HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Nguyễn Anh Thư

NGHIÊN CỨU PHƯƠNG PHÁP TẤN CÔNG KÊNH BÊN KHÔNG LẬP MẪU VÀO THUẬT TOÁN AES-128 TRÊN THIẾT BỊ IOT

Chuyên ngành: **KỸ THUẬT VIỄN THÔNG**

Mã Số: 8.52.02.08

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

HÀ NỘI – NĂM 2025

Đề án tốt nghiệp được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. Đỗ Ngọc Tuấn

TS. Vũ Thị Thúy Hà

Phản biện 1: ……

Phản biện 2: ……

Đề án tốt nghiệp sẽ được bảo vệ trước Hội đồng chấm đề án tốt nghiệp thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ……giờ ……ngày ……tháng ……năm ……

Có thể tìm hiểu đề án tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

Sự bùng nổ của hệ sinh thái Internet vạn vật (IoT) trong các lĩnh vực trọng yếu như dân sự, công nghiệp và đặc biệt là quốc phòng đã thiết lập những tiêu chuẩn khắt khe về an toàn và bảo mật thông tin. Thực tế cho thấy, các thiết bị IoT hiện nay được tích hợp sâu rộng trong đa dạng hệ thống kỹ thuật với biên độ phức tạp lớn: từ các nút cảm biến quan trắc môi trường, giải pháp nhà thông minh, cho đến những hệ thống tự hành tinh vi như phương tiện bay không người lái (UAV), robot tự hành (UGV) hay xe tự lái.

Trong kiến trúc bảo mật của các hệ thống này, các giao thức mã hóa đóng vai trò tiên quyết nhằm đảm bảo tính xác thực và toàn vẹn dữ liệu. Trong đó, tiêu chuẩn mã hóa tiên tiến AES-128 (Advanced Encryption Standard) được ưu tiên triển khai rộng rãi nhờ sự cân bằng tối ưu giữa khả năng bảo mật mạnh mẽ và hiệu suất vận hành trên các thiết bị có tài nguyên phần cứng hạn chế (CPU, bộ nhớ, năng lượng). Đặc biệt, đối với các nền tảng di động chiến lược như UAV và UGV, việc triển khai AES-128 trở thành yêu cầu bắt buộc để bảo vệ dữ liệu truyền tải trong các môi trường truyền tin không tin cậy. Tuy nhiên, bên cạnh các mô hình tấn công lý thuyết vào thuật toán, sự phát triển của các kỹ thuật tấn công kênh bên đã làm xuất hiện các nguy cơ mới, khai thác trực tiếp tín hiệu vật lý sinh ra từ thiết bị trong quá trình xử lý dữ liệu thay vì phải tấn công vào thuật toán mật mã.

Trong số các kỹ thuật SCA, nhóm tấn công không lập mẫu (non-profiled SCA) ngày càng thu hút sự quan tâm nhờ khả năng triển khai trực tiếp trên thiết bị mục tiêu mà không cần giai đoạn huấn luyện phức tạp trên thiết bị tương đương. Điều này khiến các hệ thống IoT, vốn có cơ chế bảo vệ phần cứng không mạnh như thiết bị mật mã chuyên dụng, trở thành đối tượng đặc biệt dễ bị tổn thương. Vì vậy, việc nghiên cứu và đánh giá mức độ hiệu quả của các kỹ thuật tấn công kênh bên không lập mẫu vào AES-128 trên thiết bị IoT là cần thiết nhằm nhận diện nguy cơ,

xác định ngưỡng an toàn và đề xuất các biện pháp phòng vệ tương ứng.

Đề án gồm 3 chương với nội dung tóm tắt như sau:

CHƯƠNG 1: TỔNG QUAN VỀ TẤN CÔNG KÊNH BÊN (SCA)

Chương này cung cấp nền tảng lý thuyết về tấn công kênh bên, giới thiệu chung về khái niệm, cơ chế và mức độ nguy hiểm của SCA đối với các hệ mật mã hiện đại, đặc biệt trong bối cảnh phát triển mạnh mẽ của thiết bị IoT. Tiếp đó, chương trình bày tổng quan quy trình thu thập dữ liệu kênh bên (đặc biệt là dấu vết công suất tiêu thụ) trên thiết bị mục tiêu như bo mạch ChipWhisperer, từ đó phục vụ cho các bước phân tích trong các kỹ thuật tấn công. Cuối cùng, phân loại các phương pháp tấn công SCA thành hai nhóm lớn: phương pháp dựa trên mô hình thống kê truyền thống (như DPA, CPA) và phương pháp không lập mẫu dựa trên học máy, học sâu (như DDLA).

CHƯƠNG 2: MỘT SỐ PHƯƠNG PHÁP SCA KHÔNG LẬP MẪU ĐIỂN HÌNH

Chương này tập trung trình bày ba phương pháp tấn công kênh bên không lập mẫu điển hình, trong đó mỗi phương pháp đại diện cho một hướng tiếp cận khác nhau trong việc khai thác thông tin công suất tiêu thụ từ thiết bị thực thi thuật toán AES-128. Thông qua việc phân tích cơ chế, yêu cầu đầu vào, độ phức tạp và tính khả thi triển khai thực tế của từng phương pháp, chương này cung cấp cơ sở lý thuyết vững chắc để tiến hành đánh giá kết quả mô phỏng dựa trên các tham số đánh giá trong chương tiếp theo.

CHƯƠNG 3: MÔ PHỎNG, ĐÁNH GIÁ HIỆU QUẢ KỸ THUẬT SCA KHÔNG LẬP MẪU VÀO THUẬT TOÁN AES-128 TRÊN THIẾT BỊ IOT

Chương 3 trình bày toàn bộ quá trình mô phỏng, triển khai và đánh giá hiệu quả của hai nhóm kỹ thuật tấn công kênh bên

vào thuật toán AES-128 trên thiết bị IoT, bao gồm phương pháp tương quan công suất (CPA) và phương pháp học sâu không lập mẫu (DDLA). Các mô phỏng được thực hiện trên tập dữ liệu thực thu thập từ thiết bị thật với hai kịch bản: thiết bị không áp dụng cơ chế bảo vệ và thiết bị có tích hợp kỹ thuật che dấu thời gian (desynchronization). Các kết quả thu được cho phép đưa ra cái nhìn toàn diện về khả năng của từng phương pháp trong cả hai môi trường.

CHƯƠNG 1: TỔNG QUAN VỀ TẤN CÔNG KÊNH BÊN (SCA)

1.1. Giới thiệu chung

1.1.1. Tấn công các thiết bị mã mật

1.1.2. Tấn công kênh bên

1.2. Thu thập dữ liệu SCA trên thiết bị mục tiêu

1.2.1. Dữ liệu kênh bên và thiết lập đo mức tiêu thụ năng lượng

Khi triển khai thuật toán mật mã trên phần cứng, mỗi lần thực hiện một phép toán logic đều làm xuất hiện sự thay đổi về mức tiêu thụ năng lượng. Nếu các biến trung gian liên quan đến khóa bí mật được xử lý trong các bước này thì lượng tiêu thụ năng lượng sẽ có mối liên hệ với giá trị dữ liệu cực kỳ nhạy cảm. Đây chính là cơ sở để kẻ tấn công thu thập dữ liệu kênh bên và thực hiện phân tích nhằm truy xuất khóa

1.2.2. Tiêu thụ năng lượng trong mạch CMOS

Trong các thiết kế phần cứng thực hiện thuật toán mật mã, công nghệ CMOS chiếm vai trò chủ đạo nhờ ưu điểm tiết kiệm năng lượng và dễ tích hợp trên các nền tảng nhúng như vi điều khiển hay FPGA.

