TỔNG QUAN VỀ ĐÁNH GIÁ BẢO MẬT PHẦN CỨNG SỬ DỤNG PHÂN TÍCH ĐIỆN NĂNG TIÊU THỤ

Trần Lương Hùng[1], Võ Quang Dũng[1], Nguyễn Hữu Đức[1], Đỗ Ngọc Tuấn[2*]
[1]Trường Đại học Thông tin liên lạc, [2*]Đại học Kỹ thuật Lê Quý Đôn

Thông tin chung:
Ngày nhận bài:28/04/2023
Ngày phản biện: 03/05/2023
Ngày duyệt đăng: 11/05/2023

Title: *Overview of hardware security evaluation using power consumption analysis.*

Từ khóa: *Bảo mật phần cứng, phân tích kênh bên, phân tích công suất.*

Keywords: *Hardware security, side channel analysis, power analysis.*

ABSTRACT: *Along with the development of the industrial revolution 4.0, the issue of security and safety within the electronic devices including hardware, firmware and software, is increasingly crucial. In which, hardware security is emerging as a major concern in the field of cyber security and safety. More and more important information is transmitted and stored in electronic devices. They are using widely in many aspects as traffics, medicals, telecommunication, entertainment and even in military. Therefore, hardware security becomes a real challenge. This paper presents the overview of hardware security evaluation of cryptographic devices based on power consumption analysis. Especially, some advanced analytical technique using deep neural network will be analyzed and assessed.*

TÓM TẮT: *Cùng với sự phát triển mạnh mẽ của cuộc cách mạng công nghiệp lần thứ 4, vấn đề an ninh, an toàn trong các thiết bị điện tử bao gồm cả phần cứng, firmware và phần mềm đang trở nên cấp bách. Trong đó, an ninh phần cứng nổi lên như một sự lo ngại chủ yếu trong lĩnh vực bảo mật và an toàn thông tin. Ngày càng nhiều thông tin quan trọng được truyền tải, lưu trữ trong thiết bị điện tử. Chúng được sử dụng rộng rãi trong nhiều lĩnh vực từ giao thông, y tế, thông tin liên lạc, giải trí, cho đến các thiết bị quân sự hiện đại. Chính vì thế, bảo mật phần cứng trở thành một thách thức thực sự. Bài báo này trình bày tổng quan về đánh giá an toàn thiết bị mật mã dựa trên phân tích điện năng tiêu thụ. Đặc biệt, các kĩ thuật phân tích tiên tiến hiện nay sử dụng mạng học sâu sẽ được phân tích và đánh giá.*

1. Mở đầu

Theo tiêu chuẩn quốc gia TCVN 12212:2018, việc đánh giá an toàn phần cứng cho một thiết bị điện tử thực chất là quá trình thực hiện các phương pháp tấn công để xác định mức độ an toàn của thiết bị trước các nguy cơ tấn công phần cứng. Vì vậy, nghiên cứu các phương pháp tấn công, đặc biệt với các phương pháp có hiệu quả về mặt thời gian là rất ý nghĩa trong việc đánh giá bảo mật. Các vi mạch số luôn tiêu hao năng lượng trong quá trình xử lý tính toán. Chúng tiêu thụ dòng điện từ nguồn cung cấp sau đó chuyển hóa thành nhiệt năng. Vì vậy, điện năng tiêu thụ của một vi mạch số là một vấn đề rất quan trọng. Chúng sẽ quyết định đến việc khi nào mạch cần làm mát hay không, chúng cũng quyết định đến loại nguồn nào cần sử dụng, và trong trường hợp của các bộ xử lý mã hóa/giải mã, chúng quyết định đến thiết bị có thể bị tấn công hay không.

Được đề xuất đầu tiên bởi Kocher năm 1999 [1], tấn công kênh kề (SCA) là một dạng tấn công sử dụng thông tin không mong muốn bị rò rỉ trong quá trình thực thi các thuật toán bên trong thiết bị điện tử. Trong đó, điện năng tiêu thụ là một trong những dữ liệu kênh bên được khai thác nhiều nhất và được thực thi thành công trên nhiều thuật toán từ mã khối như DES [1,2], AES [3,4] cho đến mã công khai RSA, ECC. Đặc biệt, một số thiết bị điện tử trên thực tế đã bị tấn công thành công như chip FPGA virtex-II

[5], SIM trong mạng GSM [6], hay thiết bị xác thực hai yếu tố Yubikey [7].

