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

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT
(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC
TS. ĐỖ NGỌC TUẤN
TS. VŨ THỊ THÚY HÀ

HÀ NỘI - 2025

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

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT
(*Theo định hướng ứng dụng*)

NGƯỜI HƯỚNG DẪN KHOA HỌC
TS. ĐỖ NGỌC TUẤN
TS. VŨ THỊ THÚY HÀ

HÀ NỘI - 2025

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kết quả nêu trong đề án tốt nghiệp là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tác giả



Nguyễn Anh Thư

MỤC LỤC

LỜI CAM ĐOAN i
MỤC LỤC ii
DANH MỤC CÁC KÝ HIỆU VIẾT TẮT iv
DANH MỤC CÁC CHỮ VIẾT TẮT v
Danh mục các bảng vii
Danh mục các hình vẽ viii
MỞ ĐẦU 1
CHƯƠNG 1: TỔNG QUAN VỀ TẤN CÔNG KÊNH BÊN (SCA) 4
1.1. Giới thiệu chung 4
1.1.1. Tấn công các thiết bị mã mật 4
1.1.2. Tấn công kênh bên 5
1.2. Thu thập dữ liệu SCA trên thiết bị mục tiêu 6
1.2.1. Dữ liệu kênh bên và thiết lập đo mức tiêu thụ năng lượng 6
1.2.2. Tiêu thụ năng lượng trong mạch CMOS 7
1.2.3. Thiết lập đo mức tiêu thụ năng lượng 10
1.3. Phân loại phương pháp SCA 12
1.3.1. Tấn công kênh bên lập mẫu 12
1.3.2. Tấn công kênh bên không lập mẫu 14
1.4. Kết luận chương 1 15
CHƯƠNG 2: MỘT SỐ PHƯƠNG PHÁP SCA KHÔNG LẬP MẪU ĐIỂN HÌNH 16
2.1. Phương pháp phân tích vi sai công suất tiêu thụ (DPA) 16
2.1.1. Chiến lược chung trong tấn công kênh bên không lập mẫu 16
2.1.2. Phân tích vi sai công suất tiêu thụ (DPA) 19
2.2. Phương pháp phân tích tương quan công suất tiêu thụ (CPA) 21
2.3. Phương pháp phân tích vi sai tham số mô hình học sâu (DDLA) 22
2.3.1. Học sâu trong tấn công kênh bên 22
2.3.2. Mô hình phân tích vi sai tham số học sâu 24
2.4. Kết luận chương 2 27
CHƯƠNG 3: MÔ PHỎNG, ĐÁNH GIÁ HIỆU QUẢ KỸ THUẬT SCA KHÔNG LẬP MẪU VÀO THUẬT TOÁN AES-128 TRÊN THIẾT BỊ IOT 28
3.1. Tấn công SCA sử dụng CPA 28
3.1.1. Mục tiêu và cách tiếp cận 28

3.1.2. Xử lý dữ liệu và tạo tập dataset 30
3.1.3. Thực hiện CPA trên thuật toán AES-128 32
3.1.4. Kết quả thực nghiệm CPA 33
3.2. Tấn công SCA sử dụng DDLA 37
3.2.1. Mục tiêu và cách tiếp cận 37
3.2.2. Xử lý dữ liệu và xây dựng dataset cho DDLA 38
3.2.3. Thực thi DDLA bằng Python 38
3.2.4. Kết quả thực nghiệm DDLA 38
3.3. Phân tích, đánh giá hiệu quả 42
3.3.1. So sánh CPA và DDLA theo tiêu chí hiệu năng 42
3.3.2. So sánh với các nghiên cứu trước tại Việt Nam 43
3.4. Kết luận chương 3 46
III. KẾT LUẬN 47
1. Các kết quả đạt được 47
2. Nhận xét, đề xuất và khuyến nghị 48
3. Hướng nghiên cứu tiếp theo 50

DANH MỤC CÁC KÝ HIỆU VIẾT TẮT

VIẾT TẮT	TIẾNG ANH	TIẾNG VIỆT
x	Vector x	Véc-tơ x
X	Matrix X	Ma trận X
d_i	Plaintext byte d_i	Byte văn bản rõ thứ i
k_j	Key byte k_j	Byte khóa thứ j
$\boldsymbol{t}_i$	Power trace $\boldsymbol{t}_i$	Vệt công suất thứ i
ρ	Correlation value	Giá trị tương quan
r	Correlation coefficient$r$	Hệ số tương quan
σ	Standard deviationσ	Độ lệch chuẩn
Δ_F	Power consumption difference	Hiệu năng lượng tiêu thụ của hai vệt công suất
R_ρ	Correlation matrix	Ma trận các giá trị tương quan
$\mathbb{R}^D$	D-dimensional real space	Không gian mẫu giá trị thực gồm D chiều
θ	Trainable parameters	Các tham số có thể huấn luyện của mô hình học sâu
$\mathcal{L}_X(\theta)$	Loss function	Hàm mất mát
η	Learning rate	Tốc độ học
∇	Gradient	Gradient

DANH MỤC CÁC CHỮ VIẾT TẮT

VIẾT TẮT	TIẾNG ANH	TIẾNG VIỆT
ACT	Activation Layer	Lớp kích hoạt
ADAM	Adaptive Moment Estimation	Phương pháp tối ưu ADAM
AES	Advanced Encryption Standard	Tiêu chuẩn mã hóa tiên tiến
AI	Artificial Intelligence	Trí tuệ nhân tạo
ANN	Artificial Neural Network	Mạng nơ-ron nhân tạo
ASIC	Application-Specific Integrated Circuit	Mạch tích hợp chuyên dụng
ASCAD	AES Side-Channel Attack Dataset	Bộ dữ liệu tấn công kênh bên AES
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
CMOS	Complementary Metal-OxideSemiconductor	Bán dẫn oxit kim loại bù
CPA	Correlation Power Analysis	Phân tích tương quan năng lượng tiêu thụ
CW	ChipWhisperer	Bo mạch tấn công kênh bên ChipWhisperer
DDLA	Deep Learning Differential Analysis	Phân tích vi sai tham số học sâu
DPA	Differential Power Analysis	Phân tích vi sai năng lượng tiêu thụ
DUT	Device Under Test	Thiết bị đang kiểm thử
ELU	Exponential Linear Unit	Hàm kích hoạt ELU
FC	Fully Connected Layer	Lớp liên kết hoàn toàn

VIẾT TẮT	TIẾNG ANH	TIẾNG VIỆT
FPGA	Field Programmable Gate Array	Mạch tích hợp khả trình
HD	Hamming Distance	Khoảng cách Hamming
HW	Hamming Weight	Trọng số Hamming
LSB	Least Significant Bit	Bit có trọng số nhỏ nhất
ML	Machine Learning	Học máy
MLP	Multi-Layer Perceptron	Perceptron nhiều lớp
MOC	Multi-Output Classification	Phân loại đa đầu ra
MOR	Multi-Output Regression	Hồi quy đa đầu ra
MSB	Most Significant Bit	Bit có trọng số lớn nhất
NMOS	N-type Metal-Oxide-Semiconductor	Bán dẫn oxit kim loại loại N
PL	Pooling Layer	Lớp lấy mẫu
PMOS	P-type Metal-Oxide-Semiconductor	Bán dẫn oxit kim loại loại P
ReLU	Rectified Linear Unit	Hàm kích hoạt ReLU
RNN	Recurrent Neural Network	Mạng nơ-ron hồi tiếp
SBox	Substitution Box	Hộp thay thế trong AES
SCA	Side Channel Attack	Tấn công kênh bên
SNR	Signal-to-Noise Ratio	Tỷ số tín hiệu trên nhiễu
TL	Transfer Learning	Học chuyển giao

Danh mục các bảng

Bảng 1. 1. Loại năng lượng tiêu thụ của các trạng thái chuyển tiếp.....9
Bảng 3. 1. Các file lưu kết quả CW1173 trả về phục vụ cho phân tích trên Python 31

Danh mục các hình vẽ

Hình 1. 1. Phân loại các kỹ thuật tấn công các thiết bị mã mật 5
Hình 1. 2. Mạch nghịch đảo sử dụng CMOS 8
Hình 1. 3. Thiết lập để đo năng lượng tiêu thụ 11
Hình 1. 4. Tấn công kênh bên lập mẫu 13
Hình 2. 1. Tấn công DPA với khóa chính xác (a) và khóa không chính xác (b) 20
Hình 2. 2. Tấn công CPA với khóa chính xác (a) và khóa không chính xác (b) 22
Hình 2. 3. Các thước đo đánh giá hiệu quả của mô hình DDLA trong tấn công kênh bên không lập mẫu 26
Hình 3. 1. Vòng mã hóa số 1 thích hợp cho xây dựng mô hình tiêu thụ công suất .. 29
Hình 3. 2. Quá trình thu thập power trace trên ChipWhisperer W1173 31
Hình 3. 3. Một vệt công suất thu thập được từ mạch CW1173 32
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 33
Hình 3. 5. Kết quả mô phỏng tấn công SCA sử dụng CPA vào Byte 2 đến Byte 10 với các thiết bị không được bảo vệ tấn công kênh bên 34
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 35
Hình 3. 7. Kết quả mô phỏng tấn công SCA sử dụng CPA vào byte 1 (a), Byte 2 (b) 35
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 36
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 36
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 37
Hình 3. 11. Kết quả tấn công Byte 1 (a); Byte 2 (b) 38
Hình 3. 12. Kết quả khi tấn công Byte 9 (a); Byte 10 (b) 39
Hình 3. 13. So sánh thời gian tấn công trung bình và tỷ lệ tấn công thành công của CPA và DDLA 40
Hình 3. 14. So sánh khả năng tấn công trong điều kiện chống tấn công mạnh 41

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, mặc dù AES-128 sở hữu độ an toàn lý thuyết rất cao về mặt toán học, cấu trúc này đang đối mặt với những thách thức nghiêm trọng từ các phương thức tấn công thực thể hiện đại. Trong đó, tấn công kênh bên (Side-Channel Attack - SCA) nổi lên như một mối đe dọa trực tiếp và nguy hiểm nhất. Thay vì tập trung phá giải thuật toán bằng các phương pháp truyền thống, kỹ thuật này khai thác sự rò rỉ thông tin qua các đặc tính vật lý của thiết bị trong quá trình thực thi dữ liệu – như công suất tiêu thụ, bức xạ điện từ hoặc độ trễ thời gian – từ đó cho phép kẻ tấn công khôi phục khóa bí mật một cách hiệu quả mà không cần phá vỡ cấu trúc toán học của hệ thống.

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.

Cấu trúc của đề án gồm ba chương, phản ánh quá trình nghiên cứu theo hướng từ lý thuyết – phương pháp – thực nghiệm. Chương 1 trình bày tổng quan toàn diện về SCA, bao gồm khái niệm, cơ chế rò rỉ thông tin vật lý và mức độ nguy hiểm đối với các thuật toán mật mã hiện đại. Chương này cũng mô tả quy trình thu thập dữ liệu kênh bên, đặc biệt là dấu vết công suất tiêu thụ trên các nền tảng phần cứng như ChipWhisperer, cũng như hệ thống phân loại các kỹ thuật SCA theo hai hướng tiếp cận chính: các phương pháp thống kê truyền thống (DPA, CPA) và phương pháp không lập mẫu dựa trên học máy – học sâu (DDLA).

Tiếp đó, Chương 2 đi sâu phân tích ba phương pháp SCA không lập mẫu điển hình: phân tích vi sai công suất (DPA), phân tích tương quan công suất (CPA) và phương pháp vi sai dựa trên mô hình học sâu (DDLA). Mỗi phương pháp đại diện cho một hướng khai thác khác nhau, từ thống kê cổ điển đến trí tuệ nhân tạo hiện đại. Chương này làm rõ nguyên lý hoạt động, yêu cầu dữ liệu đầu vào, độ phức tạp, ưu – nhược điểm và mức độ phù hợp khi triển khai tấn công lên AES-128 trong môi trường IoT.

Dựa trên nền tảng lý thuyết đó, Chương 3 tập trung thực hiện mô phỏng và đánh giá hiệu quả hai kỹ thuật tấn công tiêu biểu gồm CPA và DDLA trên các tập dữ liệu kênh bên của thuật toán AES-128. Quy trình thực nghiệm bao gồm: tiền xử lý dữ liệu, xây dựng tập dataset, lập trình mô phỏng tấn công bằng Python và phân tích các tham số đánh giá như: số lượng dấu vết cần thiết để khôi phục khóa, độ chính xác tấn công, độ tin cậy, thời gian thực thi và mức độ ổn định trong điều kiện nhiễu. Kết quả thực nghiệm được so sánh với các nghiên cứu trong và ngoài nước, qua đó xác định mức độ dễ tổn thương của AES-128 trên thiết bị IoT khi đối mặt với tấn công kênh bên không lập mẫu.

Bên cạnh tập dữ liệu chuẩn, đề án xem xét khả năng triển khai thực nghiệm trên một số thiết bị IoT phổ biến như Arduino, ESP32, bộ xử lý ARM trên các nền

tảng điều khiển UAV, UGV, kết hợp cùng hệ thống đo chuyên dụng (oscilloscope, đầu dò EM, shunt điện trở). Việc đánh giá dựa trên những tham số chuẩn hóa giúp kết quả có cơ sở khoa học, đồng thời phản ánh mức độ rủi ro khi AES-128 được triển khai trong môi trường thực tế.

Với cách tiếp cận toàn diện từ lý thuyết đến thực nghiệm, đề án hướng tới mục tiêu đưa ra cái nhìn rõ ràng và chính xác về hiệu quả và mức độ nguy hiểm của các kỹ thuật tấn công SCA không lập mẫu đối với AES-128, từ đó góp phần xây dựng cơ sở khoa học cho việc thiết kế, cải tiến các biện pháp bảo vệ thiết bị IoT trong bối cảnh an ninh mạng ngày càng phức tạp.

Quá trình học tập và nghiên cứu, bản thân em đã nhận được rất nhiều sự quan tâm, dìu dắt của quý Thầy Cô giáo, sự hỗ trợ, giúp đỡ của các bạn cùng lớp. Và đặc biệt gửi lời cảm ơn sâu sắc đến Thầy giáo Tiến sĩ Đỗ Ngọc Tuấn – Trường Sĩ quan Thông tin và Cô giáo Tiến sĩ Vũ Thị Thúy Hà – Học viện Công nghệ Bưu Chính Viễn thông đã tận tình chỉ bảo, hỗ trợ, động viên để em hoàn thành nội dung đề án của mình, cũng như trong suốt thời gian học tập, nghiên cứu tại Học viện. Nghiên cứu được thể hiện trong đề án là bước khởi đầu, đặt nền móng để bản thân tiếp tục học tập và nghiên cứu khoa học, từng bước ứng dụng vào thực tiễn công tác sau này.

Đề án còn nhiều hạn chế, thiếu sót do năng lực bản thân còn ở mức độ, do vậy, kính mong nhận được sự chỉ bảo của quý Thầy, Cô và các thành viên trong Hội đồng bảo vệ để em tiếp tục hoàn thiện đề án của mình.

Em xin chân thành cảm ơn!

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

Chương này cung cấp nền tảng lý thuyết về tấn công kênh bên, nhằm hình thành cơ sở cho các nội dung nghiên cứu chuyên sâu ở các chương sau. Mở đầu chương là phần 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).

1.1. Giới thiệu chung

1.1.1. Tấn công các thiết bị mã mật

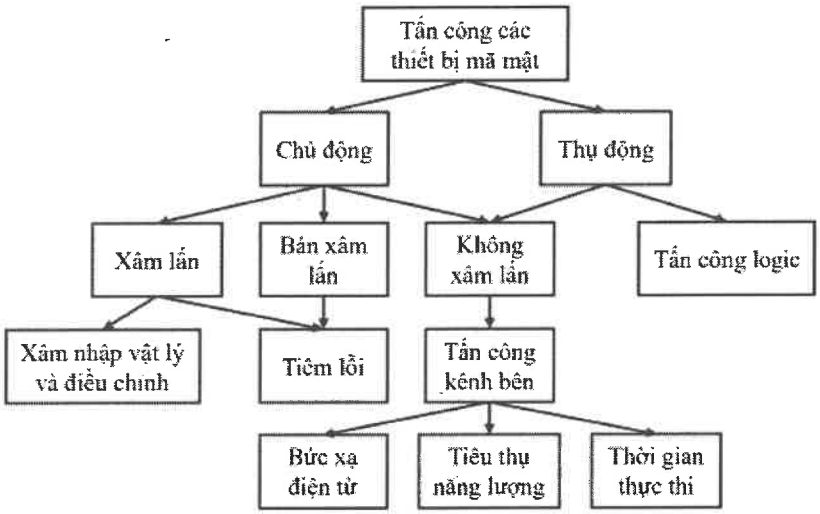
Việc đưa các thuật toán mã hóa lên phần cứng được xem là giải pháp quan trọng nhằm tăng mức độ an toàn cho thông tin trong các hệ thống truyền thông và thiết bị nhúng. Khi thuật toán được thực thi trực tiếp trên chip, tốc độ xử lý cao hơn và dữ liệu nhạy cảm được bảo vệ tốt hơn. Thế nhưng, song song với ưu điểm đó, các thiết bị mã mật lại vô tình trở thành đích nhắm của những kiểu tấn công mới.

Các kỹ thuật tấn công hiện đại không chỉ khai thác lỗ hổng trong thuật toán, mà còn thu thập những thông tin "rò rỉ" từ quá trình phần cứng vận hành. Những tín hiệu vô tình bị lộ – như thời gian xử lý, điện năng tiêu thụ, bức xạ điện từ – lại phản ánh trạng thái xử lý bên trong thiết bị. Từ những dữ liệu phụ này, kẻ tấn công vẫn có thể suy luận ra khóa bí mật mà không cần truy cập trực tiếp vào bộ nhớ hay phá vỡ thuật toán mã hóa [1].

