Báo cáo: Cấu trúc chữ ký trong PDF và vị trí lưu thời gian

I. Tóm tắt (one-page executive summary)

Tài liệu này tóm tắt cách PDF lưu trữ chữ ký điện tử (digital signatures), các object quan trọng liên quan, và nơi thông tin thời gian (signing time / timestamp) có thể xuất hiện. Mục tiêu là cung cấp bức tranh kỹ thuật cho việc phân tích, xác minh và đánh giá rủi ro bảo mật khi xử lý PDF đã ký.

Điểm chính:

- PDF lưu chữ ký dưới dạng *signature field* (AcroForm widget) và *signature dictionary* chứa /Contents (PKCS#7/CMS) và /ByteRange (phạm vi dữ liệu được bảo vệ).
- PKCS#7 chứa certificate chain, signedAttrs (trong đó có messageDigest) và có thể chứa RFC3161 timestamp token (timeStampToken) như một unsigned attribute.
- PAdES-LTV dùng thêm DSS (Document Security Store) để nhúng chứng cứ OCSP/CRL và timestamp nhằm hỗ trợ long-term validation.
- Thời gian ký có thể xuất hiện ở nhiều nơi: /M (human-readable), timeStampToken (RFC3161), document timestamp object, và trong DSS. /M không có giá trị pháp lý so với timestamp RFC3161.

II. Cấu trúc PDF liên quan chữ ký

Các thực thể/objects chính

- Catalog (Root): entry point của tài liệu; tham chiếu tới *Pages* và *AcroForm*.
- Pages tree: Catalog → /Pages → /Kids → Page object.
- Page object: chứa /Resources (fonts, XObjects), /Contents (content streams).
- Resources: chứa XObject (image, Form XObject), fonts, color spaces.
- Content streams: dữ liệu hiển thị; chữ ký hình ảnh thường được chèn như XObject hoặc vẽ lên content stream của một trang (không phải phần chữ ký số).
- AcroForm: cấu trúc form của PDF; chứa /Fields array với các widget field (text, signature...).
- Signature field (Widget): field có /FT /Sig hoặc widget annotation có /FT /Sig. Tham chiếu tới Signature dictionary qua trường /V.
- Signature dictionary (/Sig): dictionary chứa metadata chữ ký: /Type /Sig, /Filter, /SubFilter, /Contents, /ByteRange, /M, /Name, /Reason, /Location.

- /Contents: chứa PKCS#7/CMS (thường DER-encoded) — có thể được padding bằng 0x00.
- /ByteRange: mảng 4 số [off1, len1, off2, len2] xác định hai đoạn file PDF được băm (đoạn giữa là placeholder /Contents).
- /M: thông tin thời gian dạng text (PDF Date string) — dùng cho hiển thị, không tương đương timestamp RFC3161.

Incremental updates

- PDF hỗ trợ *incremental update*: khi ký (hoặc chỉnh sửa) ứng dụng có thể **append** một revision mới (một xref + trailer mới). Chữ ký băm ByteRange thường chỉ bảo vệ một revision cụ thể; các revision sau đó có thể append nội dung mới — cần kiểm tra incremental update để phát hiện sửa đổi không mong muốn.

DSS và PAdES

- **DSS (Document Security Store)**: cấu trúc (thường lưu trong xref ở phần cuối hoặc như một object riêng) theo ETSI PAdES, chứa: certificates, revocation values (OCSP/CRL responses), và có thể timestamp tokens. DSS cho phép validator kiểm tra chữ ký mà không cần truy vấn network.

III. Liệt kê object refs quan trọng & vai trò

Sơ đồ đơn giản (object references):

Catalog (/Root)

├─ /Pages → Pages tree

│ └─ /Kids → Page (N)

│ ├─ /Resources

│ │ └─ /XObject (signature image optional)

│ └─ /Contents (content streams)

└─ /AcroForm

└─ /Fields [SigField1, ...]

└─ SigField (widget)

└─ /V → Signature dictionary (SigDict)

├─ /Contents (PKCS#7/CMS) <-- signed data, certificates

├─ /ByteRange (offsets to hashed bytes)

├─ /M (display time)

└─ /FT, /SubFilter, /Filter, /Name, /Reason

Optional (PAdES):

└─ /DSS (Document Security Store) -> contains certs, OCSP/CRL, timestamps

Giải thích role:

- Catalog — entry point; cần để locate AcroForm và pages.
- AcroForm + SigField — nơi lưu metadata form; SigField định danh vị trí hiển thị chữ ký.
- SigDict (/V) — chứa bản thực chữ ký số: /Contents (PKCS#7) và /ByteRange xác định dữ liệu đã được băm.
- /Contents — chứa SignedData: certificates, signerInfo (signedAttrs), signatureValue, có thể timestamp token trong unsignedAttrs.
- /ByteRange — quan trọng nhất để tái tạo dữ liệu đã được ký và kiểm tra messageDigest.
- DSS — cho phép LTV: nếu có, validator có thể xác minh chữ ký offline.