Trên thực tế việc nâng cao hiệu quả của tấn công SCA có thể giúp làm giảm thời gian đánh giá độ an toàn của thiết bị luôn là một trong những thách thức lớn. Vì vậy, việc nghiên cứu, cập nhật và hiểu các phương pháp tấn công phân tích điện năng tiêu thụ là cần thiết. Bài báo này cung cấp các nội dung tổng quan về đánh giá an toàn phần cứng dựa trên điện năng tiêu thụ. Trong đó một số phương pháp tấn công điển hình hiện nay sẽ được trình bày. Nội dung còn lại của bài báo được tổ chức như sau. Phần II trình bày các nội dung về điện năng tiêu thụ, mô hình điện năng tiêu thụ. Các phương pháp tấn công sử dụng điện năng tiêu thụ sẽ được trình bày ở phần III. Bên cạnh đó, ưu nhược điểm của các phương pháp tấn công và phạm vi ứng dụng cũng được trình bày trong phần này. Kết luận của bài báo được trình bày trong phần IV.

2. Phương pháp nghiên cứu

Phương pháp phân tích và tổng hợp lý thuyết được sử dụng để phân tích tổng quan về đánh giá bảo mật phần cứng sử dụng phân tích điện năng tiêu thụ

Phương pháp thống kê, so sánh để đánh giá kết quả thực nghiệm của các phương pháp đề xuất so với các công trình khác đã công bố.

3. Kết quả nghiên cứu và bàn luận

3.1. Điện năng tiêu thụ của mạch CMOS

Như được trình bày trong [8], điện năng tiêu thụ của mạch CMOS được tính bằng tổng điện năng tiêu thụ của các phần tử logic cấu thành lên vi mạch CMOS đó. Vì vậy, điện năng tiêu thụ của mạch phụ thuộc vào số lượng phần tử logic, các kết nối giữa chúng và các kĩ thuật tổng hợp lên phần tử logic. Khi một mạch logic CMOS hoạt động, nguồn điện cung cấp cho mạch là nguồn cố định một chiều. Lúc này các phần tử logic (cell) sẽ xử lý các tín hiệu đầu vào và tiêu thụ một phần dòng điện được cung cấp. Giả sử i_{DD} là dòng điện tức thời, p_{cir} là điện năng tiêu thụ tức thời của mạch CMOS. Ta có thể dễ dàng tìm được điện năng tiêu thụ trung bình của mạch trong thời gian T như sau:

$$P_{\text{cir}} = \frac{1}{T}\int_0^T p_{\text{cir}}(t)\,dt = \frac{V_{\text{DD}}}{T}\int_0^T i_{\text{DD}}(t)\,dt \quad (1)$$

$$P_{stat} = I_{leak}.V_{\text{DD}} \quad (2)$$

Trên thực tế, điện năng tiêu thụ của phần tử logic CMOS là tổng của điện năng tiêu thụ tĩnh P_{stat} và điện năng tiêu thụ động P_{dyn}. Điện năng tiêu thụ tĩnh được tính bằng công thức (2), trong đó I_{leak} là dòng rò của một cell. Thông thường, P_{stat} của mạch logic thường rất nhỏ, P_{stat} sẽ tăng đáng kể đối với các cấu trúc vi mạch hiện đại, khi kích thước của chúng rất nhỏ.

Đối với P_{dyn}, điện năng tiêu thụ động thường xảy ra trong quá trình phần tử logic có sự biến đổi từ đầu vào đến đầu ra, ví dụ: chuyển logic từ 0 sang 1 hay 1 về 0. Trong trường hợp này, có thể coi điện năng tiêu thụ động của cell CMOS xảy ra do điện năng tiêu thụ dòng điện nạp và điện năng tiêu thụ dòng ngắn mạch P_{sc} . P_{chrg} và P_{sc} có thể được xác định lần lượt qua công thức (3) và (4).

$$P_{chrg} = \frac{1}{T}\int_0^T p_{chrg}(t)\,dt = \alpha \cdot f \cdot C_L \cdot V_{\text{DD}}^2 \quad (3)$$

$$P_{sc} = \frac{1}{T}\int_0^T p_{sc}(t)\,dt = \alpha \cdot f \cdot I_{peak} \cdot t_{sc} \quad (4)$$

Trong đó, α là hệ số chuyển (nếu đầu vào và đầu ra chuyển mạch mỗi chu kì đồng hồ thì α =1), f là tần số hoạt động của mạch, C_L là thành phần dung kháng đầu ra của mạch, I_{peak} và t_{sc} kí hiệu cho dòng ngắn mạch và thời gian xảy ra ngắn mạch.

Từ những phân tích trên, dễ dàng nhận thấy điện năng tiêu thụ động phụ thuộc vào dữ liệu mà mạch CMOS xử lý, điện năng tiêu thụ của mạch bằng tổng P_{dyn} và P_{stat}. Các trường hợp đầu vào và đầu ra không đổi, có thể coi chỉ tồn tại điện năng tiêu thụ tĩnh.