Một cách phân loại thường dùng là dựa vào cách thức tác động lên thiết bị:

- Tấn công chủ động: kẻ tấn công tác động thay đổi môi trường hoạt động hoặc tạo lỗi trên thiết bị để buộc nó tạo ra hành vi bất thường.

- Tấn công thụ động: không can thiệp vào thiết bị, chỉ lặng lẽ quan sát các tín hiệu rò rỉ bên ngoài và phân tích để tìm ra khóa mật Hình 1.1



Hình 1. 1. Phân loại các kỹ thuật tấn công các thiết bị mã mật

Trong tấn công chủ động, kẻ tấn công cố gắng tạo ra sai lệch trong quá trình mã hóa để thu được thông tin có lợi cho việc suy luận khóa. Ví dụ điển hình là tấn công tiêm lỗi, trong đó các tham số vận hành như nhiệt độ, xung nhịp hoặc điện áp được điều chỉnh gây lỗi tính toán. Kết quả sai lệch thu được sau khi mã hóa có thể được phân tích để rút ra khóa bí mật. Một dạng khác là tấn công khởi động lạnh, tận dụng khả năng lưu trữ tạm thời của bộ nhớ để thu thập dữ liệu như khóa bí mật sau khi nguồn điện bị ngắt.

Ngược lại, tấn công thụ động không can thiệp vào hoạt động của thiết bị mà chỉ quan sát các thông tin rò rỉ phát sinh trong quá trình mã hóa. Tấn công kênh bên là phương pháp nổi bật của nhóm này. Phương pháp tận dụng các tín hiệu vật lý như mức tiêu thụ năng lượng, bức xạ điện từ hoặc thời gian thực thi để suy luận khóa bí mật. Đây là mối đe dọa đáng kể bởi các tín hiệu này vô tình được tạo ra trong quá trình hoạt động bình thường và có thể thu thập bằng các thiết bị đo tương đối đơn giản.

1.1.2. Tấn công kênh bên

Tấn công kênh bên là nhóm kỹ thuật khai thác các tín hiệu phụ sinh ra trong quá trình thiết bị thực hiện thao tác mật mã, thay vì dựa trực tiếp vào thuật toán được cài đặt bên trong. Những tín hiệu phụ này không phản ánh bản chất của thuật toán,

mà là các hiện tượng phát sinh do hoạt động vật lý của thiết bị, chẳng hạn như thời gian xử lý [12], biến thiên năng lượng tiêu thụ [11], hay mức điện áp/điện từ rò rỉ [19] trong lúc tính toán. Khi những thông tin này bị thụ thập và phân tích bằng công cụ phù hợp, chúng có thể tiết lộ các dữ liệu nhạy cảm, bao gồm khóa bí mật.

Các loại thông tin rò rỉ nói trên thường được sử dụng để hiện thực các dạng tấn công kênh bên phổ biến. Bản chất của dạng tấn công này là phân tích sự tương quan giữa thông tin phụ thu được và các thao tác logic diễn ra bên trong thiết bị. Ví dụ, mức năng lượng tiêu thụ hay tín hiệu điện từ trong mỗi chu kỳ xử lý thường thể hiện sự khác biệt rõ rệt theo từng trạng thái dữ liệu mà hệ thống xử lý. Điều này cho phép kẻ tấn công lựa chọn kỹ thuật khai thác phù hợp với từng loại thiết bị đích để suy luận khóa mật [2].

Trong phạm vi của đề án, dữ liệu thu được từ kênh bên thường liên quan tới sự thay đổi công suất tiêu thụ của thiết bị. Mặc dù mỗi thiết bị có đặc điểm tiêu thụ điện năng khác nhau, các phương pháp phân tích đã được phát triển nhằm xử lý loại dữ liệu này một cách hiệu quả nhằm hỗ trợ tấn công.

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

Trong các mạch điện tử, đặc biệt là mạch số, năng lượng được tiêu thụ trong quá trình xử lý tín hiệu. Dòng điện đi từ nguồn qua các phần tử logic, sau đó chuyển hóa thành nhiệt năng. Vì vậy, quản lý và đo mức tiêu thụ năng lượng không chỉ nhằm đảm bảo hiệu suất hoạt động mà còn liên quan đến công tác đánh giá an toàn cho các thiết bị mã hóa.

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.

Do đó, quá trình thu thập dữ liệu không chỉ là bước chuẩn bị kỹ thuật, mà còn là yếu tố quyết định chất lượng mô hình tấn công. Dữ liệu càng phản ánh rõ mối quan hệ giữa hoạt động xử lý và mức tiêu thụ năng lượng thì khả năng khôi phục khóa càng cao.

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. Cấu trúc cơ bản của CMOS là sự phối hợp giữa hai transistor PMOS và NMOS để hình thành các khối logic, trong đó PMOS đảm nhiệm việc kéo điện áp lên mức cao, còn NMOS đảm nhiệm kéo tín hiệu xuống mức thấp.

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.

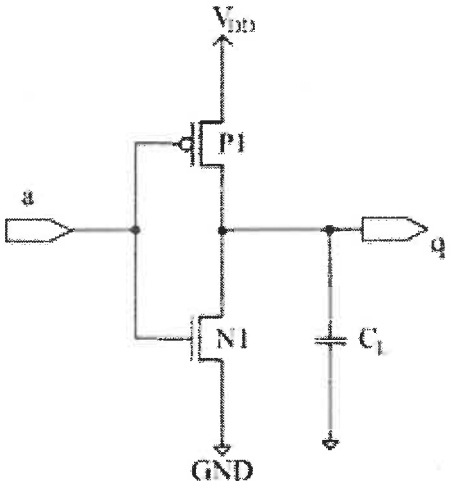
Dựa trên nguyên tắc đó, các kỹ thuật tấn công kênh bên thường không quan tâm đến giá trị logic tại đầu ra, mà chủ yếu khai thác sự biến thiên năng lượng xuất hiện trong quá trình chuyển trạng thái. Ở giai đoạn chuyển trạng thái này, mỗi transistor của cổng CMOS sẽ tạo ra một lượng năng lượng khác nhau 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.

Do đó, chính quá trình chuyển trạng thái là nguồn phát sinh thông tin phụ quan trọng, cung cấp cơ sở cho việc suy luận dữ liệu nội bộ của mạch. Điều này lý giải lý do tại sao CMOS tuy hiệu quả về mặt năng lượng, nhưng lại vô tình trở thành mục tiêu ưu tiên của các hình thức tấn công kênh bên khi triển khai thuật toán mật mã.

Trong các thiết bị mã hóa sử dụng phần cứng chuyên dụng như FPGA hay ASIC, phần tử logic thường được hình thành bởi các transistor loại PMOS và NMOS,

kết hợp với các kết nối bên trong để thực hiện chức năng xử lý logic. Một công nghệ phổ biến dùng để thiết kế các phần tử logic đó là CMOS, hoạt động dựa trên cấu trúc gồm mạng pull-up và pull-down. Trong cấu trúc này, mạng pull-up sử dụng các transistor PMOS nối với nguồn điện áp dương V_{DD}, có nhiệm vụ kéo đầu ra lên mức logic cao (mức 1) khi dẫn điện. Trái lại, mạng pull-down được tạo bởi các transistor NMOS nối với đất (GND) và sẽ đưa đầu ra về mức logic thấp (mức 0) khi được kích hoạt.

Hai mạng PMOS và NMOS làm việc đối lập nhau, đảm bảo rằng chỉ có một trong hai loại transistor dẫn điện tại một thời điểm. Nhờ vậy, các mạch CMOS chỉ tiêu thụ đáng kể năng lượng khi xảy ra chuyển đổi trạng thái logic, trong khi mức tiêu thụ ở trạng thái ổn định là rất nhỏ. Hình 1.2 minh họa cấu trúc của một cổng đảo sử dụng CMOS, trong đó P_1 sẽ dẫn điện còn N_1 bị ngắt khi đầu vào là GND (mức 0); ngược lại P_1 tắt và N_1 dẫn điện khi đầu vào được đặt ở mức V_{DD}.



Hình 1. 2. Mạch nghịch đảo sử dụng CMOS

Ký hiệu $i_{DD}(t)$ là dòng điện tại thời điểm t và $p_{tt}(t)$ công suất mà mạch tiêu thụ tức thời. Giá trị công suất tiêu thụ trung bình của mạch trong khoảng thời gian T được ước lượng thông qua biểu thức:

$$P_{tb} = \frac{1}{T}\int_0^T p_{tt}(t)dt = \frac{V_{DD}}{T}\int_0^T i_{DD}(t)dt \quad (1.1)$$

Năng lượng tiêu thụ của mạch điện tử bao gồm hai thành phần: năng lượng động và năng lượng tĩnh [2]. Trong đó, năng lượng động xuất hiện khi tín hiệu nội bộ

hoặc tín hiệu đầu ra của mạch có sự chuyển đổi trạng thái giữa 0 và 1, chủ yếu do quá trình nạp/xả điện dung và hiện tượng dẫn tạm thời qua các transistor. Ngược lại, năng lượng tĩnh xuất hiện ngay cả khi tín hiệu không thay đổi, chủ yếu do dòng rò trong transistor, nhưng giá trị thường rất nhỏ nên có thể bỏ qua trong hầu hết các tình huống tấn công kênh bên.

Xét tại một thời điểm bất kỳ, đầu ra của một cổng logic có thể rơi vào bốn kiểu chuyển trạng thái khác nhau: từ 1 sang 1, từ 0 sang 0, từ 0 sang 1 hoặc từ 1 sang 0. Trong hai tình huống không xảy ra thay đổi ($0 \rightarrow 0$ và $1 \rightarrow 1$), mạch chỉ tồn tại năng lượng tĩnh và mức tiêu thụ rất nhỏ. Ngược lại, các chuyển tiếp $0 \rightarrow 1$ hoặc $1 \rightarrow 0$ yêu cầu nạp hoặc xả điện tích, dẫn tới phát sinh năng lượng động đáng kể. Chính sự sai khác này khiến hoạt động chuyển trạng thái trở thành mục tiêu quan trọng để quan sát trong tấn công kênh bên. Ngoài ra, mức tiêu thụ năng lượng động trong mạch CMOS còn phụ thuộc vào loại dữ liệu mà mạch đó đang xử lý.

Bảng 1. 1. Loại năng lượng tiêu thụ của các trạng thái chuyển tiếp

Loại chuyển tiếp	Năng lượng tiêu thụ	Loại năng lượng tiêu thụ
$0 \rightarrow 0$	P_{00}	Tĩnh
$0 \rightarrow 1$	P_{01}	Tĩnh và động
$1 \rightarrow 0$	P_{10}	Tĩnh và động
$1 \rightarrow 1$	P_{11}	Tĩnh

Có hai nguyên nhân chính dẫn đến việc tiêu thụ năng lượng động trong các phần tử CMOS. Nguyên nhân thứ nhất là tải điện dung của phần tử logic cần được nạp điện. Gọi p_{chrg} là điện năng tiêu thụ tức thời do dòng nạp tạo ra. Điện năng nạp trung bình P_{chrg} do phần tử CMOS tiêu thụ trong khoảng thời gian T có thể được tính theo công thức sau:

$$P_{chrg} = \frac{1}{T}\int_0^T p_{chrg}(t)dt = \beta \times f \times C_L \times V^2{}_{DD} \quad (1.2)$$

Trong đó, C_L là điện dung đầu ra, có độ lớn phụ thuộc vào đặc tính vật lý và công nghệ chế tạo phần tử CMOS, f là tần số xung nhịp, β được gọi là hệ số hoạt động của phần tử CMOS tương ứng với số lần chuyển trạng thái từ $0 \rightarrow 1$ trung bình

xuất hiện trong mỗi chu kỳ xung nhịp tại đầu ra của phần tử CMOS (nếu đầu ra của phần tử CMOS chuyển trạng thái một lần trong một chu kỳ xung nhịp thì β = 1). Có thể thấy rằng, P_{chrg} tỉ lệ với bình phương điện áp nguồn cung cấp nên cách hiệu quả nhất để giảm mức tiêu thụ năng lượng động của mạch CMOS là giảm điện áp của nguồn cung cấp.

Thành phần thứ hai dẫn tới việc tiêu thụ năng lượng động của các phần tử logic là trạng thái ngắn mạch tạm thời xuất hiện tại phần tử CMOS trong quá trình chuyển trạng thái tại đầu ra. Trạng thái ngắn mạch tạm thời này xảy ra trong một khoảng thời gian rất ngắn khi hai transistor P1 và N1 của mạch nghịch đảo trong Hình 1.2 cùng dẫn điện. Điều này xảy ra khi chuyển trạng thái $0 \rightarrow 1$ và $1 \rightarrow 0$. Gọi p_{sc} là mức tiêu thụ năng lượng động tức thời của phần tử CMOS khi xảy ra ngắn mạch. Khi đó, mức tiêu thụ năng lượng động trung bình do ngắn mạch P_{sc} trong khoảng thời gian T được tính như sau:

$$P_{sc} = \frac{1}{T}\int_0^T p_{sc}(t)dt = \beta \times f \times I_{peak} \times t_{sc} \quad (1.3)$$

Trong đó, t_{sc} được gọi là thời gian xảy ra ngắn mạch và I_{peak} là giá trị của dòng điện đỉnh gây ra hiện tượng ngắn mạch trong quá trình chuyển trạng thái.

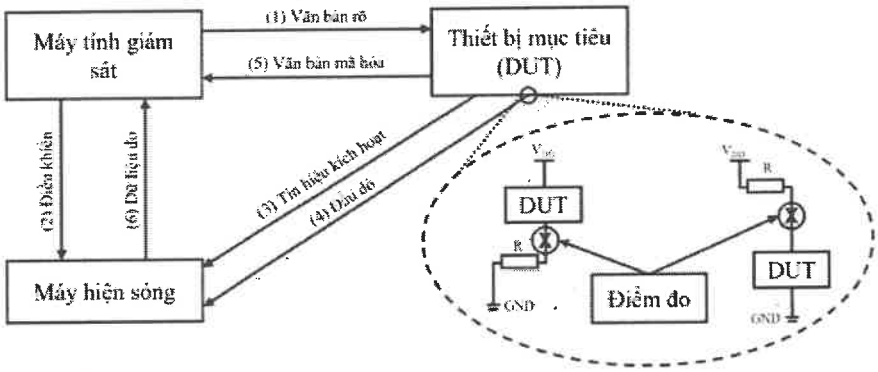
Không như dòng điện nạp chỉ xuất hiện khi có sự chuyển trạng thái $0 \rightarrow 1$ ở đầu ra, ngắn mạch xảy ra trong cả hai trường hợp chuyển trạng thái $1 \rightarrow 0$ và $0 \rightarrow 1$. Do đó, điện năng tiêu thụ khi chuyển trạng thái $0 \rightarrow 1$ lớn hơn khi chuyển trạng thái $1 \rightarrow 0$.

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 mô tả một dạng thiết lập điển hình dùng để đo mức năng lượng tiêu thụ với các thành phần cần thiết bao gồm máy tính giám sát, thiết bị mục tiêu (DUT) và máy hiện sóng.

Thiết bị mục tiêu (DUT): đây chính là thiết bị thực hiện các thuật toán mã mật. Thiết bị này có các chuẩn giao tiếp để liên lạc với máy tính giám sát (thường là GPIO) và tạo ra tín hiệu kích hoạt cho máy hiện sóng bắt đầu thu thập dữ liệu tiêu thụ năng lượng. DUT có thể là một thiết bị nhúng thực hiện mã hoá/giải mã dữ liệu bằng phần mềm hoặc là các phần cứng thực thi thuật toán mã mật trên FPGA hoặc ASIC. Để đo được mức năng lượng tiêu thụ của thiết bị mục tiêu, cần phải sử dụng một điện trở nhỏ (có giá trị khoảng 1 Ω đến 50 Ω) nối vào chân GND hoặc V_{DD} của DUT. Các điểm đo dữ liệu được xác định như mô tả trong Hình 1.3.



Hình 1. 3. Thiết lập để đo năng lượng tiêu thụ

Máy hiện sóng số: được sử dụng để đo dữ liệu kênh bên về mức năng lượng tiêu thụ khi mục tiêu thực hiện các thuật toán mã mật. Máy hiện sóng được cấu hình bao gồm hai kênh thu nhận tín hiệu tương tự. Hai kênh này được kết nối với hai đầu dò thụ động để thu thập các vệt công suất từ thiết bị mục tiêu. Trong đó, một đầu dò được sử dụng để thu thập tín hiệu tương tự từ chân V_{DD} hoặc GND của DUT, đầu dò thứ hai được dùng để phát hiện tín hiệu kích hoạt do thiết bị mục tiêu cung cấp qua chân GPIO. Máy hiện sóng được điều khiển từ xa thông qua máy tính giám sát.

Máy tính giám sát: sử dụng để vận hành toàn bộ hệ thống đo tự động. Nó giao tiếp với máy hiện sóng thông qua cổng LAN và với thiết bị mục tiêu thông qua cổng USB.