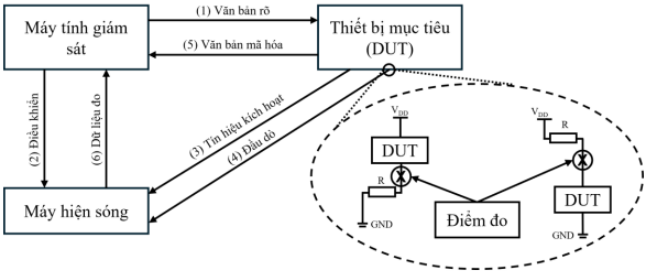
Điểm đáng chú ý của CMOS là năng lượng tiêu hao không chỉ xuất hiện khi mạch hoạt động theo chức năng logic thông thường, mà phần lớn năng lượng bị tiêu tốn trong quá trình

chuyển đổi trạng thái của tín hiệu. Khi một cổng logic thay đổi từ mức thấp sang mức cao (hoặc ngược lại), sẽ có dòng điện tạm thời chạy qua các transistor, làm phát sinh năng lượng động; ngược lại, khi tín hiệu ổn định tại một mức logic, mạch gần như không tiêu hao năng lượng đáng kể và chỉ tồn tại mức năng lượng tĩnh rất nhỏ.

Dựa trên nguyên tắc đó, các kỹ thuật tấn công kênh bên thường không quan tâm đến giá trị logic tại đầu ra, mà chủ yếu khai thác sự biến thiên năng lượng xuất hiện trong quá trình chuyển trạng thái. Ở giai đoạn chuyển trạng thái này, mỗi transistor của cổng CMOS sẽ tạo ra một lượng năng lượng khác nhau tùy thuộc vào dữ liệu đang xử lý, từ đó hình thành nên đặc trưng rò rỉ mà kẻ tấn công có thể theo dõi, đo đạc và phân tích.

1.2.3. Thiết lập đo mức tiêu thụ năng lượng

Để tấn công kênh bên dựa trên việc phân tích năng lượng tiêu thụ, cần phải đo được năng lượng tiêu thụ của thiết bị mục tiêu khi nó thực hiện thuật toán mã mật. Phần này trình bày quy trình đo mức tiêu thụ năng lượng bao gồm việc thiết lập và các thành phần cần thiết sử dụng cho quá trình đo.



Hình 1.3. Thiết lập để đo năng lượng tiêu thụ

1.3. Phân loại phương pháp SCA

1.3.1. Tấn công kênh bên lập mẫu

Tấn công lập mẫu thường được sử dụng khi biết đầy đủ các

thông tin về thiết bị mục tiêu cần tấn công. Trong trường hợp này, người tấn công sẽ sử dụng các thiết bị tham chiếu giống hệt (hoặc gần giống) với thiết bị mục tiêu để xây dựng cơ sở dữ liệu lưu trữ các thông tin về mức tiêu thụ năng lượng riêng cho từng thiết bị trước khi thực hiện tấn công.

1.3.2. Tấn công kênh bên không lập mẫu

Khác với tấn công lập mẫu, tấn công không lập mẫu được sử dụng khi không thể tiếp cận và có rất ít thông tin về hoạt động của thiết bị mục tiêu. Phương pháp này khai thác sự tương quan giữa mô hình tiêu thụ năng lượng và mức năng lượng tiêu thụ thực tế thu thập được từ thiết bị mục tiêu. Có nhiều phương pháp tấn công không lập mẫu khác nhau như phân tích vi sai năng lượng tiêu thụ (DPA) [11], phân tích tương quan năng lượng tiêu thụ (CPA) [6] hay phân tích vi sai tham số học sâu (DDLA) [23]… Tuy nhiên, tất cả các phương pháp này đều sử dụng một chiến lược chung để thực hiện tấn công đó là "chia để trị" gồm hai giai đoạn: giai đoạn xây dựng mô hình tiêu thụ năng lượng và giai đoạn tấn công.

Ngoài ra, tấn công không lập mẫu cũng phù hợp để đơn giản quá trình đánh giá bảo mật cũng như giảm chi phí quy trình kiểm tra của các thiết bị mã mật (do không cần thiết bị tham chiếu). Do vậy, đề án này tập trung vào nghiên cứu về kĩ thuật tấn công kênh bên không lập mẫu.

1.4. Kết luận chương 1

CHƯƠNG 2: MỘT SỐ PHƯƠNG PHÁP SCA KHÔNG LẬP MẪU ĐIỂN HÌNH

2.1. Phương pháp phân tích vi sai công suất tiêu thụ (DPA)

2.1.1. Chiến lược chung trong tấn công kênh bên không lập mẫu

Quy trình tấn công có thể được miêu tả như sau :

Bước 1: Tính toán giá trị trung gian giả thuyết của thuật toán mã mật

Bước 2: Tính toán các giá trị tiêu thụ năng lượng giả thuyết

Bước 3: Thu thập dữ liệu về năng lượng tiêu thụ của thiết bị mã mật.

Bước 4: Xác định khóa chính xác

2.1.2. Phân tích vi sai công suất tiêu thụ (DPA)

Phân tích vi sai năng lượng tiêu thụ (DPA) là phương pháp thống kê dùng để phân tích một tập hợp các giá trị năng lượng tiêu thụ đo được để xác định các mối tương quan phụ thuộc vào dữ liệu. Cơ sở để thực hiện tấn công DPA là mức tiêu thụ năng lượng khi chuyển một bit về trạng thái cao (bit 1) sẽ khác với mức năng lượng cần tiêu thụ khi chuyển bit đó về trạng thái thấp (bit 0). Tại đầu ra của thiết bị mục tiêu, nếu thuật toán mã mật tạo ra một bit 1 thì mức tiêu thụ năng lượng của thiết bị đó có thể sẽ lớn hơn mức tiêu thụ năng lượng khi tạo ra một bit 0. Tuy nhiên, sự khác biệt về mức tiêu thụ năng lượng này rất khó để quan sát được bằng mắt thường. Chính vì vậy, cần phải thu thập một lượng lớn các vệt công suất để tính toán giá trị tiêu thụ năng lượng trung bình của thiết bị. Trên cơ sở đó, tấn công DPA thực hiện phân chia các vệt công suất thành các tập hợp con, sau đó tính toán sự khác biệt về giá trị trung bình của các tập hợp con này. Nếu một tập hợp con chứa các vệt công suất không có sự tương quan về mức năng lượng tiêu thụ thì việc tính toán giá trị tiêu thụ năng lượng trung bình của các vệt công suất trong tập hợp con này sẽ tiến về 0 khi số lượng các vệt công suất đủ lớn.

Ngược lại, nếu tập hợp con chứa các vệt công suất có tương quan về giá trị năng lượng tiêu thụ, giá trị trung bình sẽ tiến dần đến một giá trị khác 0. Chỉ cần thu thập đủ số lượng vệt công suất, một sự tương quan rất nhỏ cũng có thể được phát hiện ra kể cả khi có sự xuất hiện của nhiễu trong phép đo.

