BÁO CÁO BÀI TẬP: CHỮ KÝ SỐ TRONG FILE PDF

Môn học: An toàn và Bảo mật thông tin
Chủ đề: Chữ ký số trong file PDF
Sinh viên thực hiện: Lê Tuấn Anh
Mã sinh viên: K225480106001
Lớp: 58KTPM
Giảng viên hướng dẫn: Đỗ Duy Cốp
Ngày nộp: 31/10/2025

I. GIỚI THIỆU

Chữ ký số trong file PDF là một trong những ứng dụng quan trọng của mật mã học hiện đại, giúp đảm bảo **tính toàn vẹn, xác thực và không thể chối bỏ** của tài liệu điện tử. Trong bài này, sinh viên tiến hành nghiên cứu, thực hiện và phân tích quy trình **ký và xác thực chữ ký số trong file PDF**, sử dụng công cụ Python kết hợp với chứng chỉ số định dạng X.509.
Kết quả gồm: báo cáo lý thuyết, file PDF đã ký và công cụ xác thực chữ ký.

II. CẤU TRÚC PDF LIÊN QUAN ĐẾN CHỮ KÝ SỐ

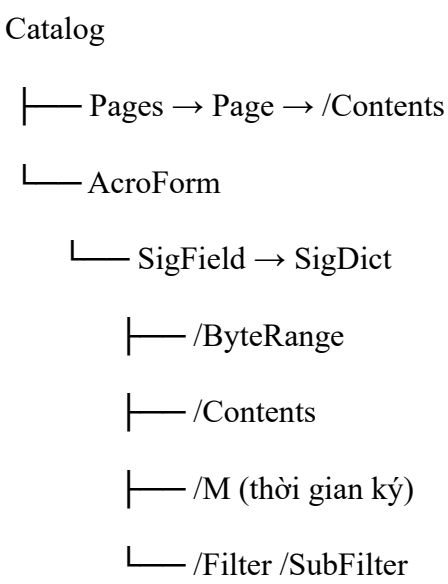
Một file PDF chứa nhiều thành phần (object) có liên hệ chặt chẽ. Trong đó, phần chữ ký số được nhúng trong cấu trúc **AcroForm** thông qua một **Signature Field**.

1. Các thành phần chính:

- **Catalog:** Đối tượng gốc, tham chiếu đến toàn bộ tài liệu.
- **Pages:** Quản lý danh sách các trang trong tài liệu.
- **Page:** Mỗi trang riêng lẻ, có thể chứa nội dung và các trường biểu mẫu.
- **AcroForm:** Nơi lưu trữ thông tin về các form field, bao gồm cả chữ ký số.
- **Signature Field:** Vùng hiển thị chữ ký, thường chứa thông tin về vị trí, hình ảnh và trạng thái ký.
- **Signature Dictionary (/Sig):** Chứa thông tin kỹ thuật của chữ ký như thuật toán, thời gian ký, chứng chỉ, dữ liệu hash,…
- **/ByteRange:** Chỉ định các vùng dữ liệu được tính toán khi ký để đảm bảo tính toàn vẹn.
- **/Contents:** Vùng lưu trữ dữ liệu chữ ký số (dạng nhị phân PKCS#7 hoặc CAdES).

- **DSS (Document Security Store):** Khu vực tùy chọn, lưu thông tin phục vụ xác minh lâu dài (chứng chỉ, OCSP, CRL,…).

2. Sơ đồ cấu trúc:



Cấu trúc này giúp PDF có thể lưu trữ nhiều chữ ký, mỗi chữ ký được ghi thêm dưới dạng **incremental update**, không làm thay đổi nội dung gốc.

III. THỜI GIAN KÝ TRONG FILE PDF

Thông tin thời gian ký có thể được lưu tại nhiều vị trí khác nhau, tùy theo mức độ yêu cầu pháp lý.

