BÀI TẬP VỀ NHÀ SỐ 2 CHỮ KÝ SỐ

GV giảng dạy Thầy: Đỗ Duy Cốp

SV: Vũ Đức Tú

MSV: K225480106068

BÀI LÀM

1.

(Catalog) Là gốc (root object) của tài liệu PDF, tham chiếu đến các đối tượng cấp cao khác như /Pages (cây trang) và /AcroForm (biểu mẫu). Nếu tài liệu có chữ ký, Catalog sẽ chứa khóa /AcroForm trỏ đến form có trường chữ ký.

(Pages tree) Là cấu trúc cây quản lý các trang trong PDF. Gồm một node gốc /Pages chứa danh sách các /Page (trang con). Không chứa dữ liệu chữ ký, nhưng mọi nội dung hiển thị (kể cả khung chữ ký) đều nằm trong Page.

(Page object) Mỗi trang PDF là một /Page object, mô tả bố cục, tài nguyên, và nội dung của trang. Nếu có khung chữ ký hiển thị (signature widget appearance), thì widget này sẽ được tham chiếu trong /Annots của Page.

(Resources) Tập hợp các tài nguyên đồ họa được dùng trong trang: font, ảnh, pattern, XObject… Phần này phục vụ việc hiển thị (render) khung chữ ký, không chứa dữ liệu ký.

(Content streams) Chứa các lệnh vẽ (PDF drawing commands) mô tả nội dung trang. Nếu chữ ký có hiển thị hình ảnh, logo, hoặc text (“Đã ký bởi…”), nội dung đó được mô tả trong Content stream.

(Xobject) Là đối tượng đồ họa con có thể tái sử dụng (như template hoặc ảnh). Trong chữ ký, XObject có thể lưu hình nền khung chữ ký hoặc hình ảnh con dấu.

(AcroForm) Là dictionary mô tả toàn bộ form fields của tài liệu. Chữ ký trong PDF là một loại trường form đặc biệt (/Sig field). Catalog trỏ tới /AcroForm.

(Signature field (widget)) Là một form field có type /Sig, biểu diễn một vùng chữ ký (có thể hiển thị hoặc ẩn). Trường này trỏ tới một Signature dictionary thực tế qua khóa /V (value).

(Signature dictionary) Chứa dữ liệu chi tiết về chữ ký: người ký, thời gian, lý do, phạm vi dữ liệu được ký, và chữ ký nhị phân. Các khóa quan trọng: /Filter, /SubFilter, /ByteRange, /Contents, /Name, /M, /Reason, v.v.

(/ByteRange) Mảng mô tả vị trí byte của dữ liệu được ký trong file PDF. Cấu trúc [start1, length1, start2, length2]. Phần giữa bị bỏ trống để chứa chữ ký số (/Contents). Khi xác minh, PDF đọc lại các vùng này để tính hash.

(/Contents) Chứa chữ ký số nhị phân (PKCS#7 hoặc CMS) ở dạng Base16/Hex. Đây là phần dữ liệu thực được tạo khi ký — bao gồm chứng thư số, chữ ký, và có thể chứa timestamp.

(Incremental updates) Cơ chế PDF cho phép thêm chữ ký mà không sửa phần cũ. Khi ký mới, phần dữ liệu ký được thêm vào cuối file (append-only). Nhờ đó, có thể chứa nhiều chữ ký tuần tự mà không phá vỡ tính toàn vẹn.

(DSS (Document Security Store) Là phần mở rộng theo chuẩn PAdES, chứa các dữ liệu xác minh lâu dài: chứng chỉ, CRL, OCSP, timestamp,… để đảm bảo khả năng kiểm chứng về sau (long-term validation). DSS được thêm như incremental update cuối cùng.

- **Các object refs quan trọng và vai trò của chúng.**

Catalog : Điểm bắt đầu của toàn bộ cấu trúc PDF. Trỏ đến /Pages (cây trang) và /AcroForm (biểu mẫu chứa trường chữ ký).