2.2. Phương pháp phân tích tương quan công suất tiêu thụ (CPA)

Mục tiêu của CPA là xây dựng được một mô hình tiêu thụ năng lượng chính xác của thiết bị mục tiêu và tìm ra được sự tương quan giữa giá trị tiêu thụ năng lượng giả thuyết với giá trị tiêu thụ năng lượng thực tế. Với số lượng vệt công suất thu thập được đủ lớn, khóa chính xác có thể được xác định dựa trên việc quan sát khóa giả thuyết nào tạo ra mức độ tương quan lớn nhất. Các mô hình tiêu thụ năng lượng thường được sử dụng trong CPA là mô hình trọng số Hamming hoặc khoảng cách Hamming.

Phương pháp thường được sử dụng để tính toán mức độ tương quan giữa mô hình tiêu thụ năng lượng và mức tiêu thụ năng lượng thực tế là ước lượng hệ số tương quan Pearson. Giả sử quá trình tấn công thu thập được D vệt công suất, mỗi vệt cộng suất bao gồm N mẫu. Gọi $t_{i,j}$ là giá trị tiêu thụ năng lượng tại mẫu j của vệt công suất thứ i với ($1 \leq i \leq D$ và $1 \leq j \leq N$). Gọi $h_{i,k}$ là giá trị tiêu thụ năng lượng giả thuyết của vệt công suất thứ i tương ứng với khóa con k (với $0 \leq k \leq K$ và K là số lượng các khóa giả thuyết có thể có với mỗi byte của khóa bí mật). Khi đó, hệ số tương quan Pearson giữa mức tiêu thụ năng lượng giả thuyết và mức tiêu thụ năng lượng thực tế tại mẫu thứ j của vệt công suất được tính như sau:"

$$r_{k,j} = \frac{\sum_{i=1}^{D}(h_{i,j} - \overline{h_k})(t_{i,j} - \overline{t_j})}{\sqrt{\sum_{i=1}^{D}(h_{i,j} - \overline{h_k})^2 \sum_{i=1}^{D}(t_{i,j} - \overline{t_j})^2}} \quad (2.6)$$

Trong đó, $\overline{h_k}$ và $\overline{t_j}$ lần lượt là các giá trị trung bình của mô hình tiêu thụ năng lượng và mức năng lượng tiêu thụ thực tế thu thập được tại thời điểm xuất hiện của mẫu thứ j trong vệt công suất.

Hệ số tương quan Pearson được tính cho mỗi giá trị của k

và j, kết quả tính toán sẽ là ma trận hệ số tương quan $R = r_{1\ldots K,1\ldots N}$. Cuối cùng, bằng cách lấy giá trị lớn nhất của $r_{k,j}$ có thể tìm ra khóa chính xác từ tập hợp các khóa giả thuyết. Trong thực tế, để xây dựng các chương trình tính toán hiệu quả hơn, cho phép tính toán trên các dữ liệu từ vệt công suất mới thêm vào mà không cần phải tính lại các giá trị đã xử lý trước đó (tức là mỗi giá trị trong tập dữ liệu đầu vào chỉ cần duyệt qua một lần trong quá trình tính toán.

2.3. Phương pháp phân tích vi sai tham số mô hình học sâu (DDLA)

2.3.1. Học sâu trong tấn công kênh bên

2.3.2. Mô hình phân tích vi sai tham số học sâu

Một trong những phương pháp tấn công kênh bên không lập mẫu sử dụng mô hình hình học sâu đầu tiên được đề xuất bởi Timon trong [23] được gọi là phân tích vi sai tham số học sâu (DDLA). Ý tưởng cốt lõi của phương pháp này là chia các vệt công suất $(T_i)_{1\le i\le N}$. thu thập được từ thiết bị mục tiêu thành các nhóm tương tự như phương pháp DPA. Tuy nhiên, thay vì tính toán sự khác biệt giữa các vệt công suất trong từng nhóm, phương pháp này thực hiện phân loại các nhóm vệt công suất sử dụng mô hình học sâu. Quá trình tấn công có thể được miêu tả như sau. Đầu tiên, hàm mục tiêu của thuật toán cần tấn công sẽ được xác định. Phương pháp này xác định hàm mục tiêu là $Sbox(d_i \oplus k_j)$, với $(d_i)_{1\le i\le N}$ là các giá trị ngẫu nhiên đã biết (văn bản rõ đầu vào của thuật toán mã hóa) và $k_j \in \mathcal{K}$ là các giá trị khác nhau của khóa giả thuyết. Với mỗi khóa giả thuyết, các giá trị trung gian giả thuyết được tính toán sử dụng một hàm phân loại h. Hàm phân loại h trong phương pháp này được định nghĩa là $HW\left(Sbox\left(d_i \oplus k_j\right)\right)$. Các giá trị trung gian $h_{i,j} = HW\left(Sbox\left(d_i \oplus k\right)\right)$ sau khi tính toán được sử dụng để làm các nhãn phân loại. Tiếp theo, các vệt công suất T_i được chia vào các

nhóm dựa vào giá trị của nhãn $h_{i,j}$. Quá trình này chính là quá trình gán nhãn dữ liệu trong học sâu. Sau đó, các dữ liệu huấn luyện T_i cùng với nhãn phân loại $h_{i,j}$ tương ứng được đưa vào huấn luyện trong một mô hình học sâu Net. Khi khoá đúng k_{cr}được sử dụng trong huấn luyện thì chuỗi các giá trị trung gian (nhãn) $h_{i,k_{cr}}$ được tính toán chính xác và khi đó, việc phân loại vệt công suất thu thập được và các nhãn dùng để huấn luyện mô hình học sâu sẽ khớp với nhau. Ngược lại, với các khoá giả thuyết $k \neq k_{cr}$ thì các nhãn dùng cho mô hình huấn luyện sẽ không khớp với các vệt công suất. Việc tìm ra giá trị của khoá chính xác k_{cr} sẽ được thực hiện thông qua việc sử dụng các thước đo để đánh giá khoá nào làm cho mô hình huấn luyện đạt kết quả tốt nhất. Các thước đo (loss và accuracy) thường sử dụng để đánh giá hiệu quả của mô hình học sâu là các trọng số thu được ở đầu ra của mô hình huấn luyện. Thuật toán 1 tóm tắt quá trình tấn công sử dụng phương pháp DDLA.

Thuật toán 1: Phân tích vi sai tham số học sâu (DDLA) [23]

Đầu vào: N vệt công suất $(T_i)_{1\leq i\leq N}$, các văn bản rõ tương ứng $(d_i)_{1\leq i\leq N}$, các khóa giả thuyết $k_j \in \mathcal{K}$. Mạng học sâu *Net* và số lần huấn luyện n_e.

Đầu ra $k_{cr} \in \mathcal{K}$

1: Thiết lập dữ liệu huấn luyện là $X = (T_i)_{1\leq i\leq N}$

2: **for** $k_j \in \mathcal{K}$ **do**

3: khởi tạo lại các tham số huấn luyện của *Net*

4: Tính toán các chuỗi của giá trị khóa giả thuyết $(h_{i,j})_{1\leq i\leq N}$

5: Thiết lập các nhãn huấn luyện là $Y_i = (h_{i,j})_{1\leq i\leq N}$

6: Thực hiện huấn luyện: $DL\ (Net, X, Y_i, n_e\ )$

7: **end for**

8: return khóa k_{cr} tạo ra thước đo huấn luyện tốt nhất.