3.2. Mô hình điện năng tiêu thụ

Mô hình điện năng tiêu thụ là mô hình có thể diễn tả được sự liên quan của quá trình chuyển trạng thái tại đầu ra của một phần tử logic với điện năng tiêu thụ thực tế. Những mô hình này thường được đi kèm với các thư viện chuẩn trong phần mềm thiết kế chuyên dụng. Tuy nhiên, cũng có các mô hình điện năng tiêu thụ khác được sử dụng để so sánh sự liên quan giữa các sự kiện chuyển trạng thái logic với các vệt điện năng (power trace) như Hamming Weight (HW) và Hamming Distance (HD).

Mô hình HD: đây là mô hình biểu diễn số lần chuyển trạng thái từ 0 sang 1 hoặc từ 1 sang 0 trong một khoảng thời gian nhất định. Số lượng chuyển trạng thái này sau đó được sử dụng để miêu tả điện năng tiêu thụ trong khoảng thời gian trên. Khi sử dụng mô hình HD để mô tả điện năng tiêu thụ, một số giả thiết sau được sử dụng. Thứ nhất là năng lượng tiêu thụ của các cell là như nhau đối với trạng thái chuyển $0 \rightarrow 1$ và $1 \rightarrow 0$, các trường hợp $0 \rightarrow 0$ và $1 \rightarrow 1$ cũng tiêu thụ năng lượng như nhau nhưng nhỏ hơn nhiều so với hai trường hợp trước. Mô hình HD vì thế bỏ qua các vấn đề như sự khác nhau của dung kháng kí sinh của các đường kết nối cũng như của các cell. Mô hình này giả thiết tất cả các cell tiêu tốn lượng điện năng là như nhau và có thể bỏ qua phần điện năng tiêu thụ tĩnh. Dựa theo định nghĩa, HD của hai biến v_0 và v_1 được tính như sau:

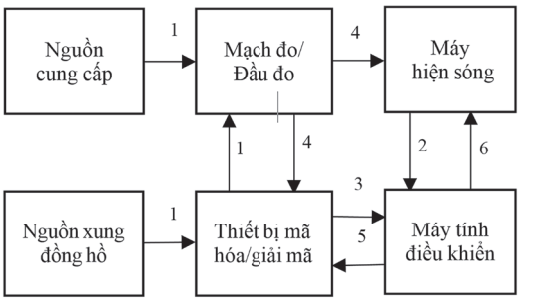
$$HD(v_0, v_1) = HW(v_0 \oplus v_1) \quad (5)$$

Trong đó kí hiệu của Hamming Weight, dùng cho việc xác định số bit 1 trong phép tính $(v_0 \oplus v_1)$.

Mô hình HW: Mô hình HW đơn giản hơn HD. Mô hình này thường được sử dụng khi mô hình HD không thể xác định trong một số trường hợp. Ví dụ, người sử dụng không biết về cấu trúc mạng của mạch, hoặc trong trường hợp không thể xác định được hai giá trị v_0 và v_1 liên tiếp nhau. Một trường hợp khác khi mô hình HD không thể áp dụng đó là người sử dụng chỉ xác định được một giá trị v_0, lúc này mô hình HW có thể được sử dụng. Trong trường hợp này, giả thiết về sự tương đương điện năng tiêu thụ giữa $1 \rightarrow 0$ và $0 \rightarrow 1$ sẽ không còn được sử dụng. Trên thực tế, quá trình chuyển trạng thái từ $0 \rightarrow 1$ thường tiêu hao năng lượng nhiều hơn so với chuyển từ $1 \rightarrow 0$. Chính vì vậy, trung bình điện năng tiêu thụ của dữ liệu có trọng số (số bít 1) nhiều hơn sẽ lớn hơn dữ liệu có trọng số nhỏ. Mô hình HW sẽ xác định số bit 1 để thể hiện sự tương quan với điện năng tiêu thụ thực tế.

3.3. Thiết lập đo điện năng tiêu thụ

Để thực hiện tấn công phân tích điện năng tiêu thụ, việc đầu tiên là thiết lập đo và thu thập vệt điện năng của bộ xử lý. Thông thường, thiết lập đo vệt điện năng có thể được thực hiện như minh họa trong Hình 1.



Hình 1: *Qui trình thu thập dữ liệu điện năng tiêu thụ [8]*

Chức năng của các thành phần như sau:

Nguồn cung cấp: Cung cấp nguồn một chiều cho thiết bị mã hóa như bộ vi xử lý hoặc lõi mã hóa chuyên dụng.

Mạch đo/đầu dò: Điện năng tiêu thụ của mạch mã hóa có thể được đo trực tiếp bằng việc gắn mạch đo (thường có thể là một điện trở) được gắn giữa nguồn cung cấp và mạch mã hóa. Ngoài ra, có thể đo gián tiếp bằng đầu dò EM.