Quy trình thu thập dữ liệu về mức tiêu thụ năng lượng của thiết bị mục tiêu trong thiết lập ở trên được thể hiện trong Hình 1.3, các số thứ tự chỉ trình tự giao tiếp của các thành phần trong hệ thống đo. Đầu tiên, máy tính giám sát gửi một văn bản rõ đến thiết bị mục tiêu (1) để thực hiện mã hóa và điều khiển máy hiện sóng thu thập

các vệt công suất (2). Trong thiết lập này, máy hiện sóng số được cấu hình ở chế độ TRIGGER (chế độ đo dữ liệu khi có tín hiệu kích hoạt xuất hiện). Khi bộ xử lý của thiết bị mục tiêu thực hiện mã hoá văn bản do máy tính giám sát gửi đến, tín hiệu kích hoạt được phát ra (3). Sau đó, máy hiện sóng ghi lại dữ liệu về mức năng lượng tiêu thụ của thiết bị mục tiêu khi nó nhận được tín hiệu kích hoạt (4). Sau khi hoàn thành việc mã hoá, thiết bị mục tiêu gửi văn bản đã mã hóa tương ứng với văn bản rõ mà nó nhận được đến máy tính giám sát (5). Cuối cùng, dữ liệu về mức tiêu thụ năng lượng mà máy hiện sóng thu thập được sẽ được chuyển đến máy tính giám sát (6).

Dữ liệu kênh bên về mức tiêu thụ năng lượng của thiết bị mục tiêu có thể được xem như các véc-tơ với giá trị tỉ lệ với dòng điện tiêu thụ trong thiết bị. Các dữ liệu này được ghi lại bởi máy hiện sóng số và được gọi là các "vệt công suất". Trong phần còn lại của đề án, khái niệm "vệt công suất" sẽ được sử dụng để miêu tả dữ liệu mức tiêu thụ năng lượng thu được từ thiết bị mục tiêu.

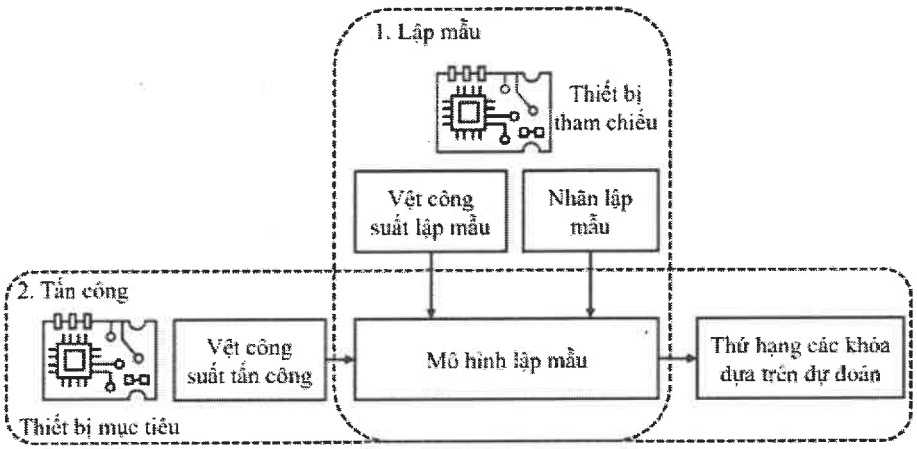
1.3. Phân loại phương pháp SCA

Trong phạm vi đề án này, tấn công kênh bên được phân loại dựa trên sự khác nhau trong chuẩn bị dữ liệu và sử dụng phương tiện để thực hiện tấn công. Theo hướng tiếp cận này, tấn công kênh bên có thể được phân loại thành: tấn công lập mẫu và tấn công không lập mẫu.

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. Kĩ thuật tấn công này lần đầu tiên được đề xuất trong [9], sau đó được phát triển thành một số kĩ thuật tấn công phổ biến bao gồm "Tấn công mẫu" [7] và "Tấn công ngẫu nhiên" [20]. Cụ thể, tấn công lập mẫu được thực hiện qua hai giai đoạn: giai đoạn lập mẫu và giai đoạn tấn công như minh họa trong Hình 1.4. Trong giai đoạn lập mẫu, một số lượng lớn các vệt công suất từ thiết bị tham chiếu được ghi lại và sử dụng để xây dựng mô hình lập mẫu cho mỗi khoá giả thuyết dựa trên phân

bố Gauss đa biến. Sau đó, ở giai đoạn tấn công hay giai đoạn trích xuất khoá bí mật, một lượng nhỏ các vệt công suất (nhỏ hơn nhiều so với giai đoạn lập mẫu) thu thập được từ thiết bị mục tiêu được sử dụng để tìm ra khoá bí mật dựa vào các mô hình mẫu đã được xây dựng trong giai đoạn lập mẫu. Phương pháp trích xuất khóa bí mật này dựa vào thước đo xác suất lớn nhất. Nếu mô hình lập mẫu ánh xạ các vệt công suất đầu vào thành các dữ liệu đầu ra có độ chính xác cao thì khi tấn công chỉ cần thu thập rất ít các vệt công suất từ thiết bị mục tiêu nhưng vẫn đạt được hiệu quả cao trong việc khôi phục lại khóa bí mật.



Hình 1. 4. Tấn công kênh bên lập mẫu

Mặc dù tấn công lập mẫu được coi là kỹ thuật tấn công mạnh nhất và tạo ra nguy cơ bảo mật cao nhất trong tấn công kênh bên nhưng kỹ thuật này ít có khả năng áp dụng do các điều kiện để thực thi nó nhiều khi không thực hiện được trong thực tế. Cụ thể, đối với các thiết bị mục tiêu linh hoạt và có tính tuỳ biến cao dựa trên các thiết kế và kiến trúc mã nguồn mở như FPGA thì việc tạo ra các thiết bị tham chiếu để xây dựng mô hình lập mẫu tương đối khó khăn. Một trường hợp khác là các sản phẩm đóng như thẻ thông minh dùng trong các ứng dụng ngân hàng. Dù có thể thu thập được các thẻ thông minh nhưng người tấn công không có quyền điều khiển các khoá bí mật và thường bị giới hạn số lượt truy cập thể nên không thể thu thập được số lượng các vệt công suất đủ lớn dùng cho xây dựng mô hình mẫu. Trong những trường hợp đó, không thể thực hiện tấn công lập mẫu được. Tuy nhiên, các thiết bị bảo mật vẫn có nguy cơ bị đe doạ bằng phương pháp khác gọi là tấn công không lập mẫu.

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. Trong giai đoạn xây dựng mô hình tiêu thụ năng lượng, người tấn công sử dụng tất cả các giá trị giả thuyết của khóa bí mật để tính toán giá trị trung gian tương ứng dựa trên thuật toán mã hóa (với thuật toán AES-128 thì giá trị trung gian có thể là giá trị đầu vào hoặc đầu ra S-Box). Đối với thuật toán mã hóa hiện đại thì việc tính toán toàn bộ các khóa giả thuyết là bất khả thi nên cần phải chia nhỏ khóa thành các phần nhỏ để tấn công lặp lại nhiều lần. Ví dụ, khi tấn công thiết bị mã mật sử dụng thuật toán AES-128, tổng số lượng khóa giả thuyết cần tính toán là 2^{128}, nếu chia khóa thành các byte, mỗi byte là một khóa con thì số lượng khóa giả thuyết cần tính toán là 16×2^8. Việc tấn công khi này sẽ được thực hiện trên các khóa con để tìm ra từng phần của khóa bí mật. Sau đó, các mô hình tiêu thụ năng lượng sẽ được sử dụng để dự đoán giá trị rò rỉ từ giá trị trung gian đã tính toán để tạo nên mô hình tiêu thụ năng lượng chuẩn của thiết bị. Trong giai đoạn tấn công, bằng cách tính toán mối tương quan giữa mô hình tiêu thụ năng lượng chuẩn và các giá trị tiêu thụ năng lượng ghi lại được từ thiết bị mục tiêu, tấn công không lập mẫu có thể khôi phục lại được khoá bí mật. Do vậy, tấn công không lập mẫu không yêu cầu sử dụng bất kì thiết bị tham chiếu nào.

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 1 phân tích bản chất, nguyên lý và quy trình của tấn công kênh bên – một hướng tấn công đặc biệt nguy hiểm trong lĩnh vực bảo mật phần cứng. Các nội dung về tiêu thụ năng lượng trong mạch CMOS, thiết lập đo công suất và phân loại các phương pháp tấn công đã tạo nên khung lý thuyết vững chắc cho nghiên cứu. Từ đó, có thể thấy rằng SCA không chỉ đòi hỏi hiểu biết sâu về thuật toán mã hóa mà còn yêu cầu năng lực phân tích vật lý thiết bị. Đây chính là tiền đề quan trọng để nghiên cứu chi tiết các kỹ thuật tấn công không lập mẫu được trình bày trong Chương 2.

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

Chương 2 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.

Phương pháp phân tích vi sai công suất tiêu thụ (DPA): Trình bày nguyên lý hoạt động của DPA – kỹ thuật dựa trên thống kê sự khác biệt trung bình của tín hiệu công suất theo hai nhóm dữ liệu đầu vào có liên quan đến một giả thiết khóa. Phương pháp này không yêu cầu mô hình hóa chi tiết thiết bị và có thể triển khai trực tiếp trên dữ liệu thu thập được từ thiết bị thực.

Phương pháp phân tích tương quan công suất tiêu thụ (CPA): Giới thiệu chi tiết phương pháp CPA – một cải tiến so với DPA, sử dụng hệ số tương quan giữa công suất đo được và giá trị dự đoán của một hàm rò rỉ (leakage model) để xác định khóa bí mật. CPA có hiệu quả cao hơn DPA trong nhiều trường hợp, đặc biệt khi có thể xây dựng một mô hình rò rỉ hợp lý.

Phương pháp phân tích vi sai tham số mô hình học sâu (DDLA): Trình bày phương pháp tấn công hiện đại sử dụng mạng nơ-ron sâu để học trực tiếp mối quan hệ giữa tín hiệu công suất và các tham số liên quan đến khóa mà không cần xây dựng mô hình rò rỉ cụ thể. DDLA được đánh giá là tiềm năng trong các tình huống mà mô hình toán học truyền thống không đủ chính xác, đặc biệt là với dữ liệu nhiễu cao.

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.

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

Phần lớn các phương pháp tấn công không lập mẫu đều sử dụng nguyên tắc "chia để trị", trong đó khóa bí mật sẽ được chia thành các khóa con, việc tấn công

được tiến hành độc lập trên từng khóa con. Từ đó, có thể khôi phục lại toàn bộ khóa bí mật của thuật toán mã mật. Điều này có thể thực hiện được là do các thuật toán mã mật phổ biến như AES hay DES đều xử lý các phần riêng lẻ của khóa bí mật một cách độc lập tại một thời điểm nhất định trong quá trình mã hóa hoặc giải mã. 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.

Giá trị trung gian của thuật toán mã mật là các kết quả của một hàm $f_{(d,k)}$ được tính toán dựa trên dữ liệu ngẫu nhiên đã biết d và một phần (một byte) của khóa bí mật k. Trong hầu hết các phương pháp tấn công không lập mẫu, dữ liệu ngẫu nhiên d có thể là văn bản rõ hoặc văn bản đã được mã hóa.

Các giá trị trung gian giả thuyết là các giá trị trung gian được tính toán dựa trên những lựa chọn có thể có của k, tức là các giá trị giả thuyết của khóa bí mật. Tất cả các giá trị giả thuyết này có thể được biểu diễn bởi vector $\boldsymbol{k} = (k_1, \dots, k_n)$, trong đó K thể hiện tổng số lượng các giá trị giả thuyết của k. Nếu biểu diễn tập hợp các dữ liệu ngẫu nhiên đã biết dưới dạng véc-tơ dữ liệu d cùng với véc-tơ khoá giả thuyết k, có thể dễ dàng tính được các giá trị trung gian giả thuyết $f_{(d,k)}$trong D lần mã hoá với K khoá giả thuyết. Việc tính toán này cho ra ma trận V có kích thước $D \times K$.

$$\boldsymbol{V = v_{i,j} = f(d_i, k_j)} \qquad \boldsymbol{i = 1, \dots, D} \qquad \boldsymbol{j = 1, \dots, K} \qquad (2.1)$$

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

Các giá trị tiêu thụ năng lượng giả thuyết $p_{i,j}$ là các giá trị được tính toán sử dụng hàm dự đoán $f_{pred}(.)$ dựa trên các giá trị trung gian giả thuyết. Các giá trị tiêu thụ năng lượng giả thuyết được tính như sau:

$$\boldsymbol{p_{i,j} = f_{pred}(v_{i,j})} \qquad (2.2)$$

Việc lựa chọn hàm dự đoán phụ thuộc phần lớn vào những hiểu biết của người tấn công về thiết bị mục tiêu. Hàm dự đoán được lựa chọn càng phù hợp với các đặc tính tiêu thụ năng lượng thực tế của thiết bị mục tiêu thì hiệu quả tấn công sẽ càng cao. Hàm dự đoán còn có tên gọi khác là mô hình tiêu thụ năng lượng.

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

Giả sử thiết bị mục tiêu xử lý N hàm f_n^{di} liên tiếp trong quá trình mã hóa dữ liệu ngẫu nhiên đầu vào d_i. Trong mỗi lần thực hiện mã hóa, khi các hàm được xử lý thì mức năng lượng t_i mà thiết bị tiêu thụ được thu thập, đồng thời dữ liệu đầu ra tương ứng cũng được ghi lại. Quá trình xử lý trong thiết bị có thể miêu tả bởi các ma trận (2.3) và (2.4). Trong đó, mỗi vector cột f_n thuộc ma trận F thực thi cùng một hàm. Ma trận T chứa thông tin về các vệt công suất t_i thu thập được sau D lần thực hiện quá trình mã hóa. Mỗi vệt công suất bao gồm S mẫu ($S > N$) sao cho mỗi hàm f_n xử lý ít nhất một mẫu trong vệt công suất.

$$\boldsymbol{F} = \left(\left| \begin{matrix} f_1^{d_1} & f_2^{d_1} & \cdots \\ f_1^{d_2} & f_2^{d_2} & \cdots \\ f_1^{d_3} & f_2^{d_3} & \cdots \\ \vdots & \vdots & \cdots \\ f_1^{d_D} & f_2^{d_D} & \cdots \end{matrix} \right| f_n \left| \begin{matrix} \cdots & f_N^{d_1} \\ \cdots & f_N^{d_2} \\ \cdots & f_N^{d_3} \\ \vdots & \vdots \\ \cdots & f_N^{d_4} \end{matrix} \right| \right) \tag{2.3}$$

$$\boldsymbol{T} = \left(\begin{array}{ccc|c|cc} t_{1,1} & t_{1,2} & \cdots & t_n & \cdots & t_{1,S} \\ t_{2,1} & t_{2,2} & \cdots & & \cdots & t_{2,S} \\ \vdots & \vdots & \cdots & & \cdots & \vdots \\ \hline t_i & & & & \cdots & \\ \hline \vdots & \vdots & \cdots & & \cdots & \vdots \\ t_{D,1} & t_{D,2} & \cdots & & \cdots & t_{D,S} \end{array} \right) \tag{2.4}$$

Bước 4: Xác định khóa chính xác

Trong tấn công kênh bên không lập mẫu, dù tấn công dựa trên phương pháp thống kê toán học hay dựa vào các mô hình học sâu, việc khôi phục khóa bí mật của thuật toán mã mật đều có thể thực hiện được bằng cách so sánh các giá trị tiêu thụ năng lượng của mỗi khoá giả thuyết với các giá trị tiêu thụ năng lượng thực tế của thiết bị mục tiêu thu thập được (vệt công suất) tại mỗi vị trí.

Ví dụ, trong tấn công phân tích tương quan năng lượng tiêu thụ, các giá trị tương quan ρ_k sẽ được sử dụng để xác định khóa chính xác k_{cr}. Nếu lựa chọn hàm dự đoán $f_{pred}(.)$ phù hợp với các vệt công suất thu thập được từ mục tiêu thì ma trận V

sẽ có mối tương quan lớn so với các vệt công suất và khi đó, giá trị tương quan $\rho_{k_{cr}}$ tương ứng với khoá chính xác sẽ lớn hơn các giá trị tương quan ρ_{k_j} còn lại ($k_j \neq k_{cr}$).

Đối với tấn công kênh bên sử dụng mô hình học sâu, mô hình huấn luyện sẽ thực hiện xấp xỉ hàm Net (t_i , $p_{i,j}$) dung để dự đoán giá trị của $p_{i,j}$ tương ứng với khóa giả thuyết k_j từ các vệt công suất t_i. Sau đó, các thước đo huấn luyện của mô hình học sâu như độ chính xác hay mất mát được sử dụng để xác định khóa chính xác thay vì sử dụng giá trị tương quan ρ_k.

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

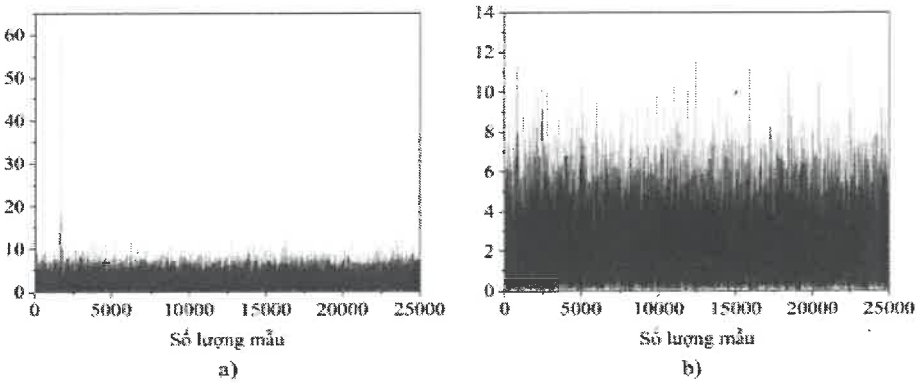
Phân tích vi sai năng lượng tiêu thụ (DPA) là phương pháp thống kê dùng để phân tích một tập hợp các giá trị năng lượng tiêu thụ đo được để xác định các mối tương quan phụ thuộc vào dữ liệu được giới thiệu bởi Kocher và cộng sự [11]. Cơ sở để thực hiện tấn công DPA là mức tiêu thụ năng lượng khi chuyển 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.