Khi sử dụng DDLA trong tấn công kênh bên không lập mẫu, các thước đo huấn luyện thu được trên tập dữ liệu được xây dựng từ khóa chính xác sẽ cao hơn do dữ liệu đầu vào được phân loại khớp với các nhãn. Đối với các khóa không chính xác, các nhãn dùng cho huấn luyện sẽ không khớp với dữ liệu đầu vào nên các thước đo huấn luyện sẽ cho ra kết quả thấp. Theo tác giả trong [23], DDLA có thể được thực hiện trên bất kỳ mô hình học sâu nào. Điều này tạo nên sự linh hoạt trong quá trình tấn công do có thể thay đổi các kiến trúc của mô hình dựa trên thiết bị và hoạt động cần nhắm tới. Tuy nhiên, theo tác giả thì kiến trúc MLP thường được sử dụng khi các vệt công suất thu thập được có sự đồng bộ với nhau do kiến trúc này đủ tốt để đạt được kết quả mong muốn. Kiến trúc CNN chủ yếu được sử dụng khi tấn công các mục tiêu có sử dụng biện pháp chống tấn công kênh bên là tạo nên sự bất đồng bộ giữa các vệt công suất.

2.4. Kết luận chương 2

CHƯƠNG 3: MÔ PHỎNG, ĐÁNH GIÁ HIỆU QUẢ KỸ THUẬT SCA KHÔNG LẬP MẪU VÀO THUẬT TOÁN AES-128 TRÊN THIẾT BỊ IOT

3.1. Tấn công SCA sử dụng CPA

3.1.1. Mục tiêu và cách tiếp cận

CPA được sử dụng để đánh giá khả năng thu hồi khóa AES-128 trên thiết bị IoT từ dữ liệu rò rỉ công suất. Đây là phương pháp không lập mẫu truyền thống, dựa trên tương quan giữa mô hình công suất dự đoán và dấu vết công suất thực tế.

3.1.2. Xử lý dữ liệu và tạo tập dataset

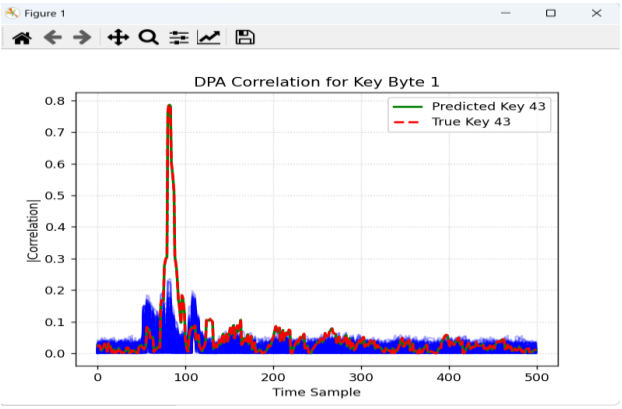
3.1.3. Thực hiện CPA trên thuật toán AES-128

Tấn công CPA được thực hiện bằng Python với các thư viện:

NumPy, SciPy, Matplotlib và ChipWhisperer API. Thuật toán CPA về bản chất có thể thực hiện trên một lượng lớn vệt công suất, tuy nhiên độ phức tạp tính toán sẽ tỉ lệ thuận với số lượng mẫu của mỗi vệt công suất.

3.1.4. Kết quả thực nghiệm CPA

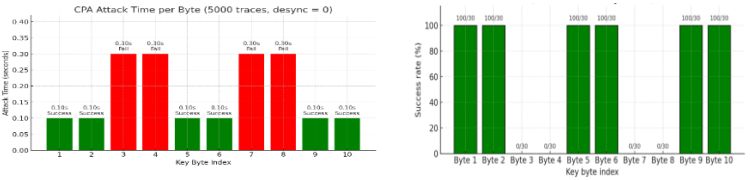
a) Tấn công thiết bị không được bảo vệ



Hình 3. 4. Kết quả mô phỏng tấn công SCA sử dụng CPA vào byte 1 với các thiết bị không được bảo vệ tấn công kênh bên

Tiếp tục khảo sát với các byte tiếp theo chỉ ra rằng các byte 1, 2, 5, 6, 9, 10 được xác định chính xác giá trị khóa, trong khi các byte còn lại thì không. Dựa trên mô hình đã được huấn luyện sẵn khi tấn công byte 1 với 5000 vệt công suất, mô hình CPA có thể tìm ra khóa chính xác khi tấn công các byte 2, byte 5, byte 6, byte 9, byte 10 của khóa bí mật trong khoảng 100-500 lượt huấn luyện đầu tiên.

Thực nghiệm tiếp theo được thực hiện bằng cách lặp lại quy trình tấn công đã miêu tả ở trên 30 lần nhằm đánh giá hiệu quả về thời gian tấn công và tỉ lệ tấn công thành công của mô hình tấn công SCA sử dụng CPA.

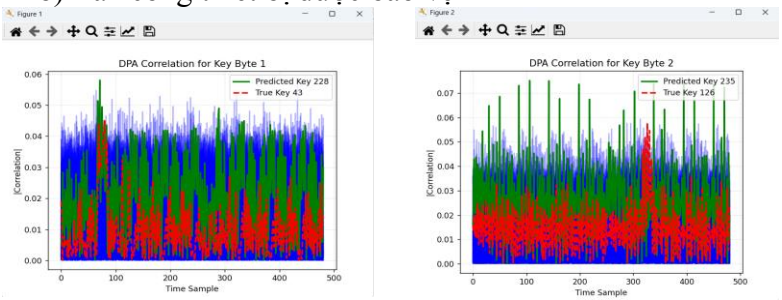


a) b)

Hình 3.6. Thời gian tấn công trung bình (a) và tỷ lệ tấn công thành công (b) của mô hình tấn công SCA sử dụng CPA

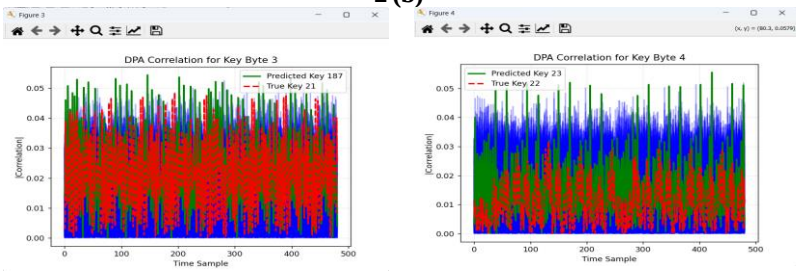
CPA hoàn toàn phát huy tác dụng với thời gian thực thi nhanh chóng và ổn định ≈ 0,10s với tỷ lệ tin cậy cho khóa chính xác là 100%. Nhìn chung, CPA hiệu quả trong điều kiện lý tưởng.