Oscilloscope: Thiết bị đo chuyển đổi tín hiệu đo điện năng tiêu thụ từ mạch đo/ đầu dò thành tín hiệu số. Thiết bị đo này có khả năng giao tiếp với máy tính để nhận lệnh điều khiển, truyền dữ liệu đã chuyển đổi đến máy tính.

Máy tính điều khiển: Có chức năng giao tiếp với cả thiết bị mã hóa và máy

oscilloscope. Đối với thiết bị mã hóa, máy tính điều khiển thực hiện truyền bản tin cần mã hóa (plaintext) và nhận về bản tin đã mã hóa (ciphertext). Đối với oscilloscope, máy tính thực hiện điều khiển chế độ đo đối với oscilloscope, đồng thời nhận dữ liệu đo từ oscillosope truyền về.

Thiết bị mật mã: Là thiết bị cần thực hiện đánh giá bảo mật (tấn công). Trong trường hợp này, thiết bị mã hóa/giải mã có thể là một vi xử lý (ARM, AVR) thực thi thuật toán mã hóa (phần mềm) hoặc có thể thiết bị mã hóa cứng trên FPGA hoặc ASIC.

Nguồn xung đồng hồ: Là thành phần quan trọng, cung cấp tín hiệu xung đồng hồ cho thiết bị mã hóa/giải mã hoạt động.

3.4. Đánh giá an toàn phần cứng dựa trên phân tích điện năng tiêu thụ

3.4.1. Phân loại

Được giới thiệu lần đầu bởi Kocher [1], tấn công phân tích kênh bên đã cho thấy mối đe dọa thực tế đến thiết bị bảo mật khi khóa bí mật có thể dễ dàng lấy được. Tấn công kênh bên thường được chia thành hai nhóm chính đó là tấn công lập mẫu (Profile attack) và không lập mẫu (Non-profile attack).

Đối với tấn công lập mẫu, điển hình là Template attack [9], người tấn công cần thực hiện hai bước quan trọng. Thứ nhất là lập mẫu, người tấn công phải có một thiết bị có chức năng và cấu tạo giống hệt như thiết bị mục tiêu và có đầy đủ quyền thực thi trên thiết bị. Sau đó người tấn công sẽ điều khiển thiết bị mục tiêu thực thi mã hóa/giải mã. Tiếp theo một lượng lớn vệt công suất tương ứng với từng giá trị khóa dự đoán sẽ được ghi lại. Thứ hai là áp dụng mẫu đã có lên thiết bị mục tiêu. Người tấn công ghi lại một lượng nhỏ vệt công suất quá trình mạch mục tiêu thực thi thuật toán mã hóa/giải mã, so sánh các khóa dự đoán với mẫu đã lập và xác định key với mẫu nào giống nhất.

Ngược lại, với tấn công không lập mẫu, ví dụ như tấn công phân tích vi sai điện năng tiêu thụ (DPA: Differential Power Analysis) [1], hoặc tấn công phân tích tương quan điện năng (CPA: Correlation Power Analysis) [3], người tấn công chỉ cần sử dụng thiết bị mục tiêu để ghi lại vệt công suất và dùng các kỹ thuật thống kê để tìm ra khóa bí mật.

Trên phương diện của người thiết kế, các phương pháp tấn công trên là các công cụ hữu hiệu để đánh giá mức độ an toàn của thiết kế trước các nguy cơ tấn công phần cứng. Trong phần tiếp theo, một số kĩ thuật tấn công phổ biến sẽ được trình bày chi tiết.

3.4.2. Phương pháp phân tích tương quan điện năng tiêu thụ

CPA là một dạng tấn công không cần xây dựng mô hình mẫu để so sánh với mạch mục tiêu. CPA là một trong những hình thức tấn công có khả năng ứng dụng thực tế cao. Vì trong quá trình thực hiện, người tấn công chỉ cần một mạch mục tiêu với lượng vệt công suất đủ lớn. Nguyên lý chung của kỹ thuật CPA là thực hiện đo và thu thập vệt công suất của thiết bị mã hóa, sau đó là xác định sự tương quan giữa dữ liệu vừa đo với các phép tính được thực hiện bởi thiết bị mã hóa tại cùng một thời điểm. Kỹ thuật này yêu cầu cần phải có một mô hình về điện năng tiêu thụ dựa trên các toán tử hoặc phép tính mà thiết bị mã hóa thực hiện. Ví dụ, sử dụng biến đổi SubByte kết hợp với trọng số Hamming như sau:

