

BÀI TẬP SỐ 2

MÔN: AN TOÀN VÀ BẢO MẬT THÔNG TIN

Sinh viên: Nguyễn Nguyệt Linh – MSSV: K225480106039

1. Giới thiệu chung

Trong thời đại chuyển đổi số hiện nay, các tài liệu điện tử như hợp đồng, hóa đơn, chứng từ PDF được sử dụng ngày càng phổ biến. Tuy nhiên, việc chia sẻ file qua Internet đặt ra nhiều vấn đề về bảo mật, tính toàn vẹn và xác thực danh tính người gửi.

Giải pháp hiệu quả cho vấn đề này chính là chữ ký số (Digital Signature) – một phương pháp dùng mật mã khóa công khai (Public Key Cryptography) để xác minh người ký và bảo vệ nội dung tài liệu khỏi bị chỉnh sửa.

Chữ ký số trong file PDF giúp:

- Xác minh được người ký là ai (Authenticity).
- Đảm bảo nội dung không bị thay đổi (Integrity).
- Chống việc chối bỏ hành vi ký (Non-repudiation).

2. Cấu trúc của file PDF có chữ ký số

Một file PDF sau khi được ký số không chỉ chứa nội dung văn bản, mà còn bao gồm các vùng dữ liệu đặc biệt do phần mềm ký chèn thêm.

Các vùng này được quy định bởi chuẩn PAdES (PDF Advanced Electronic Signatures) và thường gồm:

2.1. ByteRange

- Là mảng byte thể hiện vị trí dữ liệu trong file PDF được dùng để tính giá trị băm (hash).
- Cấu trúc: [0 <start> <length> <end>].
- Mọi dữ liệu nằm ngoài ByteRange (thường là vùng chứa chữ ký) sẽ không được tính vào hash.

=> Nếu có ai sửa đổi nội dung nằm trong vùng được bảo vệ, hash sẽ thay đổi → chữ ký vô hiệu hóa ngay lập tức.

2.2. Contents

- Chứa chữ ký số đã được mã hóa bằng khóa riêng (private key).
- Dữ liệu ở đây thường ở dạng Base64 hoặc DER, kích thước từ vài trăm đến vài nghìn byte.
- Đây là phần quan trọng nhất, vì nó thể hiện kết quả mã hóa của giá trị hash.

2.3. Certificate

- Là chứng chỉ số (Digital Certificate) do Tổ chức chứng thực (CA) cấp.
- Bao gồm:
 - Tên người ký, email, tổ chức.
 - Số sê-ri chứng chỉ.
 - Ngày cấp – ngày hết hạn.
 - Khóa công khai (Public Key).
- Chứng chỉ này được dùng để xác minh danh tính và độ tin cậy của người ký.

2.4. Metadata

- Gồm các thông tin mô tả hành động ký:
 - /Name: tên người ký.
 - /Reason: lý do ký (ví dụ: xác nhận, phê duyệt, chứng thực).
 - /Location: địa điểm hoặc thiết bị ký.
 - /M: thời điểm ký.

2.5. Timestamp (Dấu thời gian)

- Là thông tin cho biết thời điểm ký số thực tế.
- Có thể được:
 - Lấy từ máy cục bộ (local time) → không đảm bảo pháp lý.
 - Lấy từ máy chủ TSA (Time Stamping Authority) → được chứng thực, có giá trị pháp lý.
- Khi xác minh, phần mềm đối chiếu thời gian ký với thời gian hiệu lực của chứng chỉ để xác định chữ ký có hợp lệ hay không.

3. Quy trình ký và xác minh chữ ký số trên file PDF

Quy trình ký số PDF tuân theo 4 bước chính:

3.1. Tạo giá trị băm (Hash)

- Phần mềm đọc toàn bộ nội dung file (trừ vùng /Contents)
- Áp dụng thuật toán băm như SHA-256 hoặc SHA-512 để tạo ra một chuỗi hash duy nhất.
- Nếu tài liệu bị thay đổi, giá trị hash sẽ khác hoàn toàn.

3.2. Mã hóa chữ ký

- Hash được mã hóa bằng khóa riêng (Private Key) của người ký bằng thuật toán RSA hoặc ECC.
- Kết quả mã hóa chính là chữ ký số.

3.3. Nhúng chữ ký vào file

- Chữ ký, chứng chỉ và các thông tin metadata được nhúng vào file PDF.
- Nếu sử dụng Timestamp Authority (TSA), dấu thời gian được thêm vào vùng /M hoặc /Prop_Build.