AcroForm : Chứa danh sách /Fields – trong đó mỗi field có thể là trường chữ ký (/FT /Sig). Là nơi tập trung tất cả chữ ký có trong tài liệu.

Signature Field : Biểu diễn vùng chữ ký (có thể hiển thị hoặc ẩn). Có khóa /V trỏ tới Signature Dictionary chứa dữ liệu chữ ký thực.

Signature Dictionary : Là object quan trọng nhất của chữ ký. Lưu thông tin ký, phạm vi dữ liệu được ký, chứng chỉ, hash, thời gian, lý do ký...

/ByteRange : Mô tả vùng byte trong file được đưa vào tính hash (vùng trắng là nơi chứa /Contents). Khi xác minh, trình đọc PDF dùng /ByteRange để kiểm tra tính toàn vẹn.

/Contents : Chứa chữ ký số thực tế (binary signature + chứng chỉ). Bao gồm: hash dữ liệu, chứng chỉ người ký, timestamp (RFC 3161 nếu có).

Page Object : Đại diện cho từng trang trong tài liệu. Nếu có hiển thị vùng chữ ký (hình ảnh, text "Đã ký…"), thì /Page sẽ chứa annotation trỏ đến Widget.

/Annots : Chứa danh sách các annotation trên trang, bao gồm widget chữ ký. Dùng để hiển thị khung chữ ký.

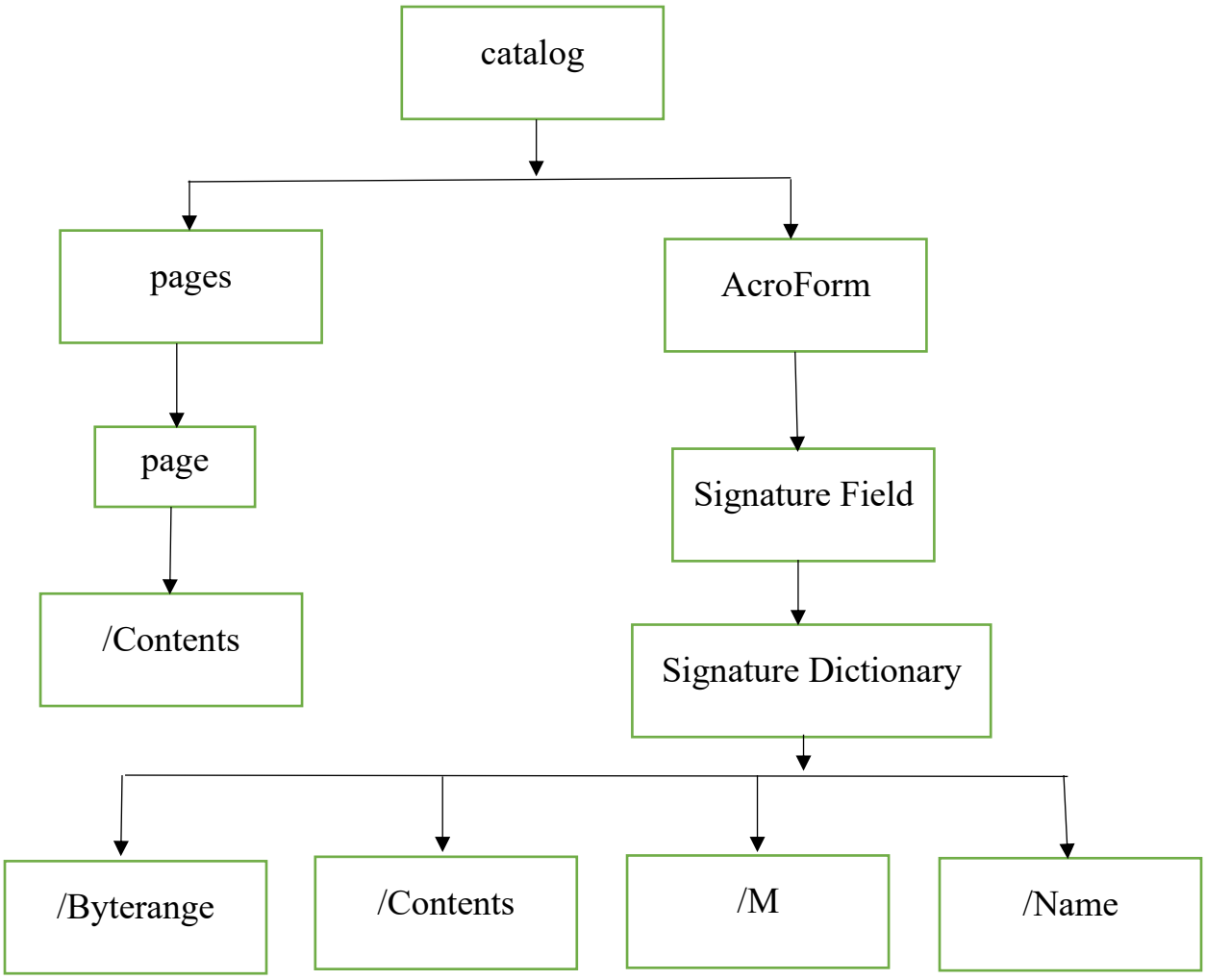
XObject (optional) : Chứa hình ảnh hiển thị trong khung chữ ký (logo, dấu mộc, hoặc text template). Không ảnh hưởng tới giá trị pháp lý.