$$h = HW\left(SubByte(d_i \oplus k)\right) \quad (6)$$

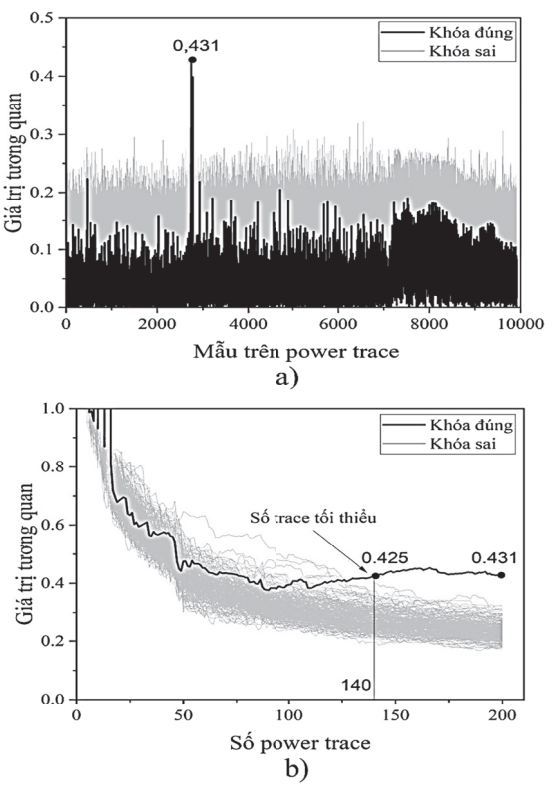
Trong đó d_i và k lần lượt là byte thứ i của bản tin đầu vào ngẫu nhiên và byte tương ứng của khóa bí mật cố định.

Như đã trình bày ở trên, HW là một trong những mô hình cơ bản nhất của mô hình tiêu thụ điện năng. Mô hình này có khả năng ước lượng khá tốt lượng điện năng tiêu thụ của thiết bị mã hóa khi mà người tấn công có thể biết chính xác về các dữ liệu đi vào và ra trong một phần thiết bị. Mô hình này coi quá trình chuyển trạng thái từ $0 \rightarrow 1$ tiêu hao năng lượng nhiều hơn so với chuyển từ $1 \rightarrow 0$. Vì vậy trong mô hình HW, điện năng tiêu thụ thực tế được cho là có mối quan hệ tuyến tính với số lượng bit 1 tại thời điểm tính toán dữ liệu. Để có thể phân tích và lấy được khóa bí

mật, người tấn công sẽ dựa trên mối quan hệ giữa vệt công suất thu được và mô hình điện năng tiêu thụ. Một phương pháp hiệu quả để tính được sự tương quan này đó chính là sử dụng hệ số tương quan Pearson. Đây là một trong những hệ số được sử dụng rộng rãi cho việc tính toán mối quan hệ tuyến tính giữa các dữ liệu. Trong lĩnh vực đánh giá bảo mật, đây là một kỹ thuật phân tích thống kê hiệu quả để thực hiện phân tích mã khóa bí mật.

Giả sử thực hiện mã hóa n bản tin (Plaintext) với cùng một khóa (Key), ta thu được n vệt công suất, mỗi vệt công suất bao gồm l sample. Kí hiệu ($1 \le i \le l, 1 \le j \le n$) là giá trị sample thứ i của trace thứ j. Hệ số Pearson biểu diễn mối quan hệ giữa điện năng tiêu thụ thực tế và mô hình điện năng tiêu thụ như sau:

$$r_{k,i} = \frac{\sum_{j=1}^{n}(h_{j,k} - \overline{h}_k)(t_{j,i} - \overline{t}_i)}{\sqrt{\sum_{j=1}^{n}(h_{j,k} - \overline{h}_k)^2 \sum_{j=1}^{n}(t_{j,i} - \overline{t}_i)^2}} \quad (7)$$



Hình 2: *Kết quả thực thi CPA lên thuật toán AES-128 trên chip XMEGA*

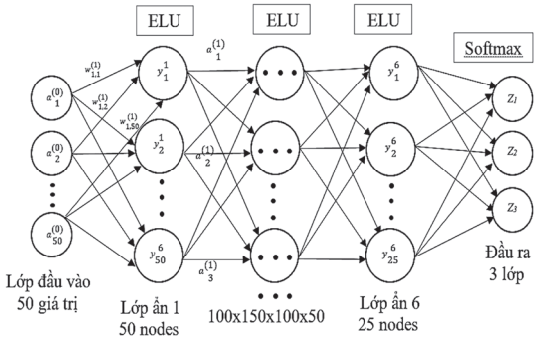
Như được mô tả trên Hình 2, kế quả của một tấn công CPA sử dụng 200 vệt công suất dựa trên công thức (7) cho thấy giá trị tương quan của khóa đúng là r = 0,43, các khóa giả thiết còn lại đều cho giá trị tương quan thấp hơn 0.3. Vì vậy, có thể thấy dựa trên hệ số tương quan giữa điện năng tiêu thụ thực tế và mô hình tiêu thụ điện năng có thể xác định chính xác khóa bí mật được xử lý trong thiết bị mã mật AES-128. Ngoài ra, cũng dựa trên CPA, người phân tích có thể xác định được số vệt công suất tối thiểu để có thể phân biệt được khóa đúng hay khoa sai như minh họa trong Hình 2.b.