Trong thực tế, để kiểm tra các khóa giả thuyết $k_K (0 \leq K < 256$ đối với mã hóa AES-128), DPA sử dụng các văn bản rõ (hoặc văn bản mã hóa) $d_{i_{1 \leq i \leq D}}$ và phân

loại $F(d_i, k_K)$. Trong tấn công DPA truyền thống, tức DPA 1 bit, hàm phân loại sẽ tính toán giá trị của bit cần kiểm tra b, bit này có thể là trọng số nhỏ nhất tại đầu ra của S-Box (LSB (Sbox$(d_i \oplus k_K)$). Sau đó, DPA tính toán hiệu Δ_F của giá trị trung bình của các vệt công suất mà hàm phân loại $F(d_i, k_K)$ có giá trị là 1 và giá trị trung bình của các vệt công suất mà hàm phân loại $F(d_i, k_K)$ có giá trị là O. Theo [11], hiệu mức tiêu thụ năng lượng $\Delta_F(b)$ tại bit b được tính theo công thức sau:

$$\Delta_F(b) = \frac{\sum_{i=1}^{D} F(d_i, k_K, b)t_i}{\sum_{i=1}^{D} F(d_i, k_K, b)} - \frac{\sum_{i=1}^{D}(1 - F(d_i, k_K, b))t_i}{\sum_{i=1}^{D}(1 - F(d_i, k_K, b)} \quad (2.5)$$

Về mặt lý thuyết, nếu các bit được xử lý trong quá trình mã hóa có phân bố đều thì $\Delta_F(b)$ sẽ có giá trị dần về 0 với số lượng vệt công suất đủ lớn khi sử dụng các khóa giả thuyết không chính xác. Ngược lại, với khóa giả thuyết chính xác thì $\Delta_F(b)$ sẽ có giá trị khác 0.



Hình 2. 1. Tấn công DPA với khóa chính xác (a) và khóa không chính xác (b)

Hình 2.1 miêu tả kết quả tấn công DPA đối với thuật toán AES-128 sử dụng 2000 vệt công suất, mỗi vệt công suất gồm 25.000 mẫu. Có thể thấy rằng với khóa giả thuyết chính xác, hiệu mức tiêu thụ năng lượng giữa hai tập hợp con tạo ra một đỉnh DPA, trong khi với khóa không chính xác, hiệu các mức tiêu thụ năng lượng không có sự khác biệt quá lớn. Trong trường hợp này, các giá trị hiệu mức tiêu thụ năng lượng không bằng 0 là do ảnh hưởng các tạp âm và các loại nhiễu xuất hiện trong quá trình thu thập các vệt công suất.

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

Tấn công phân tích tương quan năng lượng tiêu thụ (CPA) lần đầu tiên được đề xuất bởi Brier và cộng sự trong [6]. 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:

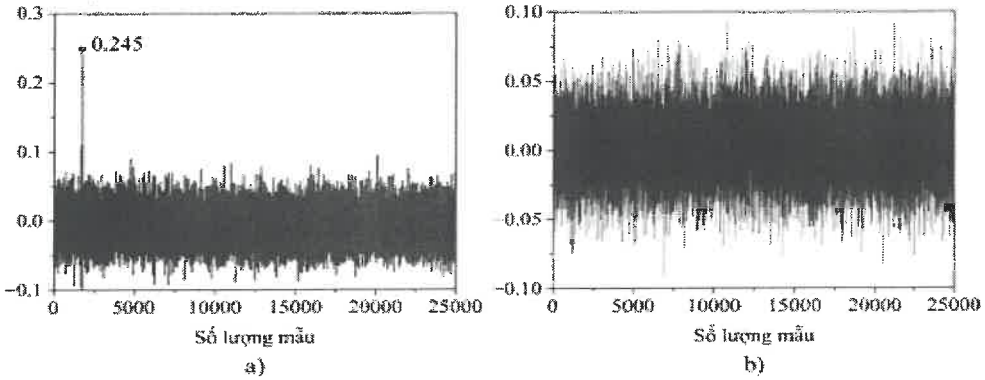
$$r_{k,j} = \frac{\sum_{i=1}^{D}(h_{i,j} - \overline{h_k})(t_{i,j} - \overline{t_j})}{\sqrt{\sum_{i=1}^{D}(h_{i,j} - \overline{h_k})^2 \sum_{i=1}^{D}(t_{i,j} - \overline{t_j})^2}} \tag{2.6}$$

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), công thức (2.6) có thể viết dưới dạng tương đương như sau:

$$r_{k,j} = \frac{D\sum_{i=1}^{D} h_{i,k}t_{i,j} - (\sum_{i=1}^{D} h_{i,k})(\sum_{i=1}^{D} t_{i,j})}{\sqrt{((\sum_{i=1}^{D} h_{i,j})^2 - D\sum_{i=1}^{D} h^2{}_{i,j})((\sum_{i=1}^{D} t_{i,j})^2 - D\sum_{i=1}^{D} t^2{}_{i,j})}} \quad (2.7)$$



Hình 2. 2. Tấn công CPA với khóa chính xác (a) và khóa không chính xác (b)

Hình 2.2 thể hiện kết quả tấn công CPA trên tập dữ liệu khi sử dụng khóa chính xác và các khóa không chính xác. Khi sử dụng khóa giả thuyết chính xác, CPA xác định được một đỉnh với giá trị tương quan là 0,245 tại mẫu thứ 1735 của vệt công suất, giá trị tương quan này cao hơn rất nhiều so với các giá trị tương quan tại các mẫu còn lại. Ngược lại, với các khóa không chính xác, các giá trị tương quan không có nhiều sự khác biệt và thay đổi trong khoảng [−0,08; 0,087]. Như vậy, bằng cách tính toán giá trị tương quan giữa mô hình tiêu thụ năng lượng và mức tiêu thụ năng lượng thực tế thu thập được từ thiết bị, CPA có thể xác định được khóa chính xác của thuật toán mã mật được thực thi trên thiết bị đó.

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

Giữa quá trình huấn luyện trong học sâu và các giai đoạn của tấn công kênh bên có một sự tương đồng làm cho học sâu có thể ứng dụng trong tấn công kênh bên. Cụ thể, về mặt lý thuyết thì hai giai đoạn của tấn công kênh bên lập mẫu bao gồm giai đoạn lập mẫu và giai đoạn tấn công có thể chuyển đổi thành giai đoạn huấn luyện và giai đoạn kiểm tra trong học sâu. Trong giai đoạn lập mẫu, người ta sẽ tìm cách xây dựng một hàm xác suất giữa mức tiêu thụ năng lượng của thiết bị mẫu và các giá

trị trung gian tương ứng. Theo đó, các vệt công suất thu được từ các thiết bị mẫu sẽ được phân loại vào các nhóm khác nhau dựa vào các giá trị trung gian tính toán được. Giai đoạn tấn công trong tấn công lập mẫu cũng như giai đoạn kiểm tra hiệu quả của mô hình học sâu được tiến hành bằng cách sử dụng mô hình đã huấn luyện trên một tập dữ liệu mới để đánh giá hiệu quả của mô hình đó dựa vào các thước đo huấn luyện đã xác định. Tuy nhiên, một trong các hạn chế của tấn công lập mẫu là khả năng áp dụng mô hình đã huấn luyện để tấn công các thiết bị sử dụng các thuật toán mã hóa khác nhau do sự khác biệt về số lượng các nhóm trong không gian phân loại của từng thuật toán. Ví dụ, trong thuật toán AES-128 thì đơn vị xử lý nhỏ nhất được sử dụng là các byte (8 bit), tức là sẽ có 256 nhãn phân loại. Tuy nhiên, đối với thuật toán PRESENT thì chỉ có 16 nhãn phân loại do thuật toán này sử dụng đơn vị xử lý nhỏ nhất là nibble (nửa byte). Chính vì sự khác biệt về kích thước của đơn vị xử lý nhỏ nhất mà một mô hình học sâu được huấn luyện trên dữ liệu của một thuật toán cụ thể sẽ không thể áp dụng trực tiếp cho các thuật toán khác. Điều này làm giảm tính tổng quát của phương pháp tấn công kênh bên lập mẫu.

Ngược lại, tấn công không lập mẫu không yêu cầu quá trình huấn luyện trước để xây dựng mô hình xác suất giữa mức tiêu thụ điện năng và các giá trị trung gian. Thay vào đó, phương pháp này sử dụng các mô hình tiêu thụ năng lượng (mô hình rò rỉ) như mô hình HW hoặc HD để ước lượng mức tiêu thụ năng lượng của các giá trị trung gian giả định (các giá trị khóa giả thuyết). Sau đó, mức độ tương quan giữa mô hình tiêu thụ năng lượng với dữ liệu tiêu thụ năng lượng thực tế thu thập được tính toán để xác định khóa chính xác. Điều này có nghĩa là sự thay đổi các bit của giá trị trung gian sẽ dẫn đến việc thay đổi của mức tiêu thụ năng lượng tương ứng. Do đó, nếu chúng ta có thể sử dụng mô hình học sâu để học các mối liên hệ giữa mức tiêu thụ năng lượng và các giá trị trung gian giả định thì sẽ cho phép mô hình có thể áp dụng được đối với các thuật toán mã mật khác nhau được sử dụng trên cùng một thiết bị [14]. Phần tiếp theo trình bày tóm tắt về một số mô hình học sâu phổ biến được sử dụng trong tấn công kênh bên không lập mẫu.

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 $(T_i)_{1\leq i\leq N}$. 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 $(d_i)_{1\leq i\leq N}$ 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: $DL\ (Net, X, Y_i, n_e)$

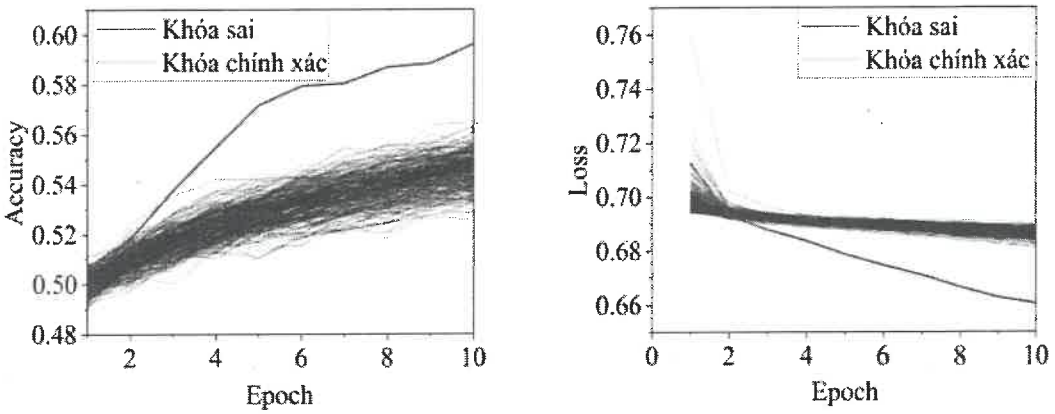
7: **end for**

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.

Hình 2.3 mô tả kết quả tấn công không lập mẫu sử dụng DDLA sử dụng kiến trúc MLP tương tự như kiến trúc MLP_{exp} trong [23] bao gồm 2 lớp ẩn chứa lần lượt 20 và 10 nơ-ron sử dụng hàm kích hoạt ReLU và lớp đầu ra gồm 2 nơ-ron sử dụng hàm kích hoạt softmax. Việc tấn công được thực hiện trên tập dữ liệu gồm 20.000 vệt công suất. Theo đó, các tham số huấn luyện của mô hình trên tập dữ liệu tạo bởi các

khóa không chính xác sẽ thấp và biến động hỗn loạn (đường màu xanh trên đồ thị). Trong khi đó, độ chính xác và mất mát của mô hình khi huấn luyện trên tập dữ liệu tạo bởi khóa chính xác cao và ổn định hơn hẳn (đường màu đỏ trên đồ thị).



Hình 2. 3. Các thước đo đánh giá hiệu quả của mô hình DDLA trong tấn công kênh bên không lập mẫu

Một nhược điểm lớn của phương pháp tấn công kênh bên sử dụng DDLA là phải thực hiện huấn luyện với số lần bằng số lượng khoá các khóa giả thuyết k. Ví dụ, đối với thuật toán AES-128 sử dụng phương pháp tấn công "chia để trị", số lượng khóa giả thuyết cho một byte là 256, như vậy số lần huấn luyện để tìm ra toàn bộ khóa là 4096 lần. Ngoài ra, nếu mạng nơ-ron cần phải huấn luyện n_e lần (số lượng epoch) mới tìm ra được khóa chính xác thì tổng số lần xuất hiện sẽ là $k \times n_e$. Tổng số lần huấn luyện này tạo ra chi phí nhiều hơn so với tấn công lập mẫu [13]. Bên cạnh đó, thuật toán 1 chưa đề cập đến việc xử lý dữ liệu ở đầu vào và đầu ra như việc tính toán các giá trị trung gian phục vụ cho gắn nhãn dữ liệu hoặc việc giám sát các thước đo sau mỗi lần huấn luyện. Đây cũng là những chi phí tính toán không thể bỏ qua khi thực hiện tấn công. Hơn nữa, việc xác định được tổng số lần huấn luyện cần thiết để phân biệt được khoá chính xác trong tập hợp các khóa giả thuyết trước khi thực hiện tấn công cũng là một thách thức rất lớn. Cuối cùng, so với tấn công lập mẫu thì tấn công không lập mẫu có thể yêu cầu một mạng nơ-ron phức tạp hơn rất nhiều để có thực hiện tấn công thành công [13]. Điều này chỉ ra rằng chi phí để thực hiện tấn công DDLA là quá cao.

2.4. Kết luận chương 2

Chương 2 đã hệ thống hóa và phân tích ba kỹ thuật tấn công không lập mẫu tiêu biểu – DPA, CPA và DDLA. Mỗi phương pháp mang ưu và nhược điểm riêng, trong đó DPA sử dụng sự khác biệt trung bình giữa các nhóm dấu vết để suy luận khóa bí mật. Phương pháp này đơn giản, dễ triển khai và không yêu cầu mô hình hóa chi tiết thiết bị. Tuy nhiên, DPA phụ thuộc mạnh vào việc phân nhóm chính xác và thường kém ổn định khi mức nhiễu cao hoặc khi thiết bị sử dụng các biện pháp che giấu năng lượng. CPA khắc phục nhiều hạn chế của DPA nhờ sử dụng hệ số tương quan Pearson giữa mô hình rò rỉ và dấu vết đo được. CPA có khả năng đánh giá mức phù hợp giữa lý thuyết và thực nghiệm tốt hơn, từ đó cho độ chính xác cao hơn trong đa số trường hợp. CPA trở thành chuẩn tham chiếu phổ biến trong phân tích công suất tiêu thụ, đặc biệt khi mô hình rò rỉ tuân theo Hamming Weight hoặc Hamming Distance. Dù vậy, hiệu năng của CPA vẫn phụ thuộc vào việc lựa chọn mô hình rò rỉ phù hợp với thiết bị. DDLA đại diện cho hướng tiếp cận hiện đại, sử dụng mô hình học sâu để tự động học đặc trưng và thiết lập mối liên hệ giữa dấu vết và khóa bí mật mà không cần giả định mô hình rò rỉ. Nhờ lợi thế từ các mạng nơ-ron sâu, DDLA có khả năng xử lý dữ liệu nhiễu cao, thích ứng tốt với thiết bị có hành vi rò rỉ phức tạp hoặc không tuyến tính. Tuy nhiên, phương pháp này đòi hỏi khối lượng lớn dấu vết huấn luyện, thời gian xử lý cao và phụ thuộc vào thiết kế mô hình học sâu.

Qua việc so sánh các phương pháp, có thể khẳng định rằng xu hướng phát triển SCA hiện nay đang chuyển dịch từ các mô hình toán học thuần túy sang các mô hình trí tuệ nhân tạo có khả năng học tự động từ dữ liệu. Các kết quả này là nền tảng quan trọng cho việc đánh giá hiệu quả mô phỏng tấn công vào thuật toán AES-128 được trình bày trong Chương 3.

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 tập trung đánh giá hiệu quả của các kỹ thuật tấn công kênh bên không lập mẫu phổ biến sử dụng các tập dữ liệu kênh bên thu thập được từ quá trình thực thi thuật toán mã hóa AES-128 trên thiết bị có tài nguyên giới hạn, tương tự các bộ xử lý trong thiết bị IoT. Hai kỹ thuật chủ yếu được khảo sát trong đề án đó là CPA (phân tích tương quan công suất) và DDLA (phân tích vi sai dựa trên mô hình học sâu). Thực nghiệm sử dụng CPA vào thuật toán AES-128 ở các thiết bị không có biện pháp bảo vệ và có biện pháp bảo vệ để phân tích kết quả về độ chính xác, độ tin cậy, thời gian và khả năng khôi phục khóa. Sau đó sẽ sử dụng DDLA để khai thác, xử lý những điều mà CPA chưa làm được. Các bước chính bao gồm: Xử lý dữ liệu, tạo tập dataset, thực hiện tấn công vào thuật toán AES-128 với hai kỹ thuật tấn công bằng ngôn ngữ lập trình Python, phân tích kết quả về độ chính xác, độ tin cậy, thời gian và khả năng khôi phục khóa, so sánh với kết quả của các nghiên cứu trước đó tại Việt Nam.