3.4. Xác minh chữ ký

- Khi người nhận mở file:
 - Phần mềm đọc khóa công khai từ chứng chỉ.
 - Giải mã chữ ký → lấy lại hash ban đầu.
 - Tính lại hash từ nội dung file hiện tại.
 - So sánh 2 giá trị hash:
 - Nếu trùng khớp → file hợp lệ.
 - Nếu khác → file đã bị sửa đổi.

4. Rủi ro bảo mật thường gặp

Rủi ro	Nguyên nhân	Hậu quả	Biện pháp phòng ngừa
Mất khóa riêng (Private Key)	Người dùng lưu khóa không an toàn hoặc bị virus	Hacker có thể giả mạo chữ ký	Lưu khóa trong USB Token hoặc HSM
Giả mạo chữ ký	Tạo chứng chỉ giả hoặc khóa giả	Mất tính xác thực tài liệu	Dùng chứng chỉ CA hợp pháp, xác thực OCSP/CRL
Sửa đổi tài liệu sau khi ký	Chỉnh sửa nội dung PDF	Làm sai lệch nội dung, chữ ký bị lỗi	Dùng thuật toán SHA-256, kiểm tra ByteRange
Gian lận thời gian ký	Thay đổi đồng hồ hệ thống	Làm sai lệch thời điểm ký	Sử dụng Timestamp Server (TSA)
Chứng chỉ hết hạn hoặc bị thu hồi	Không kiểm tra trạng thái chứng chỉ	Không thể xác minh chữ ký	Kiểm tra CRL hoặc OCSP trước khi xác thực

5.1. Tổng quan về rủi ro

Mặc dù chữ ký số trong file PDF dựa trên nền tảng mật mã học bất đối xứng (RSA/ECDSA) và chuẩn PAdES (PDF Advanced Electronic Signatures), nhưng nếu quy trình ký hoặc phần mềm xử lý PDF không chuẩn, kẻ tấn công vẫn có thể:

- Thay đổi nội dung mà không làm mất tính hợp lệ của chữ ký.
- Giả mạo chứng chỉ hoặc thời gian ký.
- Lợi dụng cơ chế incremental update của PDF để thêm nội dung.
- Khai thác lỗ hổng hiển thị trong trình đọc PDF.

Các rủi ro này có thể chia làm 3 nhóm chính: *rủi ro dữ liệu*, *rủi ro thuật toán/chứng chỉ*, và *rủi ro vận hành (operation)*.

5.2. Rủi ro ở mức dữ liệu (Data-level Risks)

◆ a. Thay đổi nội dung PDF sau khi ký (PDF Tampering Attack)

Cơ chế:

- File PDF cho phép “incremental update” — nghĩa là có thể **thêm dữ liệu mới** vào cuối file mà không cần ghi đè phần trước.
- Khi ký, phần mềm chỉ hash nội dung trước vùng ký (ByteRange). Hacker có thể **thêm nội dung mới sau ByteRange**, khiến phần hash không đổi.

Giải pháp:

- Khi xác minh, luôn kiểm tra **ByteRange** có bao trùm toàn bộ nội dung file không.
- Dùng thư viện có chức năng “Full File Validation” như PyHanko, Adobe Acrobat, DSS EU.
- Không dùng trình xem PDF mặc định (nhiều app bỏ qua byte ngoài ByteRange).

◆ b. Invisible Layer Attack (tấn công lớp ẩn)

Cơ chế:

- PDF hỗ trợ nhiều lớp hiển thị (Optional Content Groups – OCG).
- Hacker có thể **ẩn lớp nội dung thật** và **hiển thị lớp giả** khi mở bằng phần mềm đọc PDF.

Giải pháp:

- Khi ký, khóa tất cả các OCG và flatten (gộp lớp).
- Xác minh bằng phần mềm có tính năng **layer inspection**.

◆ c. Hash Collision

Cơ chế:

- Nếu dùng thuật toán yếu như **MD5** hoặc **SHA-1**, hacker có thể tạo hai file khác nhau nhưng có cùng hash.
- Khi đó, một chữ ký hợp lệ cho file A sẽ hợp lệ cho file B.

Thực tế:

- Vào năm 2017, Google công bố *SHAttered Attack* cho thấy hai PDF khác nhau có cùng SHA-1 hash.
- Điều này cho phép “tạo hợp đồng giả mà chữ ký vẫn hợp lệ”.

Giải pháp:

- Chỉ sử dụng **SHA-256, SHA-384 hoặc SHA-512**.
- Cập nhật thư viện ký (OpenSSL $\geq 1.1.1$).