Mặc dù được chứng minh hiệu quả trên nhiều thuật toán và công trình khác nhau, phương pháp CPA tồn tại một số hạn chế. Thứ nhất, CPA yêu cầu đồng bộ tốt đối với toàn bộ dữ liệu điện năng. Vì dựa trên công thức (7) có thể thấy hệ số tương quan được ước lượng dựa trên vị trí mẫu cố định của tất cả các vệt công suất. Nếu sự cố định (đồng bộ giữa các vệt) kém, thì giá trị tương quan ước lượng của khóa giả thiết đúng sẽ giảm đi đáng kể. Chính vì vậy, biện pháp để kháng tấn công CPA có thể được thực thi bằng một số biện pháp như che giá trị trung gian (dùng mặt nạ) hoặc gây bất đồng bộ (kĩ thuật ẩn) các vệt công suất. Bên cạnh phương pháp CPA truyền thống, một số kỹ thuật cải tiến của CPA đã được giới thiệu, ví dụ như công trình [10].

3.4.3. Phương pháp phân tích điện năng tiêu thụ dựa trên học máy

Được đề xuất vào năm 2019 bởi Timon [11], kĩ thuật tấn công kênh bên không lập mẫu sử dụng mạng học sâu đã cho thấy ứng dụng mạnh mẽ của trí tuệ nhân tạo trong lĩnh vực bảo mật phần cứng. Kế thừa những ưu điểm của tấn công không lập mẫu như không yêu cầu bản sao thiết bị, kiến trúc mạng đơn giản. Đặc biệt, Timon đã chứng minh kĩ thuật tấn công này có hiệu quả cao hơn hẳn với tấn công kênh bên truyền thống trong trường hợp thiết bị mật mã có trang bị những kĩ thuật chống tấn công điển hình như: kĩ thuật ẩn (bất đồng bộ thời gian),

kĩ thuật mặt nạ (bậc 1, 2) khi kĩ thuật CPA không thể thực hiện được. Hơn nữa, kĩ thuật này chỉ yêu cầu thực hiện huấn luyện mô hình và sử dụng tham số huấn luyện để xác định khóa bí mật. Chính vì vậy, nghiên cứu về kĩ thuật tấn công kênh bên không lập mẫu ứng dụng mạng học sâu đã được nhiều nhà nghiên cứu quan tâm và công bố gần đây.



Hình 3: *Kiến trúc mạng MLP sử dụng cho kỹ thuật DDLA lên thuật toán AES-128 [12]*

Về cơ bản, một mạng học sâu bao gồm có 3 lớp chính: Lớp đầu vào, lớp ẩn và lớp đầu ra. Hình 3 minh họa một ví dụ về mạng perceptron đa tầng với cấu trúc gồm có 6 lớp ẩn, đầu vào là dữ liệu có kích thước 50, đầu ra là 3.

Để thu được kết quả ở đầu ra, một quá trình thực hiện gọi là *lan truyền tiến* được thực hiện. Quá trình này được minh họa trên Hình 3, tổng trọng số của mỗi node được tính bằng công thức sau:

$$y_i^{(l)} = b_i + \sum_{j=1}^{l^{(l-1)}} a_j^{(l-1)} \times w_{ji}^{l} \qquad (8)$$

Trong đó, b_i là hệ số hiệu chỉnh của node thứ i^{th}, w_{ji}^{l} là trọng số nối giữa node i^{th} của lớp $l-1$ và node j^{th} của lớp thứ l, $a_j^{(l-1)}$ biểu diễn kết quả đầu ra của một hàm kích hoạt $F(y)$ tại node j^{th} của lớp $l-1$ và được tính như sau:

$$a_j^{(l)} = \boldsymbol{F}(y_j^{(l)}) \qquad (9)$$

Một điểm chú ý đó là công thức (8) không áp dụng cho lớp đầu vào. Giá trị đầu ra $a_j^{(l)}$ sẽ được sử dụng như đầu vào của lớp kế tiếp. Quá trình này sẽ liên tục thực hiện từ đầu vào cho đến đầu ra.

$$\text{ReLU}: F_{(y)} = \begin{Bmatrix} y : y > 0 \\ 0 : y \le 0 \end{Bmatrix} \qquad (10)$$

$$\text{ELU}: F_{(y)} = \begin{Bmatrix} y : & y > 0 \\ \alpha \cdot \left(e^{y} - 1\right) : & y \le 0 \end{Bmatrix} \qquad (11)$$

$$\text{SoftMax}: z(y)[i] = \frac{e^{y[i]}}{\sum_{j=1}^{K} e^{y[j]}} \qquad (12)$$