b) Tấn công thiết bị được bảo vệ



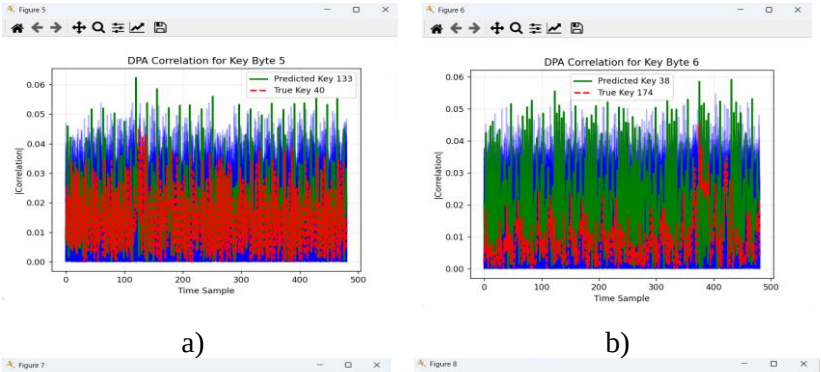
a) b)

Hình 3.7. Kết quả mô phỏng tấn công SCA sử dụng CPA vào byte 1 (a), Byte 2 (b)

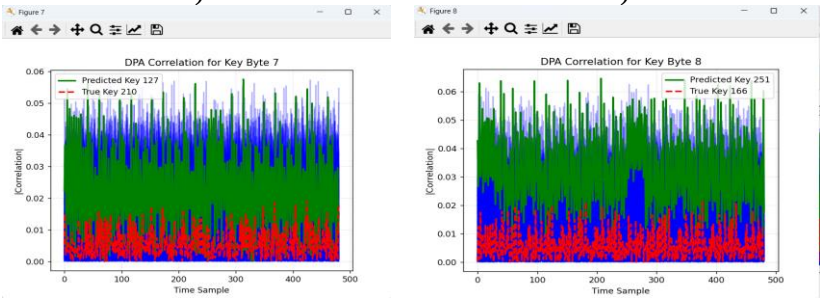


a) b)

Hình 3.8. Kết quả mô phỏng tấn công SCA sử dụng CPA vào Byte 3 (a), Byte 4 (b) với các thiết bị được bảo vệ tấn công kênh bên

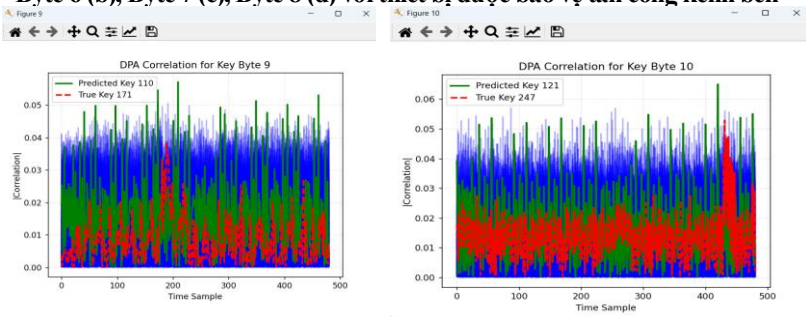


a) b)



c) d)

Hình 3. 9. Kết quả mô phỏng tấn công SCA sử dụng CPA vào Byte 5 (a), Byte 6 (b), Byte 7 (c), Byte 8 (d) với thiết bị được bảo vệ tấn công kênh bên



a) b)

Hình 3.10. Kết quả mô phỏng tấn công SCA sử dụng CPA vào Byte 9 (a), Byte 10 (b) với thiết bị được bảo vệ tấn công kênh bên

Kết quả: Trong trường hợp vệt công suất đã được desync, kết quả CPA gần như thay đổi hoàn toàn. Không còn xuất hiện đỉnh tương quan lớn. Các đường khóa thật/sai chồng lẫn, biên độ nhỏ (< 0.03). Tương quan phân bố ngẫu nhiên, không thể phân biệt được khóa đúng.

3.2. Tấn công SCA sử dụng DDLA

3.2.1. Mục tiêu và cách tiếp cận

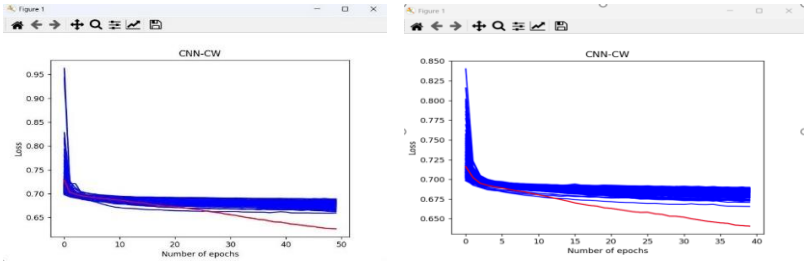
Sự thất bại của CPA ở mục 3.1 trước cơ chế bảo vệ dẫn đến nhu cầu sử dụng các mô hình học sâu nhằm trích xuất đặc trưng rò rỉ phi tuyến mà không phụ thuộc vào vị trí thời gian. DDLA chính là hướng tiếp cận này.

3.2.2. Xử lý dữ liệu và xây dựng dataset cho DDLA

3.2.3. Thực thi DDLA bằng Python

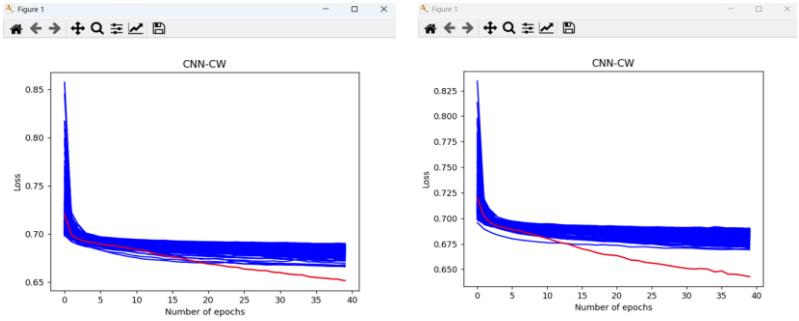
Quy“trình tấn công: Quy trình thực nghiệm cũng được tiến hành như các thực nghiệm đã được mô tả ở các phần trước, sử dụng DDLA để tấn công byte 1 của khóa bí mật, với 5000 vệt công suất, sau đó mô hình được sử dụng để tấn công ngẫu nhiên các byte khóa còn lại cụ thể là: byte 2 và byte 9 và byte 10.

3.2.4. Kết quả thực nghiệm DDLA



a) **b)**

Hình 3.11. Kết quả tấn công Byte 1 (a); Byte 2 (b)



a) **b)**

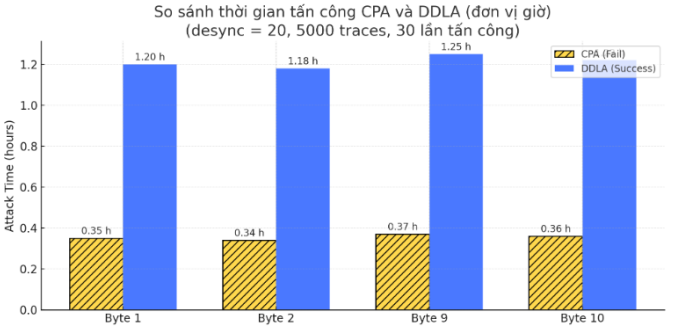
Hình 3.12. Kết quả khi tấn công Byte 9 (a); Byte 10 (b)

Nhìn vào Hình 3.11b; Hình 3.12 kết quả cho thấy các đường xanh (khóa sai) sau giai đoạn đầu ổn định quanh ~0.68–0.70. Đường đỏ (khóa thật - byte 2-10) ban đầu không khác biệt, nhưng từ epoch 8–10 trở đi, bắt đầu tách ra phía dưới và tiếp tục giảm đều. Điều này cho thấy xu thế chung giống byte 1.