5.3. Rủi ro thuật toán và chứng chỉ (Algorithmic & Certificate Risks)

◆ a. Giả mạo chứng chỉ số (Fake Certificate Attack)

Cơ chế:

- Hacker tạo chứng chỉ tự ký (self-signed) có “Subject Name” giống công ty thật.
- Người dùng không kiểm tra CA \rightarrow tin rằng chữ ký hợp lệ.

Giải pháp:

- Chỉ tin cậy CA được cấp phép (VNPT, Viettel, FPT...).
- Bật kiểm tra **trust chain** và **OCSP** khi xác minh.

◆ b. Chứng chỉ hết hạn hoặc bị thu hồi

Cơ chế:

- Một chứng chỉ hợp lệ tại thời điểm ký nhưng sau đó bị thu hồi (do lộ private key).
- Nếu hệ thống không xác minh trạng thái chứng chỉ \rightarrow chữ ký vẫn “hợp lệ”.

Giải pháp:

- Gắn **timestamp** từ **TSA (Time Stamping Authority)**.
- Khi xác minh, kiểm tra CRL/OCSP để biết chứng chỉ có bị thu hồi không.

◆ c. Khóa yếu hoặc lộ khóa

Nguyên nhân:

- RSA 1024-bit có thể bị phá bằng tấn công phân tích số hiện nay.

- Hoặc người dùng lưu khóa cá nhân .pfx mà không đặt mật khẩu.

Hậu quả:

- Hacker ký giả mạo tài liệu hợp lệ → người khác tưởng là của người ký thật.

Giải pháp:

- Dùng **RSA ≥ 2048 bit** hoặc **ECDSA 256-bit** trở lên.
- Lưu private key trong **USB Token / HSM**, không lưu trên máy tính cá nhân.

5.4. Rủi ro vận hành (Operational Risks)

◆ **a. Xác minh sai cách**

- Người dùng mở file trong Chrome hoặc trình xem PDF mặc định → không kiểm tra chữ ký số.
- Thấy nội dung đúng nhưng không biết file đã bị đổi.

Giải pháp:

- Luôn xác minh bằng phần mềm chuyên dụng (Adobe Acrobat, DigiDoc, PyHanko verify).
- Nếu hiển thị “Document modified after signing” → coi chữ ký không hợp lệ.

◆ **b. Timestamp giả hoặc bị sửa**

- Một số công cụ không kiểm tra tính hợp lệ của timestamp.
- Hacker có thể chèn timestamp cũ để tạo cảm giác “ký đúng hạn”.

Giải pháp:

- Xác minh timestamp theo chuẩn **RFC 3161**.
- Chỉ chấp nhận timestamp từ **TSA có chứng chỉ hợp lệ**.

◆ **c. Tấn công bằng PDF Form Fields**

- PDF có thể chứa các trường nhập dữ liệu (form field) có thể sửa sau khi ký.
- Dù hash không đổi, người xem thấy nội dung khác.

Giải pháp:

- Khoá tất cả các trường form (readonly = true) trước khi ký.
- Xác minh rằng phần nội dung không thể chỉnh sau khi ký.

5.5. Các kỹ thuật tấn công PDF nổi tiếng đã công bố

Tên tấn công	Mô tả	Năm công bố	Ảnh hưởng
Incremental Save Attack	Thêm nội dung mới vào cuối PDF mà không thay đổi hash	2018	Sửa nội dung hợp đồng
Signature Wrapping Attack	Bao gói nội dung thật trong XML để lừa trình xác minh	2020	Giả mạo chữ ký
Shadow Attack (Ruhr Univ. Bochum)	Tạo PDF có 2 phiên bản nội dung, một ẩn – một hiện	2021	Đánh lừa người ký
Timestamp Forgery	Giả mạo dấu thời gian ký	2022	Làm sai lệch thời điểm pháp lý

5.6. Tổng kết rủi ro và hướng phòng tránh

Loại rủi ro	Mức độ	Giải pháp khuyến nghị
Tampering (sửa nội dung)	⚠ Cao	Hash toàn bộ file, kiểm tra ByteRange
Chứng chỉ giả / hết hạn	⚠ Cao	Dùng CA tin cậy, kiểm tra OCSP
Lộ private key	⚠ Rất cao	Dùng HSM, mật khẩu mạnh
Timestamp giả	⚠ Trung bình	Dùng TSA hợp lệ
Hiển thị sai (layer/form)	⚠ Trung bình	Flatten file trước khi ký