -> /Contents -> PKCS#7 -> SignerInfo -> UnauthenticatedAttributes -> timeStampToken
- Mô tả: Đây là tiêu chuẩn vàng cho bằng chứng thời gian. Khi ký, phần mềm gửi hash của chữ ký đến một bên thứ ba tin cậy (TSA - Timestamping Authority). TSA này tạo ra một "dấu thời gian" (là một chữ ký CMS riêng biệt) và trả về. Gói tin này được nhúng vào chữ ký gốc.
- Giá trị pháp lý: Rất cao. Nó chứng minh một cách độc lập rằng chữ ký đã tồn tại trước thời điểm mà TSA xác nhận.

d. Đối tượng Document Timestamp (/DTS):

- Mô tả: Đây là một loại chữ ký PAdES riêng biệt, không phải là chữ ký của người dùng, mà là một chữ ký của TSA áp dụng cho toàn bộ tài liệu (hoặc các thay đổi của tài liệu). Nó thường được dùng để xác nhận trạng thái của tài liệu tại một thời điểm cụ thể.

e. Trong DSS (Document Security Store):

- Mô tả: Bản thân DSS không phải là một "thời gian ký", nhưng nó chứa các bằng chứng liên quan đến thời gian. Ví dụ: các phản hồi OCSP và CRLs được lưu trong DSS đều có dấu thời gian riêng (ví dụ: thisUpdate, nextUpdate), chứng minh rằng tại thời điểm đó, chứng thư vẫn còn hợp lệ.