IV. Thời gian ký được lưu ở đâu?

Các vị trí có thể lưu thông tin thời gian

1. /M trong Signature dictionary
    - Kiểu: PDF Date string (ví dụ D:20251031...).
    - Thuộc tính: human-readable, không được bảo đảm bằng cryptographic timestamp; có thể bị giả mạo nếu phần /M bị chỉnh sửa trong revision không được băm.
2. Timestamp token (RFC 3161)
    - Vị trí: thường nằm trong PKCS#7 SignedData của /Contents dưới unsignedAttrs với OID id-aa-signatureTimeStampToken (1.2.840.113549.1.9.16.2.14).
    - Thuộc tính: là evidence cryptographic thời gian (TSA ký timestamp token), có giá trị pháp lý cao hơn /M.
3. Document timestamp object / DocTimeStamp
    - Một chữ ký timestamp có thể được lưu như một signature object đặc biệt (DocTimeStamp) theo PDF spec, lưu timestamp áp dụng lên toàn bộ document.
4. DSS (Document Security Store)
    - Nếu file theo chuẩn PAdES-LTV, DSS có thể lưu cả OCSP responses, CRLs, và timestamp tokens — giúp kiểm tra long-term without network.

So sánh /M và RFC3161 timestamp

- /M là metadata dạng text: dễ đọc nhưng không là bằng chứng criptographic; có thể bị thay đổi (trong revision khác) mà không làm hỏng /Contents nếu /M không nằm trong signedAttrs.
- RFC3161 timestamp token được tạo bởi TSA (Timestamp Authority). Nó bao gồm MessageImprint = hash của data (ví dụ digest của signedAttrs) và được TSA ký bằng private key của TSA. Vì vậy timestamp RFC3161:
    - Bảo mật: cung cấp thời điểm xác thực không thể chối cãi (nếu certificate TSA tin cậy).
    - Độc lập: timestamp nằm trong PKCS#7 (unsignedAttrs) gắn trực tiếp với chữ ký — thay đổi chữ ký sẽ phá vỡ tính hợp lệ của timestamp.

V. Rủi ro bảo mật & khuyến nghị (short)

Rủi ro chính:

- ByteRange sai lệch / placeholder /Contents bị sửa → messageDigest mismatch → phát hiện lỗi.
- Nếu chỉ dựa vào /M, thời gian có thể bị giả mạo.
- Thiếu DSS/OCSP -> cần truy vấn mạng để kiểm tra revocation (có rủi ro nếu responder không khả dụng).
- Chứng chỉ tự ký (self-signed) hoặc chain không tin cậy → chữ ký không chứng thực được danh tính.
- Incremental updates: attacker có thể append content không được bảo vệ nếu validator không kiểm tra revision history.

Khuyến nghị:

- Ưu tiên PAdES-LTV: nhúng DSS để hỗ trợ xác minh offline.
- Luôn kiểm tra messageDigest vs hash(ByteRange) và detect incremental updates.
- Kiểm tra timestamp RFC3161 khi có — đây là bằng chứng thời gian mạnh.
- Xác thực chain với trusted root và kiểm tra OCSP/CRL; nếu offline, sử dụng DSS.

VI. Đầu ra và sơ đồ object

Phần này là bản tóm tắt/diagram tuân theo yêu cầu đầu ra: 1 trang tóm tắt + sơ đồ object.

(Trong báo cáo gửi nộp, bạn có thể chèn sơ đồ dạng hình SVG/PNG minh họa đường dẫn: Catalog -> Pages -> Page -> Contents và Catalog -> AcroForm -> SigField -> SigDict (/Contents, /ByteRange).)

Tài liệu tham khảo ngắn

- PDF 32000-1:2008 (ISO 32000) — PDF Reference (AcroForms, Signatures)
- ETSI TS 103 172 / EN 319 142 (PAdES)
- RFC 3161 — Time-Stamp Protocol (TSP)
- CMS (RFC 5652) — Cryptographic Message Syntax

Kết luận ngắn: hiểu rõ mối liên hệ giữa Signature dictionary (/Contents, /ByteRange), PKCS#7, signedAttrs (messageDigest), và các nguồn timestamp là then chốt để thực hiện xác minh chữ ký đúng và phát hiện sửa đổi. PAdES (với DSS) là giải pháp mạnh để hỗ trợ long-term validation.