3.1. Tấn công SCA sử dụng CPA

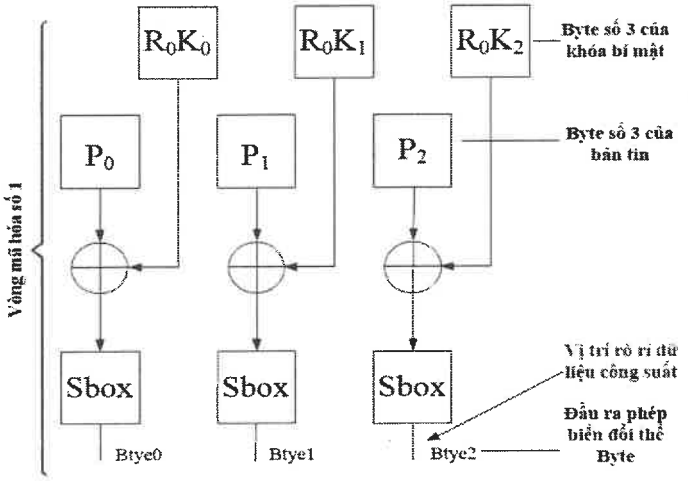
3.1.1. Mục tiêu và cách tiếp cận

Thực nghiệm trong đề án được thực hiện trong điều kiện đo gần với môi trường triển khai thực tế của thiết bị IoT. Thiết bị được cấp nguồn ổn định, không áp dụng các biện pháp che giấu hoặc làm nhiễu công suất chủ động. Môi trường đo tồn tại nhiễu tự nhiên phát sinh từ đặc tính phần cứng và các tác động bên ngoài. Tín hiệu kênh bên được thu thập với khả năng đồng bộ ở mức chức năng, cho phép xác định các thời điểm xử lý chính của thuật toán AES-128 mà không yêu cầu đồng bộ chính xác đến từng chu kỳ xung nhịp. Các điều kiện này nhằm bảo đảm tính khả thi của thí nghiệm và phản ánh sát bối cảnh triển khai thực tế của thiết bị IoT.

Trong phần này, kỹ thuật 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. CPA được lựa chọn làm cơ sở để so sánh với phương pháp DDLA ở các mục tiếp theo.

CPA sẽ xác định mối tương quan giữa giá trị trung gian trong thuật toán AES và dấu vết công suất thu được từ thiết bị IoT khi thực thi mã hóa. Giá trị trung gian này có thể được mô hình hóa bằng một mô hình công suất tiêu thụ thỏa mãn những đặc điểm sau: Thứ nhất, mô hình này liên quan trực tiếp dữ liệu đầu vào hoặc đầu ra, nhằm mục đích dễ dàng cho việc xây dựng mô hình. Thứ hai, mô hình sử dụng vị trí trong thuật toán thể hiện sự tương quan giữa công suất tiêu thụ ứng với từng giá trị của dữ liệu (vị trí rò rỉ dữ liệu công suất). Từ những đặc điểm trên, tôi lựa chọn sử dụng vị trí thực hiện biến đổi SubByte (Sbox) kết hợp với trọng số Hamming tại vòng mã hóa đầu tiên, các đặc điểm thỏa mãn điều kiện được mô tả ở Hình 3.1. Đầu ra sau đó được thực hiện tính trọng số Hamming Weight (HW) của đầu ra S-Box: $L = HW(SBOX[P_i \oplus K_i])$ với P_i là byte bản rõ vị trí i, K_i là là byte khóa AES đang thử. Nếu mối tương quan giữa mô hình rò rỉ và tín hiệu thực của thiết bị đủ lớn, đỉnh tương quan tại khóa đúng sẽ nổi bật so với 255 giá trị khóa còn lại. Sau đó, CPA tính hệ số tương quan Pearson giữa chuỗi HW và từng điểm trên vệt công suất. Nếu thiết bị không có cơ chế bảo vệ, tại khóa đúng sẽ xuất hiện đỉnh tương quan vượt trội, mức từ 0.3–0.8 tùy theo nhiễu và số lượng trace.



Hình 3. 1. Vòng mã hóa số 1 thích hợp cho xây dựng mô hình tiêu thụ công suất

3.1.2. Xử lý dữ liệu và tạo tập dataset

Dataset trong đề án được xây dựng thông qua đo đạc trực tiếp mức tiêu thụ năng lượng trên chip CW1173 bằng ChipWhisperer. Thiết lập thực nghiệm bao gồm board mục tiêu CW1173 thực hiện thuật toán AES-128, board thu thập tín hiệu ChipWhisperer tích hợp ADC lấy mẫu tốc độ cao và máy tính điều khiển kết nối qua cổng USB.

Tín hiệu công suất được đo tại đường cấp nguồn lõi xử lý của vi điều khiển, thông qua điện trở shunt tích hợp trên mạch và được ADC thu thập với tần số lấy mẫu cao. Quá trình đo được trigger đồng bộ với thời điểm bắt đầu mã hóa AES, nhằm đảm bảo mỗi lần mã hóa tương ứng với một vệt công suất hoàn chỉnh.

Quy trình thu thập dữ liệu được thực hiện như sau: firmware AES-128 được nạp lên CW1173; máy tính điều khiển cấu hình tham số đo và sinh các plaintext 128-bit gửi đến board mục tiêu. Với mỗi lần mã hóa, CW1173 thực hiện AES-128, dòng điện tiêu thụ thay đổi theo dữ liệu nội bộ và được ghi nhận thành vệt công suất. ADC thu thập toàn bộ tín hiệu công suất và máy tính lưu lại cặp dữ liệu gồm plaintext, ciphertext, khóa và vệt công suất tương ứng.

Quy trình này được lặp lại hàng nghìn lần với các plaintext khác nhau nhằm tạo ra tập dataset đầy đủ, phục vụ cho các thực nghiệm tấn công kênh bên bằng CPA và cả DDLA trong đề án.

Ở thực nghiệm này, các vệt công suất dữ liệu nguyên bản được thu từ mạch CW1173 Lite khi AES được thực thi mà không có bất kỳ cơ chế che giấu nào. Mỗi vệt công suất có 500 mẫu thời gian, độ lệch pha gần như bằng 0 giữa các trace, tín hiệu chứa cấu trúc lặp lại theo chu kỳ. Tập vệt chuẩn này là nền tảng để đánh giá khả năng thu hồi khóa của CPA trong điều kiện lý tưởng. Quá trình thực hiện thu thập power trace từ CW1173 được mô tả như Hình 3.2. Dữ liệu máy tính nhận về là một tập hợp gồm nhiều file dữ liệu có dạng file.npy được liệt kê trong Bảng 3.1.

Để mô phỏng thiết bị IoT được bảo vệ, tập vệt công suất được xử lý lại bằng thuật toán desync, theo đó mỗi trace bị dịch ngẫu nhiên từ 0 đến 20 mẫu. Kỹ thuật này làm phá vỡ tính đồng bộ theo trục thời gian. Số mẫu còn lại trong mỗi trace: 481,

cấu trúc tín hiệu AES bị biến dạng mạnh, điểm rò rỉ của S-Box không còn nằm cùng vị trí giữa các trace.

Thuật toán 2: Tạo vệt công suất bị mất đồng bộ (de-synchronized power traces)

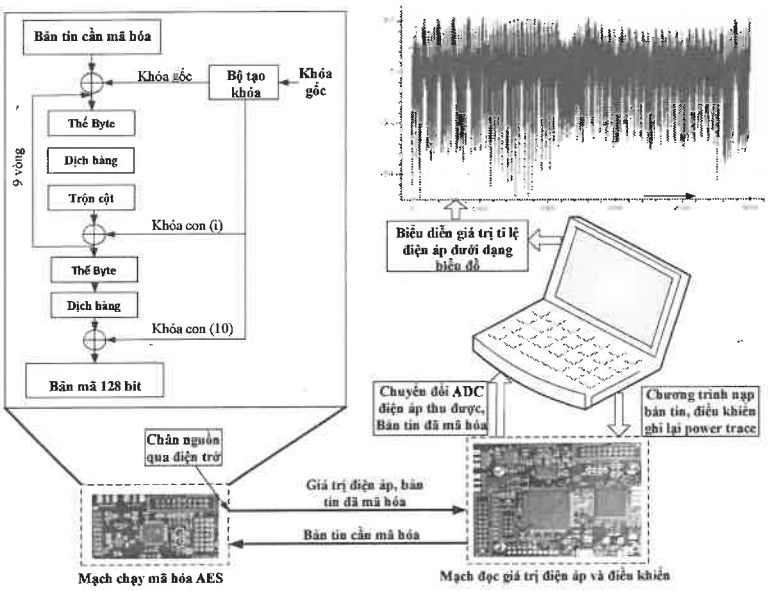
Input: N: số lượng vệt công suất; T: tập các vệt công suất ban đầu; ψ: giá trị xáo trộn (shift_value); β: số mẫu cần lấy sau khi dịch.

Output: T_{shifted}: tập vệt công suất đã bị xáo trộn.

Thuật toán:

1. **for** $i \in 1: N$ **do**
2. Chọn ngẫu nhiên một lượng dịch: sh = randi ([1: ψ])
3. Lấy đoạn tín hiệu sau khi dịch: T_{shifted} = T (i, $sh : sh + \beta$)
4. Gán vào tập dữ liệu đầu ra: T_{shifted} (i, :) = T_{shift}
5. Đặt lại giá trị dịch: sh = 0
6. **end for**

Sau khi xử lý, dataset được chuyển về dạng phù hợp cho Python (NumPy array).



Hình 3. 2. Quá trình thu thập power trace trên ChipWhisperer W1173

Để thực hiện Thuật toán 2, sử dụng 10000 power trace, mỗi power trace 500 mẫu thu thập từ CW1173. Kết quả đầu ra chương trình đã phân tích được chính xác 16 byte của khóa bí mật ở dạng thập phân như sau:

43-136-21-22-40-174-210-166-171-247-21-136-9-207-79-60

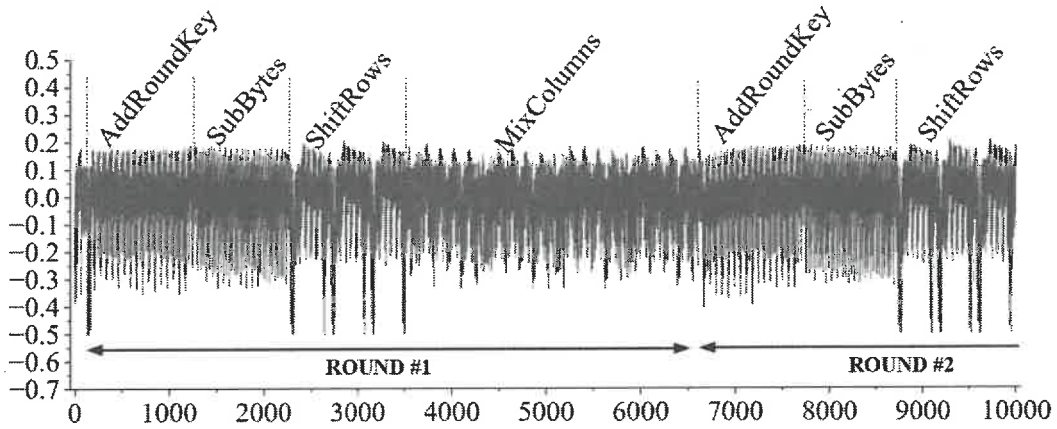
Bảng 3. 1. Các file lưu kết quả CW1173 trả về phục vụ cho phân tích trên Python

TT	Tên file	Nội dung dữ liệu
1	*_keylist.npy	Khóa bí mật (16 byte cố định)

2	*_textin.npy	*N* bản tin cho mã hóa (*N*×16 byte ngẫu nhiên)
3	*_textout.npy	N bản tin đầu ra đã được mã hóa
4	*_traces.npy	Mảng dữ liệu về tỉ lệ giá trị điện áp đo được trong quá trình mã hóa *N* bản tin

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. Hình 3.3 minh họa một vệt công suất thu thập được từ mạch CW1173, trong đó có thể quan sát thấy và phân biệt tương đối vùng của vệt công suất tương ứng với các hàm toán học thực thi thuật toán AES.



Hình 3. 3. Một vệt công suất thu thập được từ mạch CW1173

Trong phạm vi đề án này, giá trị trung gian được xác định là giá trị đầu ra của hàm Sbox, chính vì vậy có thể thấy đối với CW1173, có thể tập trung CPA vào 1000 mẫu bắt đầu từ mẫu thứ 1000 cho đến mẫu thứ 2000. Tuy nhiên, để thuận tiện cho việc tính toán mô phỏng, cũng như giới hạn tài nguyên phần cứng của máy tính thực hiện thí nghiệm, đề án lựa chọn 500 mẫu bắt đầu từ 1000 cho đến mẫu 1500 để khảo sát cho toàn bộ các thí nghiệm.

Pipeline tấn công gồm:

Tính giá trị trung gian theo mô hình AES: $v = SBox\ (p \oplus k_{guess})$ (3.1)

Xây dựng mô hình công suất dựa trên Hamming Weight: $HW(v)$

Tính hệ số tương quan Pearson giữa mô hình công suất và từng điểm mẫu trong trace.

Lựa chọn khóa có tương quan lớn nhất làm khóa đúng.

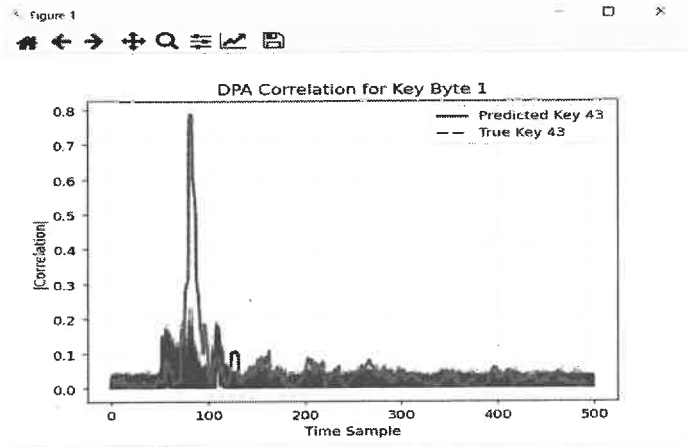
Để minh họa trường hợp tấn công thiết bị đề án sử dụng tập dữ liệu thu được từ ChipWhisperer. Tất cả các mô hình đề xuất trong đề án đều được sử dụng để tấn công trong trường hợp này theo quy trình chung như sau: sử dụng mô hình gốc để tấn công một byte của khóa bí mật; sau khi tìm được byte khóa, mô hình huấn luyện được lưu lại và được sử dụng tấn công các byte còn lại của chuỗi khóa. Hiệu quả của các hình thức tấn công được đánh giá bằng cách so sánh thời gian tấn công và tỉ lệ tấn công thành công trong cùng điều kiện tấn công.

3.1.4. Kết quả thực nghiệm CPA

a) Tấn công thiết bị không được bảo vệ

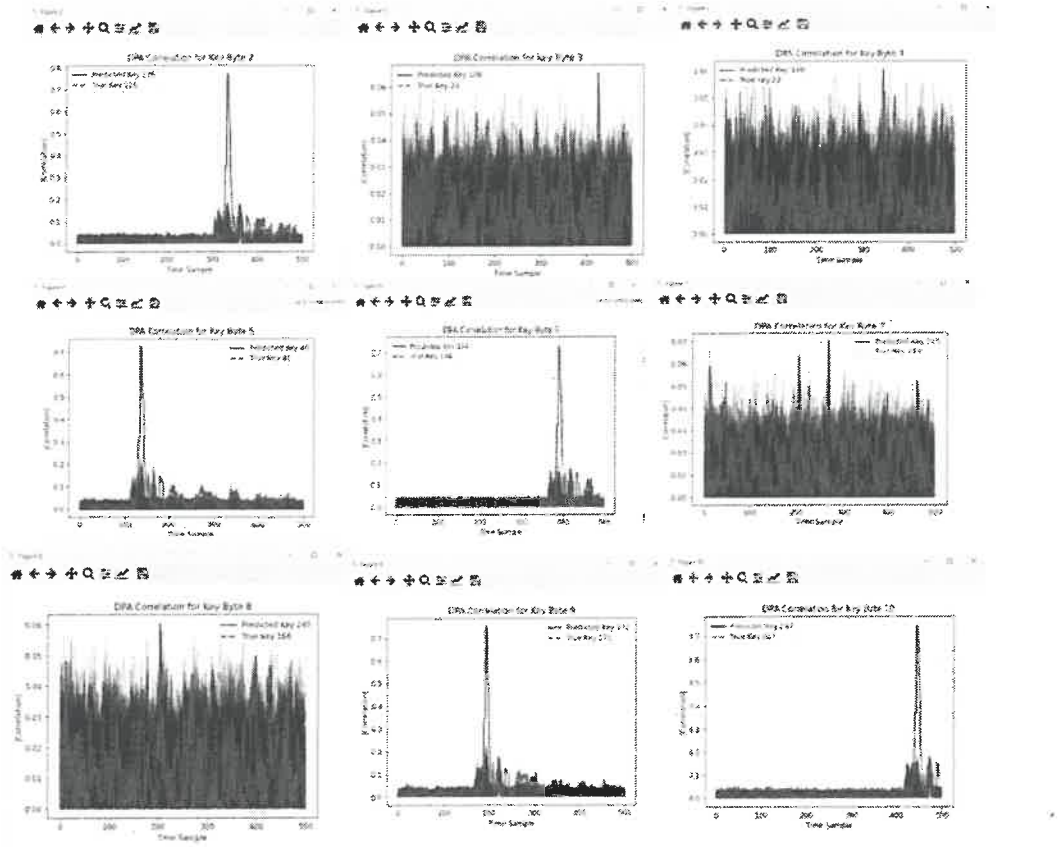
Biến desync=0

Kết quả



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

Có thể thấy, trong khoảng 500 vệt công suất, có duy nhất 1 vị trí thể hiện tương quan lớn nhất với sự biến thiên của điện năng tiêu thụ và giá trị xử lý hàm Sbox, điều này chứng minh rằng vị trí rò rỉ lớn nhất của thuật toán AES trong điều kiện thí nghiệm này là hàm xử lý phi tuyến, một trong những đặc điểm tạo nên sức mạnh cho thuật toán AES. Tuy nhiên, có thể thấy AES rất dễ bị khai thác với hình thức tấn công kênh bên.