Trong DL, một số hàm kích hoạt thường được sử dụng đó là ELU và RELU, được tính như công thức (10) và công thức (11). Riêng lớp đầu ra của mô hình, hàm kích hoạt thường sử dụng là Softmax được tính bằng công thức (12). Đến đây, mô hình mới tính được giá trị đầu ra. Tuy nhiên, so với giá trị mong muốn ở đây là nhãn (label) thì chưa thể đạt được độ chính xác. Độ lệch này được xác định bởi một hàm mất mát tính bằng công thức sau:

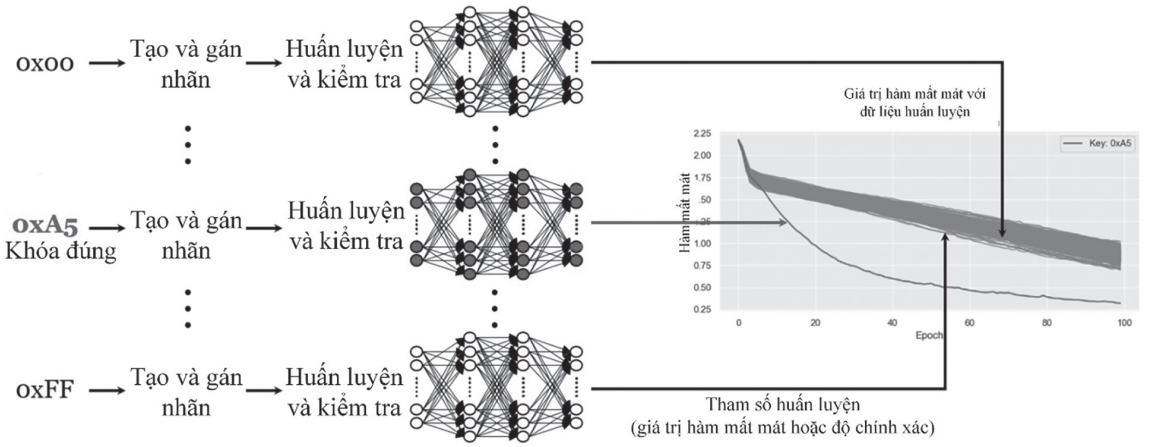
$$\mathsf{L}_X(\boldsymbol{w}) = -\sum_{j=1}^{n} y_{true} \ln(z) \qquad (13)$$

Trong đó n là số nhãn đầu ra, y_{true} là giá trị nhãn.

Vì vậy, một quá trình ngược lại có tên *lan truyền ngược* được sử dụng để tối ưu hàm mất mát, từ đó cập nhật trọng số của mô hình. Quá trình này sẽ lặp đi lặp lại cho đến khi kết quả đầu ra và nhãn đạt độ chính xác nhất định. Lúc này, mô hình sẽ hoàn tất quá trình huấn luyện.

Sau khi đã xây dựng mô hình, quá trình tấn công có thể được mô tả tóm tắt như sau: Khi khóa dự đoán là khóa chính xác được sử dụng, chuỗi các giá trị trung gian sẽ được dự đoán chính xác và do đó, các

phân vùng và các nhãn được sử dụng cho huấn luyện mạng DL sẽ khớp với các vệt điện năng tương ứng. Mặt khác, đối với tất cả các khóa dự đoán sai, các nhãn được sử dụng cho huấn luyện sẽ không phù hợp với các vệt điện năng.



Hình 4: *Quá trình thực hiện phân tích DDLA lên thuật toán AES-128*

Do đó, nếu kiến trúc mạng phù hợp với tập dữ liệu (các vệt công suất) người ta có thể quan sát quá trình huấn luyện hiệu quả hơn cho giá trị khóa chính xác hơn so với các khóa dự đoán khác. Sau đó kẻ tấn công có thể phân biệt khóa chính xác với các khóa khác bằng cách chọn khóa để tìm ra các chỉ số (Accuracy và loss) của dữ liệu huấn luyện là tốt nhất như minh họa trong Hình 4. Để bảo đảm rằng mỗi lần khóa dự đoán là độc lập nhau, điều quan trọng là phải khởi tạo lại các tham số có thể huấn luyện của mạng sau mỗi lần huấn luyện. DDLA là phương pháp tiềm năng cho đánh giá bảo mật của các thuật toán có áp dụng các biện pháp kháng tấn công. Tuy nhiên việc lặp lại quá trình huấn luyện mạng học sâu nhiều lần dẫn đến hiệu quả về chi phí của phương pháp này chưa được tối ưu. Mới đây nhất, công trình [13] đã được giới thiệu để giải quyết các nhược điểm của DDLA. Tuy nhiên, vẫn còn tồn tại những vấn đề cần nghiên cứu như độ chính xác và phương pháp xác định khóa đúng của DDLA.