Incremental Update Section : Khi tài liệu được ký, PDF không ghi đè dữ liệu cũ mà thêm phần mới ở cuối file. Mỗi chữ ký là một incremental update, giúp xác minh từng lớp ký độc lập.

DSS : Lưu trữ dữ liệu xác minh lâu dài: chứng chỉ, OCSP, CRL, và timestamp. Giúp việc xác minh vẫn hợp lệ sau nhiều năm (LTV – Long Term Validation).

Trailer dictionary : Chứa thông tin Root (Catalog), Prev (offset phần trước), và XRef. Giúp PDF reader tìm đúng phiên bản chữ ký mới nhất.

- **Sơ đồ object**



2.

- Thời gian ký trong file PDF có thể được lưu ở nhiều vị trí khác nhau. Thông thường, nó được ghi trong trường /M của Signature Dictionary (/Sig), là dạng chuỗi thời gian do phần mềm ký tạo ra, nhưng không có giá trị pháp lý. Thời gian có giá trị xác thực thật sự nằm trong timestamp token (RFC 3161) bên trong chữ ký PKCS#7, do TSA (Tổ chức cấp dấu thời gian) phát hành. Ngoài ra, trong chuẩn PAdES, thời gian còn có thể được lưu trong Document Timestamp Object (tem thời gian cho toàn bộ tài liệu) và DSS (Document Security Store), nơi chứa dữ liệu xác minh lâu dài. Như vậy, chỉ timestamp RFC 3161 và Document Timestamp mới có giá trị pháp lý, còn /M chỉ mang tính hiển thị.

- Thông tin thời gian trong chữ ký PDF có thể được lưu tại bốn vị trí chính:
(1) Trường /M trong Signature Dictionary (/Sig) – đây là thời gian do phần mềm ký ghi lại, chỉ ở dạng văn bản và không có giá trị pháp lý.
(2) Timestamp token (RFC 3161) nằm trong PKCS#7 (thuộc tính *timeStampToken*) – là dấu thời gian được cấp bởi TSA, có giá trị pháp lý.
(3) Document Timestamp Object trong chuẩn PAdES – tem thời gian cho toàn bộ tài liệu, xác nhận tài liệu tồn tại tại thời điểm đó.
(4) DSS (Document Security Store) – phần lưu trữ thông tin xác minh dài hạn, có thể chứa timestamp, OCSP, CRL và chứng chỉ phục vụ việc xác thực lâu dài.

- Sự khác biệt giữa /M và timestamp RFC 3161 nằm ở tính xác thực và giá trị pháp lý:

+ /M là thời gian được phần mềm ký chèn vào trong Signature Dictionary, chỉ mang tính thông tin hiển thị, dễ bị chỉnh sửa và không có giá trị pháp lý.

+ Timestamp RFC 3161 là dấu thời gian điện tử do TSA (Time Stamping Authority) cấp, được ký số độc lập để chứng minh tài liệu đã tồn tại tại một thời điểm cụ thể, nên có giá trị pháp lý và xác thực cao.

KẾT LUẬN: /M = thời gian do người ký ghi, còn timestamp RFC 3161 = thời gian được bên thứ ba đáng tin cậy chứng thực.

RỦI RO BẢO MẬT TRONG HỆ THỐNG CHỮ KÝ SỐ PDF

Hệ thống chữ ký số PDF đảm bảo tính xác thực, toàn vẹn và chống chối bỏ cho tài liệu điện tử. Tuy nhiên, trong quá trình ký và xác thực, vẫn tồn tại nhiều rủi ro bảo mật cần được nhận diện và phòng tránh như sau:

1. Rủi ro lộ khóa bí mật

Khóa bí mật (private key) là thành phần quan trọng nhất. Nếu bị lộ, kẻ tấn công có thể giả mạo chữ ký giống hệt người thật.
Giải pháp: Mã hóa khóa bằng mật khẩu mạnh, giới hạn quyền truy cập, hoặc lưu khóa trong thiết bị bảo mật (HSM, USB Token).

2. Rủi ro xác minh không đầy đủ

Một số hệ thống chỉ kiểm tra chữ ký hợp lệ về mặt kỹ thuật mà không xác thực chứng chỉ hay trạng thái thu hồi.
Giải pháp: Thực hiện kiểm tra OCSP/CRL, xác thực chuỗi chứng chỉ (CA chain) và đảm bảo chứng chỉ còn hiệu lực.

3. Rủi ro chỉnh sửa nội dung sau khi ký

Kẻ tấn công có thể lợi dụng định dạng incremental update hoặc hybrid xref để chèn dữ liệu mới mà không làm hỏng chữ ký.
Giải pháp: Chuẩn hóa PDF trước khi ký (dùng pikepdf), bật chế độ DocMDP để khóa tài liệu sau khi ký.

4. Rủi ro giả mạo hình ảnh chữ ký

Chèn hình ảnh chữ ký (PNG/JPG) mà không có chữ ký số thật có thể bị giả mạo dễ dàng.
Giải pháp: Luôn đính kèm PKCS#7/CMS signature và hiển thị rõ ràng chứng chỉ người ký.

5. Rủi ro từ chứng chỉ tự ký

Chứng chỉ tự ký (self-signed) không do tổ chức CA cấp nên không có giá trị pháp lý, dễ bị giả mạo danh tính.
Giải pháp: Sử dụng chứng chỉ được cấp bởi CA tin cậy (VNPT-CA, Viettel-CA, FPT-CA...) khi triển khai thực tế.

6. Rủi ro thao túng thời gian ký

Người ký hoặc kẻ tấn công có thể thay đổi trường signingTime trong PKCS#7 để làm sai lệch thời điểm ký.
Giải pháp: Gắn RFC3161 Timestamp Token (TSA) khi ký, hoặc ghi nhận thời gian ký độc lập trong hệ thống.

7. Rủi ro khi truyền tải tài liệu

Tài liệu ký được gửi qua email hoặc cloud mà không mã hóa có thể bị sao chép hoặc thay đổi.
Giải pháp: Mã hóa PDF bằng AES-256, sử dụng HTTPS/VPN khi truyền tải.