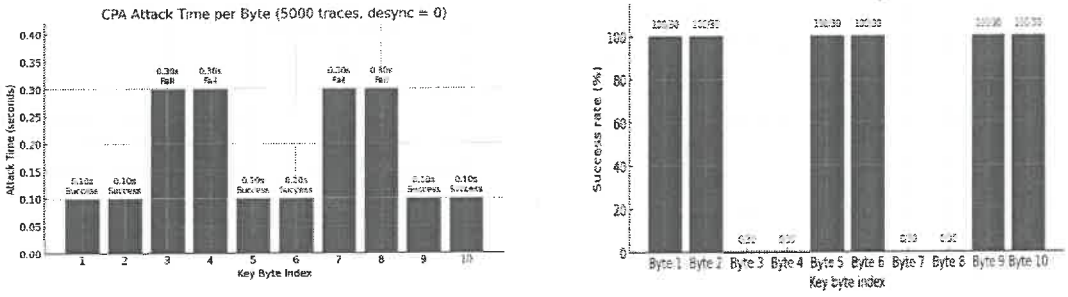
Hình 3. 5. Kết quả mô phỏng tấn công SCA sử dụng CPA vào Byte 2 đến Byte 10 với các thiết bị không được bảo vệ tấn công kênh bên

Tiếp tục khảo sát với các byte tiếp theo Hình 3.5 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. Điều này thể hiện đúng bản chất của thuật toán AES khi được thực thi bằng chương trình phần mềm. Lúc này 500 mẫu đầu tiên của quá trình Sbox tương ứng với quá trình xử lý 2 cột trong ma trận dữ liệu được thuật toán AES sắp xếp. 2 cột còn lại trong ma trận tương ứng với các Byte khóa mà CPA chưa khôi phục được. Điều này thể hiện việc khoanh vùng vệt công suất để đạt được kết quả tấn công hiệu quả.

Nhìn chung, 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. Cụ thể, sử dụng CPA để tấn công các byte 1, 2, 5, 6, 9, 10 của khóa bí mật với tập dữ liệu đầu vào 5000 vệt công suất. Thời gian tấn công trung bình và tỷ lệ tấn công thành công được thể hiện ở Hình 3.6.

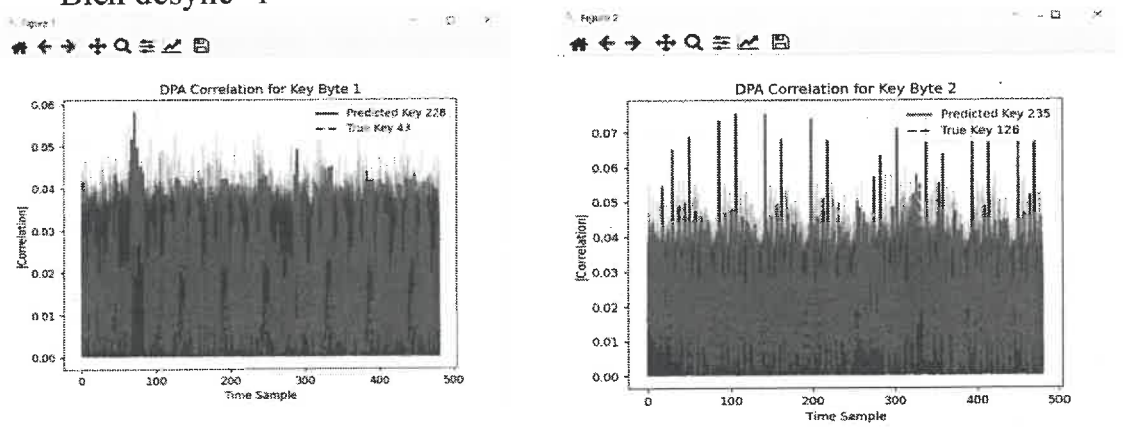


a) b)

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

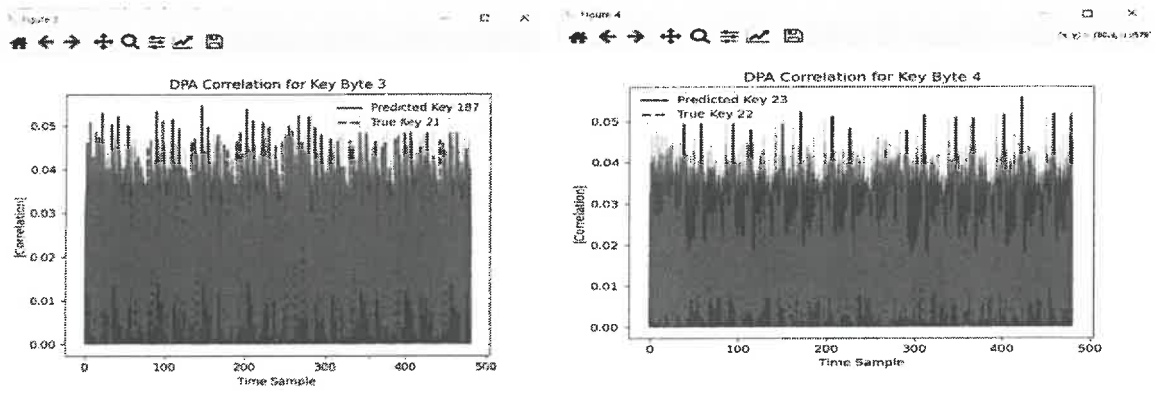
Trong Hình 3.6a,b có thể thấy rõ rằng, trong điều kiện lý tưởng tấn công các thiết bị không được áp dụng các biện pháp phòng chống tấn công kênh bên, CPA hoàn toàn phát huy tác dụng với thời gian thực thi nhanh chóng và ổn định $\approx 0{,}10s$ với tỷ lệ tin cậy cho khóa chính xác là 100%. Nhìn chung, CPA hiệu quả trong điều kiện lý tưởng.

b) Tấn công thiết bị được bảo vệ
Biến desync=1



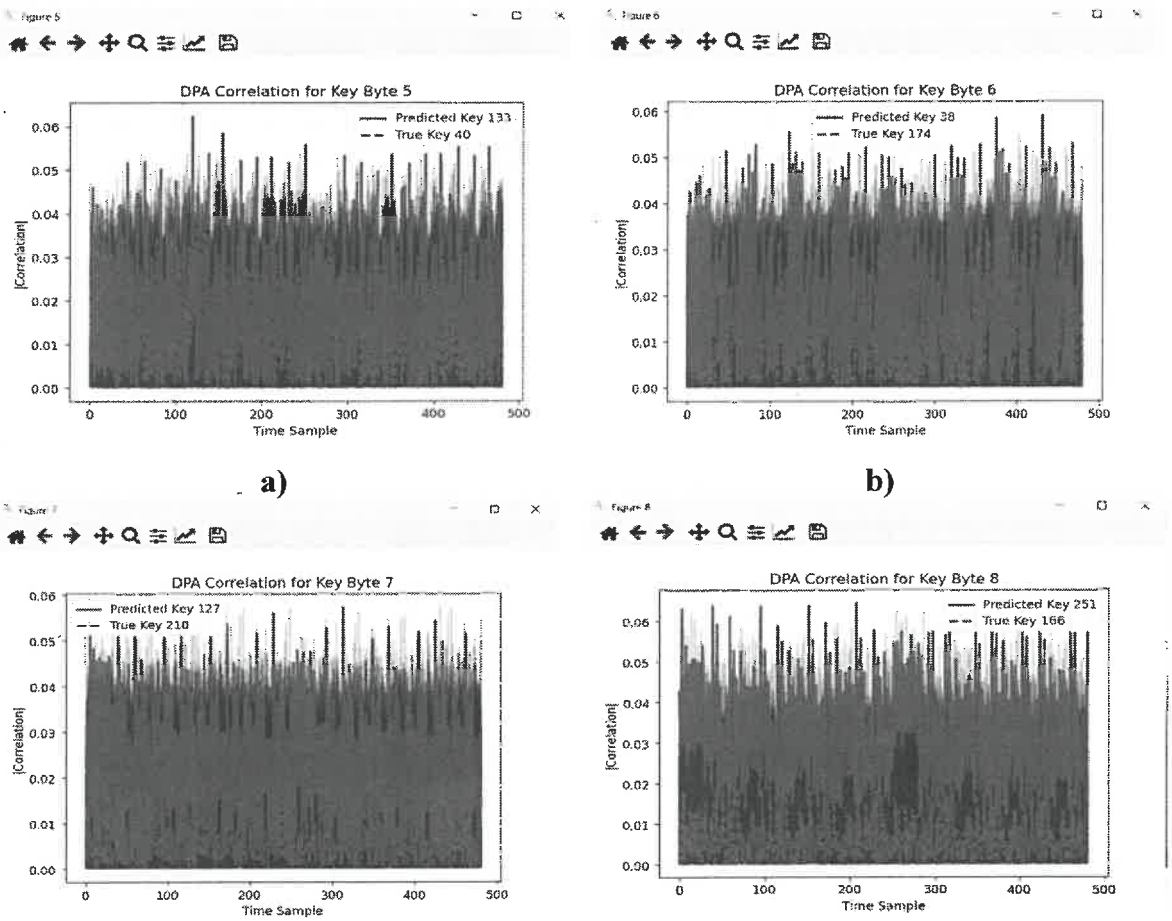
a) **b)**

Hình 3. 7. Kết quả mô phỏng tấn công SCA sử dụng CPA vào byte 1 (a), Byte 2 (b)



a) b)

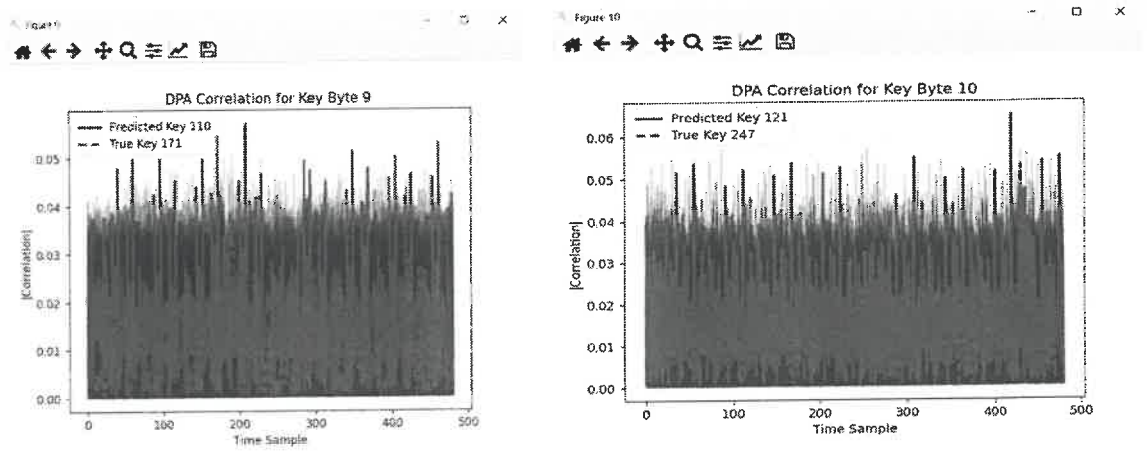
Hình 3. 8. Kết quả mô phỏng tấn công SCA sử dụng CPA vào Byte 3 (a), Byte 4 (b) với các thiết bị được bảo vệ tấn công kênh bên



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

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. Điều này là do CPA phụ thuộc tuyệt đối vào vị trí thời gian. Trong khi đó, desync phá hủy sự đồng nhất này, khiến mô hình tương quan trở nên vô nghĩa. Từ đó có thể thấy CPA hoạt động tốt trong môi trường có nhiễu thấp–trung bình nhưng giảm hiệu quả khi mức nhiễu tăng. Điều này tạo nền tảng để so sánh với phương pháp DDLA trong mục 3.2.

3.2. Tấn công SCA sử dụng DDLA

3.2.1. Mục tiêu và cách tiếp cận

DDLA (Deep-learning-based Differential Leakage Analysis) là kỹ thuật tấn công kênh bên không lập mẫu hiện đại, dựa trên học sâu để khai thác trực tiếp sự biến thiên của tín hiệu rò rỉ mà không cần mô hình công suất như CPA. 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

Dữ liệu từ ASCAD/CW305 được chuẩn hóa hoặc lọc đạo hàm bậc nhất để làm nổi bật tín hiệu rò rỉ. Sau đó, dataset được chuyển đổi sang dạng tensor phù hợp với mô hình học sâu.

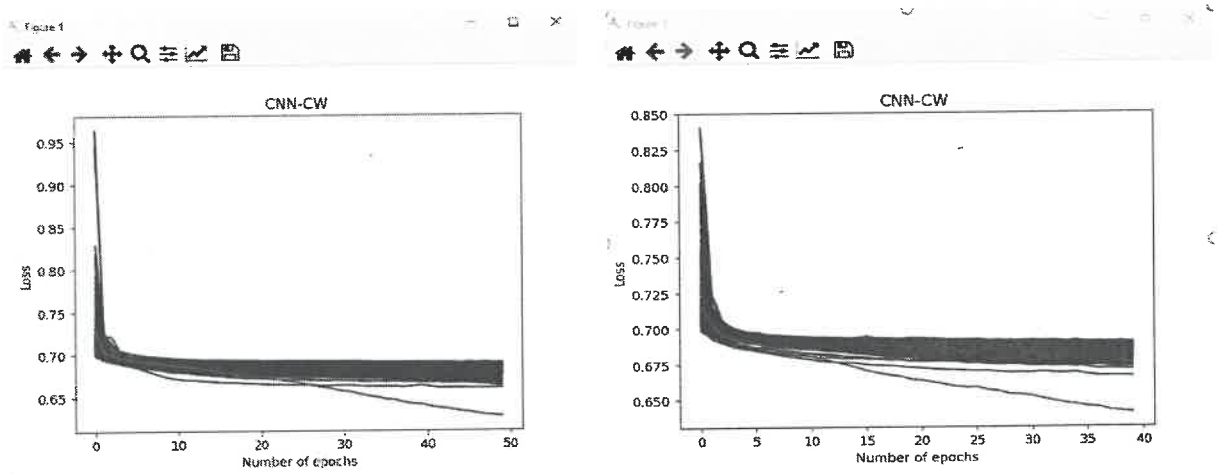
Pipeline tiền xử lý gồm: chuẩn hóa từng trace; lấy đạo hàm để tăng độ nhạy tín hiệu; giảm chiều (nếu cần); chia dữ liệu thành tập train/test (nếu mô hình DDLA có phase train nhẹ).

3.2.3. Thực thi DDLA bằng Python

Mô hình DDLA được triển khai bằng TensorFlow, gồm các lớp CNN nhằm học các đặc trưng không tuyến tính trong tín hiệu rò rỉ.

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

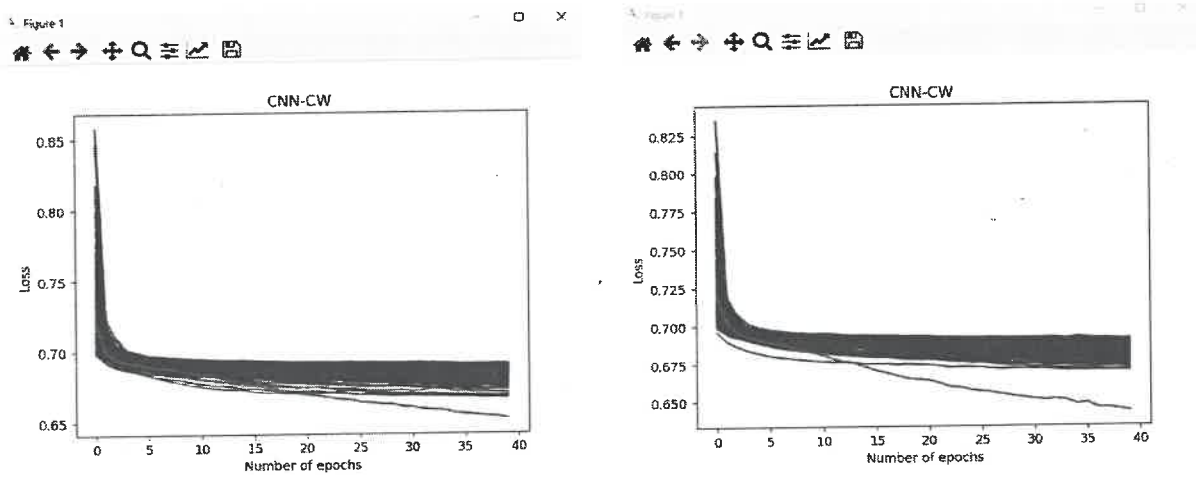


a) b)

Hình 3. 11. Kết quả tấn công Byte 1 (a); Byte 2 (b)

Nhìn vào Hình 3.11a. kết quả có thể thấy đường khóa đúng (đỏ) có xu hướng giảm loss liên tục theo số epoch. Khóa sai (xanh) dao động quanh mức loss ổn định 0.68–0.70. 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.



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.

Tại epoch 40, loss khóa đúng của byte 2-10 đạt mức khoảng 0.64–0.65, trong khi nhóm sai vẫn quanh 0.68–0.70 và khoảng cách giữa khóa đúng và phần lớn khóa sai giữ ở mức ổn định.