1. Các vị trí chứa thông tin thời gian:

- **Trường /M** trong *Signature Dictionary:* ghi lại thời gian ký ở dạng chuỗi, chỉ mang tính tham khảo.
- **Timestamp Token (RFC 3161):** là dữ liệu được cấp bởi máy chủ chứng thực thời gian (TSA), nằm trong phần chữ ký PKCS#7, có giá trị pháp lý.
- **Document Timestamp (PAdES):** dạng chữ ký đặc biệt, dùng để đóng dấu thời gian cho tài liệu mà không cần người ký.
- **DSS (Document Security Store):** có thể chứa dữ liệu timestamp, OCSP, CRL giúp xác minh tính hợp lệ trong thời gian dài.

2. So sánh:

Tiêu chí	**/M**	**Timestamp RFC3161**
Nguồn	Người ký tự chèn	Máy chủ TSA cung cấp
Tính pháp lý	Không có	Có
Dễ bị thay đổi	Có	Không
Mục đích	Ghi thời điểm ký	Chứng minh thời điểm tồn tại tài liệu

IV. QUY TRÌNH KÝ VÀ XÁC THỰC CHỮ KÝ

1. Quy trình ký

Quy trình ký bắt đầu bằng việc chuẩn bị file PDF gốc và tạo trường chữ ký (Signature Field). Sau đó, chương trình xác định vùng dữ liệu được phép tính toán bằng /ByteRange, tính toán giá trị hash (ví dụ SHA-256), và tạo chữ ký số bằng khóa riêng RSA.
Dữ liệu chữ ký được nhúng vào vùng /Contents, cùng với thông tin chứng chỉ, thuật toán, thời gian ký.

Kết quả là một file PDF mới có chứa chữ ký số, được lưu dưới dạng incremental update để giữ nguyên tài liệu gốc.

2. Quy trình xác thực

Khi xác minh chữ ký, chương trình đọc thông tin từ Signature Dictionary, tách vùng dữ liệu cần kiểm tra theo ByteRange, tính toán lại giá trị hash và so sánh với chữ ký lưu trong /Contents.
Chứng chỉ công khai (public key) được sử dụng để xác thực tính hợp lệ.
Ngoài ra, quá trình còn kiểm tra chuỗi chứng chỉ, tình trạng thu hồi (CRL, OCSP) và kiểm tra timestamp.
Nếu có bất kỳ thay đổi nào trong nội dung sau khi ký, chữ ký sẽ bị coi là không hợp lệ.

V. RỦI RO BẢO MẬT VÀ GIẢI PHÁP

Rủi ro	Mô tả	Giải pháp
Rò rỉ khóa riêng	Khóa ký bị lưu công khai hoặc chia sẻ nhầm.	Lưu trữ khóa ở nơi an toàn, dùng mật khẩu hoặc thiết bị bảo mật.
Tấn công Padding Oracle	Lợi dụng lỗi trong thuật toán RSA PKCS#1 v1.5.	Sử dụng RSA-PSS hoặc ECDSA an toàn hơn.
Replay Attack	Sao chép chữ ký hợp lệ sang tài liệu khác.	Dùng timestamp và định danh tài liệu duy nhất.
Sửa đổi sau ký	Chèn nội dung mới vào incremental update.	Kiểm tra toàn bộ ByteRange và xác minh toàn vẹn.
Lộ khóa qua mã nguồn	Private key bị ghi cứng trong file code.	Lưu riêng file khóa, hạn chế quyền truy cập.

Các biện pháp trên giúp đảm bảo tài liệu PDF ký số vẫn giữ được giá trị pháp lý và tính an toàn trong môi trường số.

VI. KẾT LUẬN

Bài thực hành đã giúp sinh viên hiểu rõ quy trình **tạo và xác thực chữ ký số trong tài liệu PDF**, từ cấu trúc đối tượng, cách lưu thông tin thời gian, cho đến các rủi ro và biện pháp bảo mật.

Việc ký số không chỉ là thao tác kỹ thuật, mà còn liên quan đến chuẩn định dạng quốc tế (PDF 1.7, PAdES) và các quy định về chữ ký điện tử có giá trị pháp lý.
Thông qua bài này, sinh viên có thể vận dụng kiến thức mật mã học vào các ứng dụng thực tế như ký văn bản, hợp đồng điện tử, và xác thực tài liệu hành chính.

Sinh viên thực hiện: