I. Cấu trúc PDF liên quan chữ ký số

Khi tích hợp chữ ký số, tệp PDF không ghi đè lên nội dung gốc mà thêm một bản cập nhật gia tăng (incremental update). Phần này chứa các đối tượng (object) mới phục vụ việc lưu và hiển thị chữ ký.

1. Thành phần chính trong cấu trúc PDF có chữ ký

Thành phần	Vai trò
Catalog	Nút gốc của tài liệu, trỏ tới cây trang (/Pages) và biểu mẫu (/AcroForm).
Pages Tree / Page Object	Quản lý các trang và nội dung hiển thị.
AcroForm	Định nghĩa các trường biểu mẫu, trong đó có trường chữ ký.
Signature Field (Widget)	Vùng hiển thị chữ ký, gắn vào vị trí cụ thể trên trang PDF.
Signature Dictionary (/Sig)	Lưu thông tin về người ký, thời gian ký và dữ liệu chữ ký (PKCS#7/CAdES).
/ByteRange	Xác định vùng byte được ký (loại trừ vùng chứa chữ ký).
/Contents	Chứa dữ liệu chữ ký số (chuỗi PKCS#7).
DSS (Document Security Store)	Lưu dữ liệu xác minh dài hạn (certificates, OCSP, CRL, timestamp).

2. Sơ đồ cấu trúc logic

Catalog

├── /Pages

│ └── Page

│ ├── /Contents

│ └── /Resources

└── /AcroForm

├── /Fields

│ └── Signature Field (Widget)

│ ├── /T (Tên trường)

│ ├── /FT /Sig

│ ├── /Rect (Tọa độ hiển thị)

│ └── /V → Signature Dictionary

└── /NeedAppearances true

3. Cấu trúc Signature Dictionary

Đây là đối tượng trung tâm của quá trình ký:

/Type /Sig

/Filter /Adobe.PPKLite

/SubFilter /adbe.pkcs7.detached

/Name (Người ký)

/M (D:20251030T123456+07'00')

/Reason (Ký báo cáo học phần)

/Location (Thai Nguyen)

/ByteRange [0 12345 56789 99999]

/Contents <3082A5...> % dữ liệu PKCS#7

/Cert (certificate chain)

Giải thích:

- /Filter và /SubFilter: xác định định dạng chữ ký (Adobe, PKCS#7, CAdES, RFC3161…).
- /ByteRange: vùng dữ liệu được hash trước khi ký.
- /Contents: khối dữ liệu PKCS#7/CMS, chứa hash, chứng chỉ và timestamp.
- /M: thời gian ký (dạng text, không có giá trị pháp lý).
- /Reason, /Location, /Name: thông tin mô tả của người ký.

4. Incremental Update và DSS

- Incremental Update: mọi thay đổi (chữ ký, chỉnh sửa) được ghi nối tiếp vào cuối file; phần mềm xác minh có thể phát hiện mọi thay đổi sau khi ký.
- DSS (Document Security Store): lưu trữ thêm chứng chỉ, OCSP, CRL và timestamp để xác minh lâu dài (LTV).
  Ví dụ:

/DSS

 /Certs [obj_ref]

 /OCSPs [obj_ref]

 /CRLs [obj_ref]

 /VRI << /sig1 {Certs, OCSPs, CRLs} >>

5. Tóm tắt

File PDF có chữ ký số là sự mở rộng của cấu trúc PDF chuẩn, bổ sung các đối tượng:

- AcroForm để chứa trường ký,

- Signature Dictionary để lưu dữ liệu chữ ký,
- DSS để đảm bảo xác minh lâu dài.
  Nhờ cơ chế incremental update, PDF có thể lưu nhiều chữ ký mà vẫn bảo toàn nội dung gốc.

II. Thời gian ký được lưu ở đâu trong PDF

Khi một file PDF được ký, thời điểm ký có thể được ghi lại ở nhiều vị trí khác nhau. Mỗi loại lưu trữ có mục đích và giá trị pháp lý khác nhau.

1. Các vị trí lưu thông tin thời gian

Vị trí	Nằm trong	Dạng lưu	Giá trị pháp lý
/M	Signature Dictionary	Chuỗi text	Không có giá trị pháp lý
signingTime	PKCS#7 signed attributes	Dữ liệu ASN.1	Tham khảo
timeStampToken	RFC 3161 trong PKCS#7	Token có chữ ký TSA	Có giá trị pháp lý
Document Timestamp Object	PAdES extension	Object riêng	Có giá trị pháp lý
DSS	Document Security Store	Metadata lưu timestamp & OCSP	Hỗ trợ xác minh lâu dài

2. Mô tả chi tiết

a. /M trong Signature Dictionary

- Dạng: /M (D:YYYYMMDDHHmmSS+07'00')
- Lấy từ đồng hồ hệ thống người ký.
- Không được ký bảo vệ → có thể chỉnh sửa → không hợp pháp.

b. signingTime trong PKCS#7

- Là thuộc tính trong phần signedAttributes của CMS (PKCS#7).
- Được bảo vệ bởi chữ ký, nhưng vẫn phụ thuộc thời gian máy người ký → chỉ có giá trị tham khảo.

c. timeStampToken (RFC 3161)

- Do TSA (Time Stamping Authority) cấp và ký, xác nhận dữ liệu tồn tại tại thời điểm cụ thể.
- Có chữ ký của TSA → chứng cứ pháp lý về thời gian ký.

d. Document Timestamp Object (PAdES)

- Là object riêng trong PDF, thường có /SubFilter /ETSI.RFC3161.
- Được TSA ký, áp dụng cho toàn bộ tài liệu (document-level timestamp).

e. DSS (Document Security Store)

- Lưu trữ timestamp, OCSP, CRL giúp xác minh lâu dài (LTV).
- Hỗ trợ xác thực ngay cả khi TSA/CA gốc hết hạn.

3. So sánh giữa /M và RFC 3161 timestamp

Tiêu chí	/M (Signature Dictionary)	RFC 3161 Timestamp
Vị trí lưu	/Sig dictionary	Bên trong PKCS#7 (unsignedAttributes)
Nguồn thời gian	Máy người ký	Máy chủ TSA
Bảo vệ bởi chữ ký	Không	Có
Có thể bị sửa	Có	Không
Giá trị pháp lý	Không	Có
Chuẩn liên quan	PDF 1.7	RFC 3161, PAdES

4. Kết luận

Chỉ tem thời gian RFC 3161 hoặc document timestamp (PAdES) mới có giá trị chứng thực pháp lý.
Các trường như /M chỉ mang tính hiển thị và không thể dùng làm bằng chứng thời gian ký.

III. Rủi ro bảo mật khi ký và xác thực PDF

Mặc dù chữ ký số trong PDF dựa trên nền tảng mật mã học an toàn, song trong thực tế vẫn tồn tại nhiều rủi ro kỹ thuật và khai thác nếu không tuân thủ chuẩn.

1. Rủi ro thường gặp

a. Thay đổi nội dung sau khi ký (Incremental Update Attack)

- PDF cho phép lưu thêm nội dung ở phần cuối mà không xóa phần cũ.
- Kẻ tấn công có thể chèn nội dung "vô hại" vào incremental update để đánh lừa người đọc, trong khi chữ ký vẫn hiện "hợp lệ".

b. Sửa vùng /ByteRange

- Nếu công cụ xác minh xử lý sai, có thể bỏ qua một phần dữ liệu nằm ngoài /ByteRange.
- Điều này cho phép chèn nội dung giả mạo mà vẫn không phát hiện thay đổi hash.

c. Tái sử dụng chữ ký (Replay Attack)

- Một chữ ký hợp lệ từ tài liệu khác có thể bị sao chép sang file mới nếu không có cơ chế ràng buộc nội dung (content binding).

d. Lộ khóa riêng (Private Key Leak)

- Người ký lưu private key không an toàn hoặc gửi cùng mã nguồn → có thể bị giả mạo toàn bộ chữ ký.
- Đề xuất: sử dụng RSA ≥ 2048 bit, lưu khóa trong HSM hoặc token bảo mật.

e. Tấn công vào định dạng PKCS#1 và padding

- Một số thư viện cũ dùng RSA/PKCS#1 v1.5 dễ bị "padding oracle".
- Khuyến nghị dùng RSA-PSS hoặc ECDSA/SHA-256.

f. Lỗi xác thực chuỗi chứng chỉ hoặc timestamp

- Nếu CA hết hạn, TSA ngừng hoạt động, hoặc dữ liệu xác minh không được lưu trong DSS → file sẽ không còn xác thực được.
- Giải pháp: sử dụng PAdES-LTV để nhúng toàn bộ dữ liệu xác minh vào file.

2. Biện pháp giảm thiểu

Nhóm rủi ro	Giải pháp đề xuất
Thay đổi nội dung	Kiểm tra incremental update và ByteRange khi xác minh.
Replay, giả mạo	Gắn metadata (Reason, Location, DocumentID) vào nội dung ký.
Private key	Lưu trữ an toàn trong HSM/token, không đưa vào repo.
Timestamp	Luôn dùng TSA đáng tin cậy (RFC 3161).
Chuỗi chứng chỉ	Nhúng OCSP/CRL và Certs vào DSS để đảm bảo LTV.

3. Kết luận

- Chữ ký số trong PDF giúp bảo đảm tính toàn vẹn, xác thực và chống chối bỏ, nhưng chỉ khi được thực hiện đúng chuẩn (PAdES, RFC 3161).
- Mọi hệ thống ký phải bảo vệ khóa riêng, lưu timestamp hợp lệ và kiểm tra đầy đủ các vùng dữ liệu được ký.
- Đảm bảo điều này sẽ giúp tài liệu PDF có giá trị pháp lý và an toàn lâu dài.