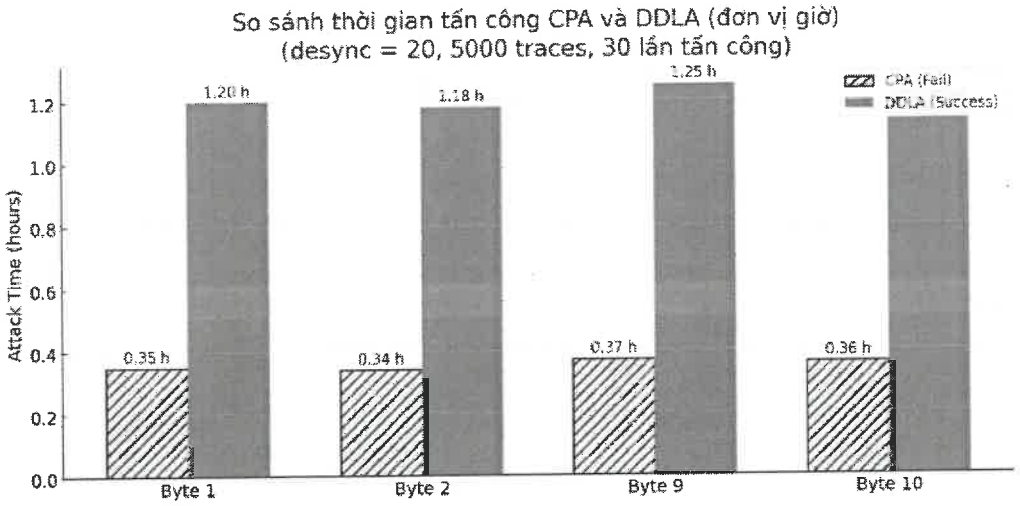
Từ phân tích các byte ngẫu nhiên 1, 2, 9, 10, có thể rút ra một số kết luận kỹ thuật quan trọng:

Đường khóa đúng luôn giảm nhanh hơn và đạt giá trị thấp hơn mọi khóa sai.

Chỉ cần quan sát loss cuối epoch (hoặc trung bình 5 epoch cuối) là đủ quyết định khóa.

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.

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 DDLA và so sánh với CPA. Cụ thể, sử dụng DDLA để tấn công các byte 1, 2, 9, 10 của khóa bí mật với tập dữ liệu đầu vào 5000 vệt công suất, desync = 20. Thời gian tấn công trung bình và tỷ lệ tấn công thành công được thể hiện ở Hình 3.13.

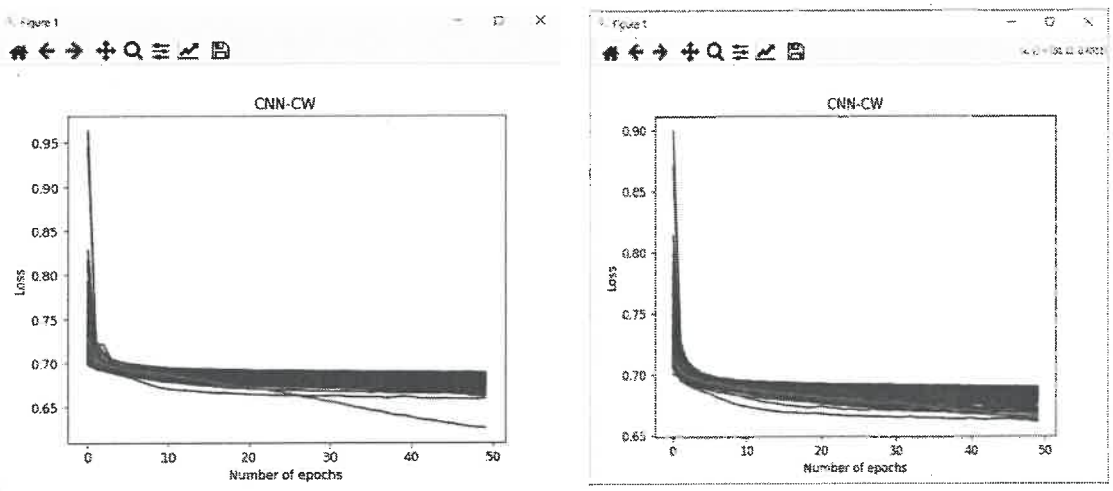


Hình 3. 13. So sánh thời gian tấn công trung bình và tỷ lệ tấn công thành công của CPA và DDLA

Trong Hình 3.13, cho thấy CPA thời gian tấn công thấp hơn (khoảng 0.34–0.37 giờ), tuy nhiên CPA luôn thất bại, nên thời gian ngắn này không mang ý nghĩa về hiệu quả, vì kết quả không thu được khóa mà chỉ thể hiện rằng lượng tính toán thấp. Đối với DDLA thời gian tấn công dài hơn, dao động 1.18–1.25 giờ, kết quả này

minh họa nhược điểm lớn nhất của DDLA đó là cần huấn luyện mô hình học sâu với các tập dataset con ứng với từng khóa giả thuyết, cụ thể là 256 lần huấn luyện tương đương 256 khóa giả thuyết cho thuật toán AES-128. 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.



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

Hình 3.14 cho thấy đường loss tương ứng với khóa đúng (đường màu đỏ) giảm sâu và giảm đều theo số epoch, trong khi đó, các đường loss của khóa sai (đường màu xanh) dao động quanh một ngưỡng cao hơn, cho thấy mô hình phân biệt được khóa đúng với sai. Sau khoảng 20–25 epoch, sự tách biệt giữa khóa đúng và sai trở nên rõ rệt. Hình 3.14b, với trace shift = 50 Đường loss của khóa đúng không còn giảm sâu như trường hợp trace shift = 20, các đường loss của khóa sai gần như chồng lên hoàn toàn với đường loss của khóa đúng, không xuất hiện sự tách biệt giữa khóa đúng và

khóa sai dù huấn luyện đến 40–50 epoch, mô hình không "học" được tín hiệu rò rỉ nào đủ ổn định để phân biệt khóa. Đây là dấu hiệu rõ ràng của hiện tượng mất đồng bộ nghiêm trọng, khiến các đặc trưng rò rỉ sau S-box không còn mang tính tuần hoàn hoặc định vị cố định theo thời gian, cụ thể cụm sample chứa leakage vẫn nguyên vẹn, nó chỉ bị dịch đi theo trục thời gian một khoảng lớn hơn, dẫn đến Kernel CNN vẫn có thể nhìn thấy leakage, nhưng ở phạm vị trải rộng hơn nhiều, Gradient tại mỗi vị trí rất yếu vì phân bố vị trí của leakage giữa các trace không còn tập trung như với Desync = 10, 20. Điều này dẫn đến kết quả attack không thành công. DDLA thất bại – không xác định được khóa chính xác khi mức xáo trộn đạt đến 50 mẫu.

3.3. Phân tích, đánh giá hiệu quả

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

Việc đánh giá hiệu năng các kỹ thuật tấn công kênh bên không lập mẫu đòi hỏi xem xét đồng thời các yếu tố thời gian xử lý, số lần lặp cần thiết để hội tụ, cũng như khả năng chịu đựng nhiễu và các biện pháp phòng vệ của thiết bị. Trong phần này, hai kỹ thuật CPA và DDLA dựa trên học sâu được đặt lên cùng một thang đánh giá nhằm chỉ ra ưu - nhược điểm của từng phương pháp trong các bối cảnh tấn công khác nhau.

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. Các đồ thị tương quan cho thấy những đỉnh sắc nét xuất hiện đúng vị trí thời gian mà vòng S-box hoạt động, minh chứng khả năng khôi phục khóa nhanh chóng và chính xác. Trong điều kiện lý tưởng này, CPA chỉ cần vài giây để hoàn tất một truy vấn và tái hiện byte khóa đúng.

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. Các đồ thị thực nghiệm trên thiết bị có bảo vệ cho thấy toàn bộ đường tương quan chìm trong nhiễu, biên độ chỉ dao động trong khoảng 0.01–0.03, không xuất

hiện bất kỳ đỉnh nhận dạng nào. Mặc dù thời gian tính toán vẫn nhanh, kết quả thu được không còn mang bất kỳ giá trị thực tiễn nào. Điều này phản ánh nhược điểm cố hữu của CPA: phương pháp phụ thuộc chặt chẽ vào sự đồng bộ của vệt công suất và không có khả năng tự thích nghi trước các biến dạng tín hiệu do cơ chế bảo vệ gây ra.

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). Trong hầu hết các byte, đường loss của khóa đúng giảm đều đặn xuyên suốt quá trình huấn luyện, trong khi khóa sai dao động quanh một mức gần như cố định. Điều này được quan sát nhất quán trong các byte 1, 2, 9 và 10. Dù trong môi trường có bảo vệ, các mô hình CNN vẫn trích xuất được đặc trưng rò rỉ bị che giấu dưới nhiễu và xáo trộn.

Ở góc độ thời gian, DDLA rõ ràng tốn nhiều tài nguyên hơn: mỗi byte yêu cầu 40–50 epoch mới đạt mức hội tụ tốt và tổng thời gian huấn luyện mỗi byte có thể kéo dài vài giờ. Tuy nhiên, điểm quan trọng là kỹ thuật này vẫn tấn công thành công trên vệt công suất đã được bảo vệ – nhiệm vụ mà CPA hoàn toàn thất bại. Như vậy DDLA thích nghi và khai thác được các đặc trưng phi tuyến mà dạng bảo vệ dựa trên xáo trộn không thể loại bỏ hoàn toàn.

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.

Không dừng lại ở đó, 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. Đây là kết quả quan trọng, góp phần mở rộng hướng nghiên cứu tấn công kênh bên không lập mẫu trên thiết bị IoT sử dụng học sâu – một chủ đề tuy đầy tiềm năng nhưng còn rất ít công trình khai thác sâu trong bối cảnh trong nước 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 nhiện 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.

Tóm lại, dù DDLA cho thấy năng lực vượt trội so với CPA trong môi trường bảo vệ, đề án vẫn còn những điểm cần mở rộng, đặc biệt là về tối ưu thời gian, đánh giá trong điều kiện ít dữ liệu và khả năng tổng quát hóa trên đa thiết bị. Đây cũng là

những hướng nghiên cứu quan trọng để tiếp tục hoàn thiện kết quả trong các giai đoạn nghiên cứu tiếp theo.

3.4. Kết luận chương 3

Chương 3 đã trình bày quá trình mô phỏng, triển khai và đánh giá hiệu quả của hai kỹ thuật tấn công kênh bên không lập mẫu vào thuật toán AES-128 trên thiết bị IoT, bao gồm phương pháp tương quan công suất (CPA) và phương pháp học sâu DDLA, dựa trên tập dữ liệu thực thu thập từ thiết bị thật trong các điều kiện có và không có cơ chế bảo vệ.

Các kết quả thực nghiệm cho thấy CPA chỉ hoạt động hiệu quả trong môi trường lý tưởng, khi vệt công suất duy trì sự đồng bộ và mức nhiễu thấp. Khi thiết bị được áp dụng cơ chế làm lệch thời gian, phương pháp này mất hoàn toàn khả năng phân biệt khóa đúng do bản chất phụ thuộc chặt chẽ vào vị trí thời gian của rò rỉ. Điều này khẳng định rằng CPA không còn phù hợp để đánh giá an toàn cho các thiết bị IoT hiện đại, vốn thường tích hợp các biện pháp bảo vệ đơn giản nhưng hiệu quả.

Ngược lại, phương pháp DDLA cho thấy khả năng khai thác rò rỉ kênh bên ngay cả trong điều kiện dữ liệu bị xáo trộn và nhiễu mạnh. Thông qua việc phân tích đồ thị loss, mô hình học sâu thể hiện sự phân tách ổn định giữa khóa đúng và các khóa giả thuyết trên nhiều byte và nhiều lần lặp, cho thấy DDLA không phụ thuộc vào sự đồng bộ theo thời gian mà học trực tiếp các đặc trưng phi tuyến của tín hiệu rò rỉ.

Bên cạnh việc so sánh hiệu quả giữa hai phương pháp, chương 3 còn xác định được giới hạn hoạt động của DDLA thông qua việc thay đổi mức độ desynchronization. Kết quả cho thấy DDLA vẫn duy trì khả năng tấn công ở mức lệch thời gian trung bình, nhưng bắt đầu thất bại khi mức bảo vệ được tăng cường đủ mạnh. Phát hiện này có ý nghĩa thực tiễn quan trọng, giúp làm rõ ngưỡng bảo vệ cần thiết để giảm thiểu rủi ro từ các mô hình học sâu không lập mẫu.

Tổng hợp các kết quả đạt được cho thấy DDLA là phương pháp tấn công kênh bên phù hợp hơn so với CPA trong bối cảnh thiết bị IoT có bảo vệ, đồng thời phản ánh xu hướng chuyển dịch từ các kỹ thuật thống kê truyền thống sang các mô hình

học sâu trong nghiên cứu tấn công kênh bên hiện đại. Đây cũng là cơ sở khoa học cho các hướng nghiên cứu tiếp theo nhằm tối ưu mô hình, mở rộng đánh giá trên nhiều thiết bị và khảo sát hiệu quả trong điều kiện dữ liệu hạn chế.

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.

Như vậy, 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.

TÀI LIỆU THAM KHẢO

[1] "Introduction," in Power Analysis Attacks: Revealing the Secrets of Smart Cards. Boston, MA: Springer US, 2007, pp. 1-13.

[2] "Power Consumption," in Power Analysis Attacks: Revealing the Secrets of Smart Cards. Boston, MA: Springer US, 2007, pp. 27-60.

[3] A. Alipour, A. Papadimitriou, V. Beroulle, E. Aerabi, and D. Hély, "On the performance of non-profiled differential deep learning attacks against an aes encryption algorithm protected using a correlated noise generation based hiding countermeasure," presented at the Proceedings of the 23rd Conference on Design, Automation and Test in Europe, Grenoble, France, 2020.

[4] R. Benadjila, E. Prouff, R. Strullu, E. Cagli, and C. Dumas, "Deep learning for side-channel analysis and introduction to ASCAD database," Journal of Cryptographic Engineering, vol. 10, no. 2, pp. 163- 188, 2020/06/01 2020, doi: 10.1007/s13389-019-00220-8.

[5] Y. Bengio, "Learning Deep Architectures for AI," Foundations, vol. 2, pp. 1-55, 01/01 2009, doi: 10.1561/2200000006.

[6] E. Brier, C. Clavier, and F. Olivier, "Correlation Power Analysis with a Leakage Model," Berlin, Heidelberg, 2004: Springer Berlin Heidelberg, in Cryptographic Hardware and Embedded Systems - CHES 2004, pp. 16-29.

[7] S. Chari, J. R. Rao, and P. Rohatgi, "Template Attacks," Berlin, Heidelberg, 2003: Springer Berlin Heidelberg, in Cryptographic Hardware and Embedded Systems - CHES 2002, pp. 13-28.

[8] N.-T. Do, P.-C. Le, V.-P. Hoang, S. Doan, G. Nguyen, and C.-K. Pham, MO-DLSCA: Deep Learning Based Non-profiled Side Channel Analysis Using Multi-output Neural Networks. 2022, pp. 245-250.

[9] P. N. Fahn and P. K. Pearson, "IPA: A New Class of Power Attacks," Berlin, Heidelberg, 1999: Springer Berlin Heidelberg, in Cryptographic Hardware and Embedded Systems, pp. 173-186.

[10] C. O. Flynn and Z. D. Chen, "Side channel power analysis of an AES256 bootloader," in 2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE), 3-6 May 2015 2015, pp. 750-755, doi: 10.1109/CCECE.2015.7129369.

[11] P. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," Berlin, Heidelberg, 1999: Springer Berlin Heidelberg, in Advances in Cryptology — CRYPTO' 99, pp. 388-397.

[12] P. C. Kocher, "Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems," Berlin, Heidelberg, 1996: Springer Berlin Heidelberg, in Advances in Cryptology — CRYPTO '96, pp. 104-113.

[13] D. Kwon, S. Hong, and H. Kim, "Optimizing Implementations of NonProfiled Deep Learning-Based Side-Channel Attacks," IEEE Access, vol. 10, pp. 5957-5967, 2022, doi: 10.1109/ACCESS.2022.3140446.

[14] L. Li and Y. Ou, "A deep learning-based side channel attack model for different block ciphers," J. Comput. Sci., vol. 72, p. 102078, 2023.

[15] H. Maghrebi, T. Portigliatti, and E. Prouff, "Breaking Cryptographic Implementations Using Deep Learning Techniques," IACR Cryptol. ePrint Arch., vol. 2016, p. 921, 2016.

[16] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," ArXiv, vol. abs/1511.08458, 2015.

[17] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2010, doi: 10.1109/TKDE.2009.191.

[18] E. Prouff and M. Rivain, "Masking against Side-Channel Attacks: A Formal Security Proof," Berlin, Heidelberg, 2013: Springer Berlin Heidelberg, in Advances in Cryptology – EUROCRYPT 2013, pp. 142- 159.

[19] J.-J. Quisquater and D. Samyde, "ElectroMagnetic Analysis (EMA): Measures and Counter-measures for Smart Cards," Berlin, Heidelberg, 2001: Springer Berlin Heidelberg, in Smart Card Programming and Security, pp. 200-210.

[20] W. Schindler, K. Lemke, and C. Paar, "A Stochastic Model for Differential Side Channel Cryptanalysis," Berlin, Heidelberg, 2005: Springer Berlin Heidelberg, in Cryptographic Hardware and Embedded Systems – CHES 2005, pp. 30-46.