4. Kết luận

Phân tích kênh bên là một công cụ mạnh mẽ để phá vỡ các thuật toán rất phổ biến hiện nay. Tuy nhiên, SCA cũng chính là biện pháp trực quan nhất để đánh giá được độ an toàn trước các nguy cơ tấn công phần cứng. Bài báo đã trình bày tổng quan về phân tích điện năng tiêu thụ, ứng dụng trong đánh giá an toàn phần cứng. Với biện pháp sử dụng thống kê toán, người đánh giá có thể thực hiện được đối với các thuật toán được thực thi không có biện pháp phòng về. Hơn nữa, phương pháp này đòi hỏi một số kĩ thuật tiền xử lý dữ liệu trước khi đưa vào thực hiện phân tích. Điều này dẫn đến thời gian đánh giá cho thiết bị lâu hơn và tốn kém. Ngược lại, với phương pháp phân tích sử dụng mạng học sâu, người đánh giá có thể bỏ qua các bước tiền xử lý. Tuy nhiên, quá trình phân tích này yêu cầu thực thi trên các thiết bị máy tính có sức mạnh tính toán lớn.

Tài liệu tham khảo

1. Kocher P, Jaffe J, Jun B, "Differential Power Analysis," *CRYPTO 1999, LNCS 1666. Springer: Heidelberg*, p. 388–397, 1999.
2. N. H. Quang, "DPA, một dạng tấn công sidechannel hiệu quả," *Tạp chí nghiên cứu Khoa học và Công nghệ Quân sự*, 2013.
3. Chari S, Rao JR, Rohatgi P, "Template Attacks," *CHES 2002,*

LNCS 2523. Springer: Heidelberg, pp. 13-28, 2002.
4. Le.T.H, Clediere, Canovas.C, Robisson.B, Serviere.C, Lacoume.J, "A Proposition for Correlation Power Analysis Enhancement," *in CHES*, 2006.
5. Josep Balasch, Benedikt Gierlichs, Roel Verdult, Lejla Batina, and Ingrid Verbauwhede, "Power analysis of atmel cryptomemory - recovering keys from secure eeproms," *in The Cryptographers' Track at the RSA Conférence, San Francisco, CA, USA*, 2012.
6. Yuanyuan Zhou, Yu Yu, François-Xavier Standaert, and Jean-Jacques Quisquater, "On the need of physical security for small embedded devices: A case study with COMP128-1 implémentations in SIM cards," *in Financial Cryptography and Data Security - 17th International Conférence, Okinawa, Japan*, 2013.
7. David Oswald, Bastian Richter, and Christof Paar, "Side-channel attacks on the yubikey 2 one-time password generator," *in Research in Attacks, Intrusions, and Defenses -16th International Symposium, RAID 2013, Rodney Bay, St. Lucia*, 2013.
8. Mangard, S., Oswald, E., Popp, T., & Stefan Mangard Elisabeth Oswald, T. P. (2007). Power Analysis Attacks: Revealing the Secrets of Smart Cards (Advances in Information Security). In *Power Analysis Attacks* (Vol. 31).
9. S. Chari, J. R. Rao, and P. Rohatgi, "Template attacks," in *CryptographicHardware and Embedded Systems - CHES 2002*, B. S. Kaliski, ¸c. K. Ko¸c, and C. Paar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003,pp. 13–28.
10. N. -T. Do, V. -P. Hoang and C. -K. Pham, "Low Complexity Correlation Power Analysis by Combining Power Trace Biasing and Correlation Distribution Techniques," in *IEEE Access*, vol. 10, pp. 17578-17589, 2022, doi: 10.1109/ACCESS.2022.3150833.
11. Timon, Benjamin. "Non-profiled deep learning-based side-channel attacks with sensitivity analysis." *IACR Transactions on Cryptographic Hardware and Embedded Systems* (2019): 107-131.
12. Do, Ngoc Tuan, et al. "On the performance of non-profiled side channel attacks based on deep learning techniques." *IET Information Security* (2022).
13. N. -T. Do, P. -C. Le, V. -P. Hoang, V. -S. Doan, H. G. Nguyen and C. -K. Pham, "MO-DLSCA: Deep Learning Based Non-profiled Side Channel Analysis Using Multi-output Neural Networks," *2022 International Conference on Advanced Technologies for Communications (ATC)*, Ha Noi, Vietnam, 2022, pp. 245-250, doi: 10.1109/ATC55345.2022.9943024.