Khoảng cách giữa hai nhóm khóa tăng dần theo thời gian. Mô hình bắt đầu phân tách rõ từ epoch thứ 10 và hội tụ ổn định sau epoch thứ 30. Sự phân biệt rõ ràng sau một số chu kỳ huấn luyện cho thấy mô hình CNN trong thí nghiệm đã tối ưu được trọng số của mô hình. Trọng số này giúp mô hình thể hiện được mối liên hệ giữa giá trị nhãn được tính toán bằng khóa đúng và dữ liệu đầu vào, lần lượt là mô hình công suất tiêu thụ LSB và các vệt công suất thu được trong quá trình mã hóa. Ngược lại, với các khóa sai giá trị nhãn và dữ liệu đầu vào không có mối liên hệ ràng buộc nhất định, vì vậy mô hình không thể tối ưu được trọng số, chính vì vậy loss chỉ có xu hướng giảm trong vài epoch đầu, sau đó gần như không đổi. Đây cũng chính là đặc trưng quan trọng của kỹ thuật DDLA giúp xác định khóa bí mật với các dữ liệu kênh bên phức tạp điều mà CPA không thể làm.

Byte 1, 2, 9, 10 đều cho hình dạng đồ thị tương tự, chứng tỏ mô hình không "overfit" vào một byte cụ thể, mà học được cơ chế chung của rò rỉ SBOX trên thiết bị. Thời gian thực thi trung

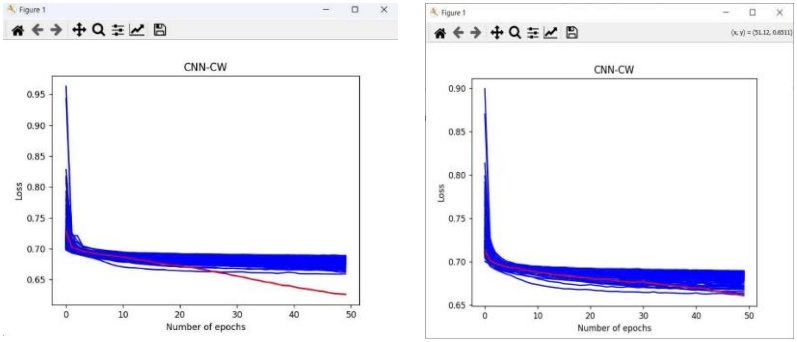
bình và tỉ lệ tìm ra khóa chính xác của các mô hình cũng được tính toán sau khi quy trình tấn công được lặp lại 30 lần. Các kết quả thực nghiệm cho thấy, tất cả các mô hình được sử dụng trong thực nghiệm này đều có khả năng phân biệt khóa chính xác trong tập hợp các khóa giả thuyết sau khoảng 40 lượt huấn luyện.



Hình 3.13. So sánh thời gian tấn công trung bình và tỷ lệ tấn công thành công của CPA và DDLA

Kết quả: DDLA tìm được khóa bí mật điều mà CPA không làm được. Tiếp tục thực nghiệm với 30 lần tấn công DDLA đều cho ra khóa bí mật. Mặc dù cần nhiều thời gian để huấn luyện, nhưng kết quả khôi phục khóa thành công và hoàn toàn ổn định sau 30 lần thử điều này hiệu quả hơn CPA. Như vậy, có thể thấy rằng DDLA vượt CPA tuyệt đối về độ tin cậy trong môi trường IoT có bảo vệ chống căn chỉnh vệt công suất, mặc dù DDLA tốn thời gian hơn, nhưng luôn thu được khóa, CPA chỉ phù hợp trong môi trường *không bảo vệ, không nhiễu,* vốn không còn phổ biến trong thiết bị IoT hiện đại.

Thực nghiệm tiếp theo được thực hiện bằng cách thay đổi mức xáo trộn đạt đến 50 mẫu, nhằm khảo sát mức độ giới hạn hoạt động DDLA không còn hiệu quả trong điều kiện chống tấn công mạnh hơn bằng cách phân tích giới hạn phòng thủ với trace shift = 50. Kết quả, DDLA thất bại – không xác định được khóa chính xác khi mức xáo trộn đạt đến 50 mẫu.



a) Thực hiện xáo trộn 20 mẫu

b) Thực hiện xáo trộn 50 mẫu

Hình 3.14. So sánh khả năng tấn công trong điều kiện chống tấn công mạnh

3.3. Phân tích, đánh giá hiệu quả

3.3.1. So sánh CPA và DDLA theo tiêu chí hiệu năng

Đối với CPA, khi áp dụng trên thiết bị không được bảo vệ, phương pháp này thể hiện đặc trưng rất mạnh về tốc độ và tính trực tiếp. Chỉ với 5000 vệt công suất, CPA có thể thực hiện tính toán hệ số tương quan giữa mô hình rò rỉ và tín hiệu thu được mà không cần huấn luyện hay tối ưu thêm. Tuy nhiên, ưu thế của CPA chỉ duy trì được khi thiết bị không có cơ chế phòng vệ. Khi các vệt công suất bị xáo trộn theo dạng desynchronization, masking hoặc được bổ sung nhiễu nhằm làm mất đồng bộ thời gian, CPA gần như hoàn toàn bị vô hiệu hóa.

Ngược lại, kỹ thuật DDLA dựa trên mạng học sâu cho thấy cách thức xử lý hoàn toàn khác. Thay vì khai thác trực tiếp sự phụ thuộc tuyến tính giữa mô hình rò rỉ và vệt công suất như CPA, DDLA học chính xác cấu trúc thống kê – hình thái – đặc trưng rò rỉ trong dữ liệu bằng quá trình huấn luyện nhiều epoch. Dù thời gian tấn công dài hơn, đồ thị loss của mô hình qua từng epoch thể hiện rất rõ khả năng phân tách hai lớp (khóa đúng và khóa sai). Ở thời điểm tổng hợp kết quả, có thể khẳng định rằng

CPA phù hợp trong môi trường lý tưởng, nơi thiết bị không có biện pháp phòng thủ, trong khi DDLA phù hợp cho môi trường thực tế hơn, nơi các cơ chế bảo vệ gần như là yếu tố mặc định trên thiết bị IoT hiện đại.

3.3.2. So sánh với các nghiên cứu trước tại Việt Nam

a) Những ưu điểm của đề án

Phần lớn các nghiên cứu SCA ở Việt Nam chủ yếu thử nghiệm trên các bộ dữ liệu chuẩn đã được căn chỉnh sẵn, như ChipWhisperer Basic, vốn ít nhiễu và hầu như không gây khó khăn cho các thuật toán kinh điển như DPA hoặc CPA. Trái lại, đề án triển khai thu thập dữ liệu trong điều kiện "thực" hơn: các vệt công suất bị mất đồng bộ, có độ dịch chuyển thay đổi theo từng trace và chứa nhiễu lớn – đúng với cơ chế bảo vệ được tích hợp trong nhiều thiết bị IoT hiện đại nhằm giảm thiểu rủi ro rò rỉ kênh bên. Việc khôi phục thành công các byte của khóa AES-128 trong một môi trường nhiễu cao như vậy không chỉ chứng minh tính khả thi của tấn công phi mẫu, mà còn cho thấy hiệu quả vượt trội của phương pháp DDLA so với các hướng tiếp cận truyền thống.