[21] D. Thapar, M. Alam, and D. Mukhopadhyay, "Deep Learning assisted Cross-Family Profiled Side-Channel Attacks using Transfer Learning," in 2021 22nd International Symposium on Quality Electronic Design (ISQED), 7-9 April 2021 2021, pp. 178-185, doi: 10.1109/ISQED51717.2021.9424254.

[22] S. Tillich, C. Herbst, and S. Mangard, "Protecting AES Software Implementations on 32-Bit Processors Against Power Analysis," Berlin, Heidelberg, 2007: Springer Berlin Heidelberg, in Applied Cryptography and Network Security, pp. 141-157.

[23] B. Timon, "Non-Profiled Deep Learning-based Side-Channel attacks with Sensitivity Analysis," IACR Transactions on Cryptographic Hardware and Embedded Systems, vol. 2019, no. 2, pp. 107-131, 02/28 2019, doi:10.13154/tches.v2019.i2.107-131.

[24] Lê Huy Thanh, Nguyễn Hữu Minh, Đỗ Ngọc Tuấn và Hoàng Văn Phúc, (2024), "Nâng cao hiệu quả phân tích kênh bên không lập mẫu cho thuật toán AES-128 sử dụng kỹ thuật học máy chuyển giao," Kỷ yếu Hội nghị Quốc gia lần thứ XXVII về Điện tử, Truyền thông và Công nghệ Thông tin (REV-ECIT 2024), tr. 351-356.

12% Tính tương đồng nói chung

Tổng cộng của tất cả các kết quả trùng khớp, bao gồm cả các nguồn trùng lặp, cho mỗi c...

Đã lọc khỏi Báo cáo

- Mục lục tham khảo
- Văn bản được trích dẫn

Nguồn hàng đầu

10% Nguồn Internet
5% Ấn bản
2% Bài tập được nộp (bài của học sinh)

Học Viên

Nguyễn Anh Thư

Người hướng dẫn khoa học

TS. Vũ Thị Thúy Hà

BỘ KHOA HỌC VÀ CÔNG NGHỆ
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

BIÊN BẢN
HỌP HỘI ĐỒNG CHẤM ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

Căn cứ quyết định số Quyết định số 3299/QĐ-HV ngày 26 tháng 12 năm 2025 của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ. Hội đồng đã họp vào hồi .9h..... giờ ...15.. phút, ngày 24 tháng 01 năm 2025 tại Học viện Công nghệ Bưu chính Viễn thông để chấm đề án tốt nghiệp thạc sĩ cho:

Học viên: **Nguyễn Anh Thư**

Tên đề án tốt nghiệp: **Nghiên cứu phương pháp tấn công kênh bên không lập mẫu thuật toán AES-128 trên thiết bị IoT**

Ngành: **Kỹ thuật viễn thông** Mã số: 8.52.02.08

Các thành viên của Hội đồng chấm đề án tốt nghiệp có mặt: ...05../ 05

TT	HỌ VÀ TÊN	TRÁCH NHIỆM TRONG HĐ	GHI CHÚ
1	**PGS.TS. Đặng Hoài Bắc**	Chủ tịch	
2	**TS. Phạm Anh Thư**	Thư ký	
3	**PGS.TS. Lê Nhật Thăng**	Phản biện 1	
4	**PGS.TS. Phạm Thanh Hiệp**	Phản biện 2	
5	**PGS.TS. Nguyễn Tiến Hòa**	Uỷ viên	

Các nội dung thực hiện:

1. Chủ tịch Hội đồng điều khiển buổi họp. Công bố quyết định của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ.
2. Người hướng dẫn khoa học hoặc thư ký đọc lý lịch khoa học và các điều kiện bảo vệ đề án tốt nghiệp của học viên (Có bản lý lịch khoa học và kết quả các môn học cao học của học viên kèm theo).
3. Học viên trình bày tóm tắt đề án tốt nghiệp.
4. Phản biện 1 đọc nhận xét (có văn bản kèm theo)
5. Phản biện 2 đọc nhận xét (có văn bản kèm theo)
6. Các câu hỏi của thành viên Hội đồng:

.....Vì sao các thiết bị IoT lại sử dụng bộ mật mã.....
.....AES-128 ?.....
.....
.....Cơ sở khoa học của việc lựa chọn mã hiệu và nội.....
.....dung của phương pháp CPA, DPA có thuật sử.....
.....phù hợp với phần cứng không ?.....

7. Trả lời của học viên:

Học viên lựa chọn AES-128 chỉ là một lựa chọn, phụ thuộc vào thiết bị IoT giải mã học viên lựa chọn

Việc lựa chọn như vậy là phù hợp phần cứng và chức năng của thiết bị IoT đáp ứng cứu

8. Thư ký đọc nhận xét về quá trình thực hiện đề án tốt nghiệp của học viên *(có văn bản kèm theo).*

9. Hội đồng họp riêng:

- Bầu Ban kiểm phiếu:
  1. Trưởng Ban kiểm phiếu: PGS.TS. Lê Nhật Thăng
  2. Uỷ viên Ban kiểm phiếu: TS. Phạm Anh Thư
  3. Uỷ viên Ban kiểm phiếu: PGS.TS. Phạm Thanh Hiệp
- Hội đồng chấm đề án tốt nghiệp bằng bỏ phiếu kín.
- Ban kiểm phiếu làm việc;
- Trưởng Ban kiểm phiếu báo cáo kết quả kiểm phiếu *(có Biên bản họp Ban kiểm phiếu kèm theo)*
- **Điểm trung bình của đề án tốt nghiệp:** 8.7

Kết luận:

1. Các nội dung cần chỉnh sửa, hoàn thiện sau bảo vệ đề án tốt nghiệp:

Chỉ sửa theo góp ý của hai phản biện và hội đồng

Rà soát các lỗi chính tả, đánh máy

2. Đề nghị Học viện công nhận (hoặc không) và cấp bằng (hoặc không) thạc sĩ cho học viên: Đề nghị HV công nhận và cấp bằng Thạc sĩ cho học viên

3. Đề án tốt nghiệp có thể phát triển thành đề tài nghiên cứu cho NCS. Nguyễn Anh Thư, có.

Buổi làm việc kết thúc vào 9h cùng ngày.

Chủ tịch	**Thư ký**
PGS.TS. Đặng Hoài Bắc	**TS. Phạm Anh Thư**

BỘ KHOA HỌC VÀ CÔNG NGHỆ
HỌC VIỆN CÔNG NGHỆ
BƯU CHÍNH VIỄN THÔNG

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

(Dùng cho người phản biện)

Tên đề tài đề án: ***Nghiên cứu phương pháp tấn công kênh bên không lập mẫu thuật toán AES-128 trên thiết bị IoT***

Ngành: Kỹ thuật viễn thông

Mã số: 8.52.02.08

Tên học viên: **Nguyễn Anh Thư**

Họ và tên người nhận xét: **PGS.TS. Lê Nhật Thăng**

Học hàm, học vị: Phó Giáo sư, Tiến sĩ

Chuyên ngành: Điện tử - Viễn thông

Cơ quan công tác: Học viện Công nghệ Bưu chính Viễn thông

NỘI DUNG NHẬN XÉT

I/ Cơ sở khoa học và thực tiễn, tính cấp thiết của đề tài:

Hiện nay, IoT sử dụng thuật toán mã hóa AES-128 được ứng dụng phổ biến trong nhiều lĩnh vực khác nhau như nhà thông minh, đô thị thông minh, trong y tế, nông nghiệp, công nghiệp... do khả năng mã hóa mạnh, hiệu năng cao, phù hợp với các thiết bị có tài nguyên hạn chế. Tuy nhiên, do các thiết bị IoT có tài nguyên 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. Ngoài ra, trong số các kỹ thuật tấn công kênh bên 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 các thiết bị tương đương. Do vậy, việc nghiên cứu, 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.

Cho nên, đề tài *"Nghiên cứu phương pháp tấn công kênh bên không lập mẫu thuật toán AES-128 trên thiết bị IoT"*, do học viên Nguyễn Anh Thư thực hiện là một đề tài có tính thời sự, khoa học và tính thực tiễn, phù hợp với tầm cỡ của một đề án tốt nghiệp thạc sĩ.

II/ Về nội dung, chất lượng của đề án, các kết quả đã đạt được (so với đề cương đã được duyệt):

Đề án tốt nghiệp thạc sĩ của học viên Nguyễn Anh Thư tập trung nghiên cứu tìm hiểu về tấn công kênh bên (SCA), một số phương pháp SCA không lập mẫu điển hình, tiến hành mô phỏng, đánh giá hiệu quả của phương pháp tấn công bên không lập mẫu vào thuật toán AES-128 trên thiết bị IoT và đưa ra các đánh giá, so sánh giữa phương pháp tấn công truyền thống CPA và phương pháp không lập mẫu dựa trên học học sâu (DDLA). Đề án được trình bày rõ ràng qua 54 trang viết A4 với tổ chức thành các chương chính, cụ thể như sau:

Mở đầu

Chương 1: Tổng quan về tấn công kênh bên (SCA).

Chương 2: Một số phương pháp SCA không lập mẫu điển hình.

Chương 3: Mô phỏng, đánh giá hiệu quả kỹ thuật SCA không lập mẫu vào thuật toán AES-128 trên thiết bị IoT.

Kết luận.

Theo đánh giá của người phản biện, nội dung nghiên cứu của đề án phù hợp với tên đề tài và phù hợp với ngành Kỹ thuật viễn thông - mã số đào tạo 8.52.02.08 và về cơ bản đạt được các kết quả theo như đề cương đã được duyệt.

III/ Những vấn đề cần giải thích thêm:

+ Đề án được trình bày rõ ràng, chi tiết. Tuy nhiên, đề án cần hiệu chỉnh, rút gọn Kết luận Chương 3 theo hướng tổng quát, ngắn gọn, súc tích hơn.

+ Làm rõ mô hình kẻ tấn công (attacker model), mức độ truy cập vật lý vào thiết bị IoT và các giả định về môi trường đo (nhiễu, điều kiện nguồn, đồng bộ thời gian).

+ Cần phân tích sâu hơn cơ sở lựa chọn mô hình Hamming Weight/Hamming Distance và mức độ phù hợp của chúng với đặc điểm phần cứng cụ thể của thiết bị mục tiêu.

+ Làm rõ mức độ đe dọa thực tế của các kết quả thu được đối với hệ thống IoT đang triển khai AES-128 trong thực tế.

IV/ Kết luận:

Đề án của học viên Nguyễn Anh Thư có chất lượng chuyên môn đáp ứng được các yêu cầu về nội dung và hình thức của một đề án tốt nghiệp thạc sĩ theo như quy định của Bộ Giáo dục và Đào tạo.

Tôi đồng ý để học viên Nguyễn Anh Thư được bảo vệ đề án trước Hội đồng đánh giá đề án tốt nghiệp thạc sĩ.

Hà Nội, ngày tháng 01 năm 2026

NGƯỜI NHẬN XÉT

PGS.TS. Lê Nhật Thăng

Câu hỏi:

Câu 1: Đề án lựa chọn hướng tấn công kênh bên không lập mẫu; vậy trong bối cảnh thiết bị IoT thực tế, những ưu điểm nổi bật và các hạn chế chính của hướng tiếp cận này so với tấn công lập mẫu là gì?

Câu 2: Cơ sở khoa học của việc lựa chọn mô hình rò rỉ công suất (Hamming Weight/Hamming Distance) và phương pháp CPA, DDLA có thực sự phù hợp với kiến trúc phần cứng và điều kiện đo của thiết bị IoT được nghiên cứu hay không

Câu 3: Từ các kết quả thực nghiệm, mức độ đe dọa thực tế của các tấn công kênh bên không lập mẫu đối với việc triển khai AES-128 trên thiết bị IoT là đến đâu, và những khuyến nghị phòng vệ quan trọng nhất là gì?

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

(Dùng cho người phản biện)

Tên đề tài đề án tốt nghiệp: 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 chuyên ngành: 8.52.02.08

Họ và tên học viên: Nguyễn Anh Thư

Họ và tên người nhận xét: Phạm Thanh Hiệp

Học hàm, học vị: PGS. TS

Chuyên ngành: Kỹ thuật viễn thông

Cơ quan công tác: Khoa Vô tuyến điện tử, Học viện Kỹ thuật quân sự

Số điện thoại: 0982.535.203 E-mail: thanhhiep@lqdtu.edu.vn

NỘI DUNG NHẬN XÉT

I/ Cơ sở khoa học và thực tiễn, tính cấp thiết của đề tài:

1. **Cơ sở khoa học:**

o Tác giả đã hệ thống hóa được các kiến thức nền tảng phức tạp: Từ mật mã học (AES-128), vật lý bán dẫn (tiêu thụ năng lượng CMOS) đến các kỹ thuật phân tích thống kê (DPA, CPA) và học sâu (Deep Learning).

o Việc so sánh giữa CPA (phương pháp thống kê truyền thống) và DDLA (phương pháp học sâu) dựa trên khả năng chống lại các biện pháp bảo vệ (như làm lệch xung/desynchronization) là một cách tiếp cận khoa học bài bản và có giá trị chứng minh cao.

2. **Tính cấp thiết:**

o An ninh cho thiết bị IoT (Internet of Things) đang là một chủ đề cực kỳ nóng hổi. Việc chuyển hướng nghiên cứu từ các lỗ hổng thuật toán toán học sang các lỗ hổng vật lý (kênh bên - Side Channel Attack) là rất sát với thực tế triển khai, khi mà các thiết bị IoT thường có lớp bảo vệ vật lý yếu hơn so với các thiết bị bảo mật chuyên dụng.

o Đề tài tập trung vào "Tấn công không lập mẫu" (Non-profiled SCA). Đây là hướng đi rất đúng đắn vì trong thực tế, kẻ tấn công thường không có quyền truy cập vào một thiết bị bản sao hoàn hảo để huấn luyện (profiling) trước. Do đó, mô hình này phản ánh đúng mối đe dọa thực tế đối với hạ tầng IoT.

II/ Nội dung của đề án tốt nghiệp, các kết quả đã đạt được:

1. **Cấu trúc và nội dung:**

o Bố cục luận văn chặt chẽ, đi từ tổng quan lý thuyết (Chương 1, 2) đến thực nghiệm mô phỏng (Chương 3).

- Tác giả đã xây dựng được một quy trình tấn công hoàn chỉnh (Pipeline): Từ thu thập dữ liệu (hoặc sử dụng dataset chuẩn) tới Tiền xử lý (Desync) và Tấn công bằng CPA/DDL. Cuối cùng là đánh giá hiệu năng.
- Việc lựa chọn ChipWhisperer làm nền tảng tham chiếu là phù hợp với chuẩn nghiên cứu chung của cộng đồng SCA thế giới.

2. **Kết quả đạt được:**

- Đã thực hiện thành công việc tấn công khôi phục khóa AES-128 bằng cả hai phương pháp CPA và DDLA trên tập dữ liệu mô phỏng.
- Đã chứng minh được sự ưu việt của mô hình Học sâu (DDLA) so với CPA trong môi trường có nhiễu và mất đồng bộ. Cụ thể, tác giả đã chỉ ra CPA thất bại khi trace bị lệch (desync), trong khi DDLA vẫn khôi phục được khóa với độ lệch 20 mẫu.
- Đã xác định được "ngưỡng thất bại" của mô hình đề xuất (thất bại khi độ lệch đạt 50 mẫu), điều này cho thấy sự trung thực và nghiêm túc trong nghiên cứu khoa học (biết rõ giới hạn của giải pháp).

III/ Những vấn đề cần giải thích thêm:

1. **Về nguồn dữ liệu thực nghiệm:**

- Trong mục 3.1.2, đề án nêu dữ liệu được thu từ mạch CW1173, tuy nhiên chưa chỉ rõ đây là dữ liệu tự thiết lập phần cứng và đo đạc trực tiếp, hay sử dụng bộ dữ liệu mẫu (public dataset) có sẵn của ChipWhisperer/ASCAD. Nếu là tự đo, cần mô tả kỹ hơn về cấu hình đo (tốc độ lấy mẫu, vị trí đặt đầu dò trên mạch). Nếu là dataset có sẵn, cần trích dẫn rõ nguồn gốc để đảm bảo tính minh bạch khoa học.

2. **Về kiến trúc mạng Nơ-ron (CNN):**

- Việc lựa chọn kiến trúc mạng CNN (Convolutional Neural Network) để giải quyết vấn đề mất đồng bộ (desynchronization) thay vì mạng MLP (Multi-Layer Perceptron) thuần túy cần được làm rõ.

3. **Về giới hạn của mô hình (Desync = 50):**

- Bổ sung biện luận cho lý do mô hình DDLA lại thất bại khi độ lệch (trace shift) đạt mức 50 mẫu và những giải pháp có thể để khắc phục giới hạn này.

4. **Về kịch bản tấn công thực tế trên IoT:**

- Đề tài thực hiện tấn công với giả định thu thập được 5.000 vệt công suất (traces) với cùng một khóa cố định. Trong thực tế triển khai IoT, các giao thức bảo mật thường xuyên thay đổi khóa phiên (Session Key) hoặc sử dụng khóa dùng một lần. Cần biện luận trường hợp này hoặc giới hạn phạm vi đề án.

IV/ Kết luận:

Đề án có hàm lượng khoa học tốt, phương pháp nghiên cứu hiện đại và kết quả thực nghiệm rõ ràng. Tác giả nắm vững kiến thức về cả Mật mã và Học sâu. Đề án đã hoàn thành các nội dung, đảm bảo tính khoa học và chất lượng, đồng ý cho phép học viên bảo vệ đề án tốt nghiệp.

Ngày 19 tháng 01 năm 2026
NGƯỜI NHẬN XÉT

PGS. TS Phạm Thanh Hiệp