Đề án còn mở rộng hướng khảo sát hơn các nghiên cứu đã công bố trong nước trước đây – chính là cách tiếp cận thực nghiệm toàn diện, không chỉ nghiên cứu khả năng tấn công DDLA so với CPA, DPA mà còn đánh giá giới hạn hoạt động DDLA không còn hiệu quả trong điều kiện chống tấn công mạnh hơn bằng cách phân tích giới hạn phòng thủ. Đây là kết quả có giá trị thực tiễn, giúp các nghiên cứu tiếp theo đánh giá đúng tiềm năng và hạn chế của DDLA.

Thay vì đánh giá một vài byte đơn lẻ như nhiều công bố hiện hành, đề án tiến hành tấn công liên tiếp trên nhiều byte khóa, theo dõi đầy đủ đồ thị hội tụ của hàm loss, phân tích độ ổn định giữa các lần huấn luyện và so sánh định lượng giữa các kỹ thuật DPA, CPA và DDLA. Quy mô dữ liệu 5000 traces/byte cũng lớn hơn đáng kể so với nhiều nghiên cứu trong nước, thường chỉ sử

dụng vài trăm đến một nghìn traces, giúp kết quả đánh giá có độ tin cậy và khả năng khái quát tốt hơn.

Đề án không chỉ mang tính kế thừa mà còn kết hợp tốt các phương pháp, mức độ khó khăn kỹ thuật và khả năng ứng dụng thực tế. Việc triển khai tấn công trong môi trường nhiễu cao, kết hợp mô hình CNN xử lý trực tiếp traces chưa đồng bộ hóa, cùng quy trình đánh giá đa dạng theo từng byte và theo từng kỹ thuật, giúp đề án có ưu điểm hơn so với mặt bằng chung của các nghiên cứu SCA tại Việt Nam hiện nay.

b) Những điểm còn hạn chế

Mặc dù mô hình DDLA cho thấy khả năng vượt trội trong việc tấn công thiết bị có cơ chế bảo vệ, đề án vẫn tồn tại một số hạn chế nhất định cần được nhìn nhận thẳng thắn. Trước tiên, do bản chất dữ liệu thực nghiệm được thu thập trong môi trường IoT có nhiều yếu tố ngẫu nhiên, mỗi byte khóa trong AES-128 được mô phỏng tấn công theo phương thức ngẫu nhiên và phải được chạy lặp lại ít nhất 30 lần để ước lượng trung bình thời gian thực thi và số epoch hội tụ. Cách làm này đảm bảo độ tin cậy của kết quả, tuy nhiên cũng đồng nghĩa với việc khối lượng, thời gian, chi phí tính toán tăng lên đáng kể.

Mặc dù kết quả của các byte được trình bày (như Byte 1, 2, 9, 10) cho thấy sự phân tách loss rất rõ ràng, các byte còn lại trong thí nghiệm - vốn được xử lý theo quy trình hoàn toàn tương tự - vẫn có thể dao động về tốc độ hội tụ tùy theo mức độ nhiễu hiện diện trong từng nhóm vệt công suất. Điều này là đặc trưng tự nhiên của tấn công kênh bên trên thiết bị có bảo vệ vì từng byte khóa có thể bị tác động khác nhau bởi mức độ mất đồng bộ và nhiễu phần cứng. Trong đề án, các byte còn lại vẫn cho kết quả khả quan nhưng độ sâu phân tách loss không hoàn toàn đồng nhất giữa các byte. Việc phải chạy lặp lại nhiều lần để lấy trung bình càng cho thấy rằng học sâu trong SCA dù hiệu quả nhưng vẫn chịu ảnh hưởng mạnh bởi chất lượng tập dữ liệu.

Bên cạnh đó, việc đánh giá trong đề án chủ yếu dựa trên 5000 vệt công suất mỗi byte. Mặc dù đây là mức dữ liệu tương đối lớn so với nhiều nghiên cứu trong nước, nhưng trong thực tế một số thiết bị IoT có thể giới hạn số lần đo hoặc tích hợp các cơ chế che dấu mạnh khiến số lượng vệt thu thập bị hạn chế đáng kể. Do đó, đề án bỏ ngỏ khả năng trong các môi trường khắt khe hơn.

Cuối cùng, việc tấn công được thực hiện theo phương thức ngẫu nhiên trên các byte độc lập cho phép đánh giá cục bộ rất tốt, nhưng chưa khai thác được mối quan hệ giữa các byte trong cấu trúc AES, chẳng hạn tính tuần tự của các vòng SubBytes hay phân bố rò rỉ giữa các byte. Một mô hình đa đầu ra hoặc mô hình học liên byte có thể rút ngắn thời gian huấn luyện đáng kể so với việc huấn luyện riêng rẽ cho từng byte 30 lần như hiện tại.

3.4. Kết luận chương 3

III. KẾT LUẬN

Đề án "Nghiên cứu phương pháp tấn công kênh bên không lập mẫu vào thuật toán AES-128 trên thiết bị IoT" đã tập trung giải quyết một vấn đề quan trọng trong an toàn thông tin hiện đại, nơi các thiết bị IoT với khả năng tính toán hạn chế nhưng được triển khai rộng rãi lại trở thành mục tiêu dễ bị khai thác thông qua các kênh rò rỉ vật lý. Trên cơ sở tổng hợp lý thuyết về tấn công kênh bên, phân tích cơ chế rò rỉ năng lượng trong AES-128 và các chiến thuật đối phó phổ biến, đề án đã triển khai các mô phỏng tấn công thực nghiệm dựa trên dữ liệu đo thực tế. Từ đó đưa ra đánh giá toàn diện giữa phương pháp tấn công truyền thống CPA và kỹ thuật học sâu DDLA - đại diện cho xu hướng tấn công không lập mẫu mới.

Toàn bộ kết quả thu được đã minh chứng rằng, trong môi trường IoT hiện đại vốn tích hợp nhiều cơ chế che giấu nhằm làm suy yếu mối liên hệ giữa vệt công suất và giá trị trung gian của AES, các phương pháp truyền thống phụ thuộc vào mức độ

đồng bộ như CPA bộc lộ rõ những hạn chế cốt hữu. Ngược lại, DDLA với khả năng học đặc trưng phi tuyến và bù trễ thời gian đã chứng minh được hiệu quả vượt trội trong cả điều kiện nhiễu mạnh. Những điều này không chỉ cho thấy sự cần thiết của học sâu trong phân tích kênh bên, mà còn chỉ ra hướng phát triển tất yếu của lĩnh vực này trong tương lai.

1. Các kết quả đạt được

Đề án đã hoàn thành các mục tiêu đề ra và đạt được một số kết quả như sau:

Xây dựng được cơ sở lý thuyết đầy đủ về tấn công kênh bên không lập mẫu: Đề án đã hệ thống hóa rõ ràng nguyên lý rò rỉ công suất, phân loại các kỹ thuật SCA, mô hình hóa quá trình tấn công AES-128 và phân tích sâu hai phương pháp đại diện: CPA và DDLA. Đây là nền tảng giúp các thí nghiệm mô phỏng được thực hiện chính xác và có tính học thuật cao.

Xử lý thành công các tập dữ liệu kênh bên công khai và xây dựng dataset chuẩn cho mô phỏng: Các tập dữ liệu đã được xử lý theo quy trình chuẩn nhằm tạo ra dataset sử dụng được cho cả CPA và DDLA.

Triển khai hai kỹ thuật CPA và DDLA bằng ngôn ngữ Python:

Đề án đã xây dựng đầy đủ pipeline tấn công SCA bao gồm:

+ Tính toán giá trị trung gian của AES-128;

+ Mô hình công suất và tính tương quan (CPA);

+ Tính toán điểm khóa và hợp nhất kết quả để thu hồi khóa chính xác.

Những mô phỏng này cho phép đánh giá định lượng hiệu quả của từng phương pháp.

Với CPA trên thiết bị không có bảo vệ, các byte rò rỉ tốt cho thấy đỉnh tương quan nổi bật đúng tại khóa thật. Đây là bằng chứng cho thấy CPA vẫn phù hợp trong môi trường lý tưởng hoặc thiết bị không có đối pháp.

CPA trên thiết bị có cơ chế desynchronization, kết quả thực nghiệm chỉ ra sự sụt giảm mạnh mẽ của tương quan, hầu như không thể tìm thấy đỉnh spike - khẳng định CPA hoàn toàn mất tác dụng trong môi trường IoT bảo vệ.

DDLA trên thiết bị có bảo vệ**,** mô hình học sâu cho thấy khả năng tách biệt khóa đúng ngay cả khi dữ liệu bị nhiễu và dịch pha mạnh. Nhiều byte khóa được tấn công (byte 1, 2, 9, 10 và các byte khác trong các lần thử) đều cho thấy đường loss của khóa đúng hội tụ nhanh và thấp hơn hẳn các khóa sai. Từ đó, cho thấy học sâu có thể "học" rò rỉ mà không cần mô hình hóa thủ công. Tuy nhiên DDLA cũng có giới hạn và chỉ duy trì hiệu quả tấn công trong phạm vi nhất định của mức độ làm lệch vệt công suất và bắt đầu thất bại khi cơ chế desynchronization được tăng cường đến mức đủ lớn.

Các giá trị trung bình về thời gian thực thi và số lần huấn luyện được thống kê chính xác, tăng độ tin cậy của đánh giá.

Kết quả cuối cùng cho thấy DDLA đạt hiệu năng cao hơn CPA rõ rệt, đặc biệt trong môi trường bảo vệ và đây cũng chính là mục tiêu trọng tâm của đề án.

2. Nhận xét, đề xuất và khuyến nghị

Từ các mô phỏng và phân tích thực nghiệm, có thể đưa ra một số nhận xét quan trọng:

Thứ nhất, CPA là kỹ thuật nền tảng, dễ triển khai, nhưng có độ nhạy rất cao với nhiễu và trễ thời gian. Chỉ cần một mức độ desynchronization nhỏ cũng đủ làm triệt tiêu mối liên hệ giữa mô hình Hamming Weight và vệt công suất thực tế. Do đó CPA chỉ phù hợp để giảng dạy, kiểm thử thuật toán, hoặc đánh giá các

thiết bị không có lớp bảo vệ phần cứng.

Thứ hai, DDLA cho thấy khả năng khái quát hóa tốt hơn bởi nó không dựa vào mô hình toán cứng mà học trực tiếp rò rỉ năng lượng. Tuy nhiên, mô hình học sâu đòi hỏi tài nguyên tính toán và thời gian huấn luyện lớn hơn, nhất là khi tấn công nhiều byte độc lập. Đây là yếu tố cần cân nhắc khi triển khai thực tế.

Thứ ba, kết quả tấn công cho thấy với 5000 vệt công suất, phương pháp học sâu hoàn toàn có khả năng vượt qua lớp bảo vệ desync, điều này gợi ý rằng các cơ chế bảo vệ hiện nay trên nhiều thiết bị IoT giá rẻ chỉ mang tính chất phòng thủ tối thiểu. Vì vậy, cần xem xét nâng cấp các cơ chế chống tấn công như: random delay nhiều cấp độ, masking mức độ cao, hoặc kết hợp nhiều lớp bảo vệ thay vì chỉ phụ thuộc vào một kỹ thuật đơn lẻ.

Cuối cùng, trong bối cảnh chuẩn IoT tiếp tục mở rộng và thiết bị tiêu dùng ngày càng phổ biến, việc nghiên cứu và đánh giá tính an toàn trước các kỹ thuật SCA là yêu cầu cấp thiết. Các tổ chức, doanh nghiệp và nhà sản xuất IoT cần chủ động áp dụng các biện pháp phòng chống SCA trong giai đoạn thiết kế thay vì xử lý khi sản phẩm đã đưa vào thực tế.

Có thể thấy rằng các kỹ thuật tấn công kênh bên không lập mẫu có thể khôi phục từng phần khóa AES-128 trên thiết bị IoT với số lượng vết đo ở mức từ vài nghìn đến vài chục nghìn mẫu, tùy thuộc vào điều kiện đo và đặc điểm phần cứng cụ thể. Điều này cho thấy nguy cơ rò rỉ khóa là hiện hữu trong các kịch bản triển khai IoT phổ biến, đặc biệt với các thiết bị IOT giá rẻ phổ biến trên thị trường và khi thiết bị có thể bị tiếp cận vật lý trong một khoảng thời gian nhất định. Tuy nhiên, mức độ đe dọa giảm đáng kể khi chất lượng tín hiệu thấp hoặc khi áp dụng các biện pháp phòng vệ cơ bản, khẳng định rằng rủi ro chủ yếu xuất phát từ điều kiện triển khai hơn là từ bản thân thuật toán AES-128.

3. Hướng nghiên cứu tiếp theo

Mở rộng mô hình DDLA sang kiến trúc sâu hơn, như học sâu chuyển giao, học sâu đa đầu ra, học sâu kết hợp chuyển giao và học sâu đa đầu ra.... Các mô hình tiên tiến có thể trích xuất đặc trưng mạnh hơn và giảm số lượng vệt công suất cần thiết.

Nghiên cứu tấn công trong điều kiện dữ liệu cực ít (few-trace SCA), đặc biệt quan trọng trong các tình huống mà thiết bị giới hạn số lần đo.

Kiểm tra khả năng tấn công "cross-device", tức huấn luyện trên thiết bị A và tấn công sang thiết bị B - một bài toán khó hơn rất nhiều nhưng là yêu cầu thực tế trong bối cảnh IoT đa dạng về phần cứng.

Xây dựng mô hình tấn công đa byte đồng thời, thay vì tách từng byte để huấn luyện; điều này hứa hẹn giảm thời gian tính toán đáng kể trong các bài toán tấn công thật.

Đề xuất các biện pháp phòng vệ mạnh hơn, dựa trên kết quả từ các tấn công học sâu, nhằm đóng góp trở lại vào lĩnh vực thiết kế phần cứng an toàn