(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ

(11) **1-0051509**

(51)[2020.01] **H04L 9/00**; G06N 3/02

(13) **B**

(21) 1-2021-07694

(22) 30/11/2021

(45) 25/09/2025 450

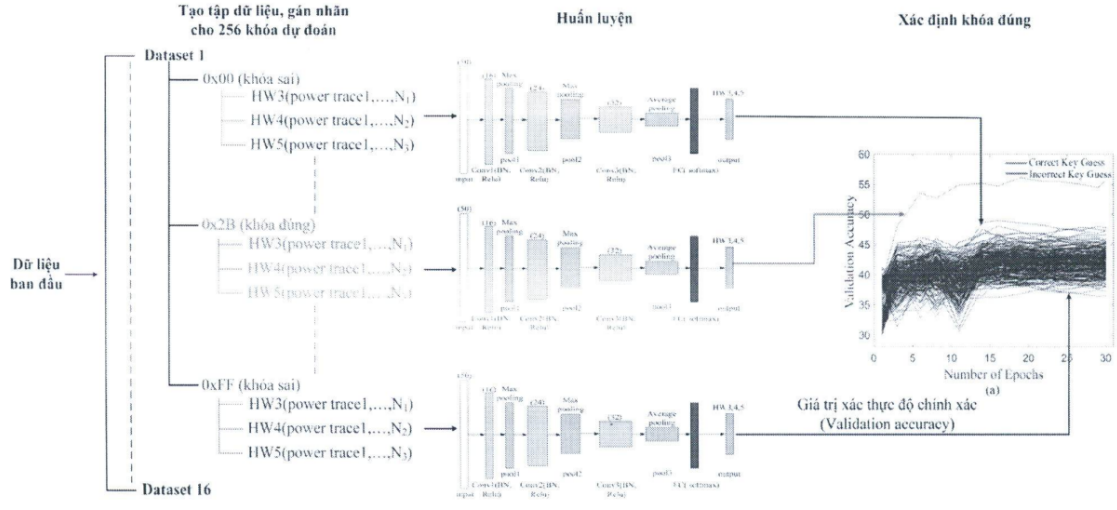
(43) 25/02/2022 407A

(76) 1. HOÀNG VĂN PHÚC (VN)
Viện Tích Hợp Hệ Thống/ Học Viện Kỹ Thuật Quân Sự, 236 Hoàng Quốc Việt, Quận Bắc Từ Liêm, Hà Nội
2. Đỗ Ngọc Tuấn (VN)
Bộ môn Điện tử số, Khoa Cơ sở, Trường Sĩ quan Thông Tin, số 101 đường Mai Xuân Thưởng, Phường Bắc Nha Trang, Tỉnh Khánh Hòa.
3. Đoàn Văn Sáng (VN)
Khoa Thông tin- ra đa/Học viện Hải quân, 30 Trần Phú, Nha Trang, Khánh Hòa

(54) PHƯƠNG PHÁP THỰC HIỆN PHÂN TÍCH KÊNH BÊN KHÔNG LẬP MẪU SỬ DỤNG MẠNG NƠ-RON TÍCH CHẬP ĐỐI VỚI THUẬT TOÁN MÃ MẬT AES 128 BIT

(21) 1-2021-07694

(57) Sáng chế đề cập đến phương pháp sử dụng mạng nơ-ron tích chập trong phân tích kênh bên không lập mẫu và ứng dụng của phương pháp này trong đánh giá bảo mật các bộ xử lý chuyên dụng thực thi thuật toán AES 128 bit. Kỹ thuật xây dựng tập dữ liệu cho phân tích bằng mạng nơ-ron sử dụng kĩ thuật phân đoạn trong phân tích tương quan công suất, hình thành nên một tập dữ liệu có số chiều nhỏ, phù hợp với mô hình mạng nơ-ron đề xuất. Sáng chế đề xuất áp dụng mô hình mạng nơ-ron tích chập cho phân tích kênh bên không lập mẫu. Nguyên lý chính của áp dụng mạng CNN trong phân tích khóa bí mật AES 128 bit là sử dụng các tham số của mô hình trong quá trình huấn luyện. Bằng việc xây dựng tập dữ liệu dựa trên giá trị tương quan cao nhất giữa power trace và giá trị trung gian HW. Nhóm tác giả đã tạo ra được mối liên hệ giữa đầu vào (các mẫu trên power trace) và giá trị đầu ra (HW). Vì vậy, khi áp dụng mạng CNN, ứng với khóa dự đoán là chính xác, mạng CNN có thể dễ dàng huấn luyện và phân loại đơn giản với 3 nhãn. Điều này dẫn đến các tham số của mạng CNN trong quá trình huấn luyện như độ chính xác (accuracy) sẽ tăng dần hoặc giá trị của hàm mất mát (loss function) sẽ giảm dần. Ngược lại, với khóa dự đoán là sai, giá trị trung gian cũng sẽ bị tính toán sai, dẫn đến mạng CNN không thể huấn luyện được. Chính từ các tham số này, mô hình sẽ quyết định được từ khóa giả thiết nào là đúng, và khóa giả thiết sai với độ tin cậy cao. Kết quả thực nghiệm cho thấy tập dữ liệu đã xây dựng và mô hình mạng đề xuất hoạt động hiệu quả, có thể phân biệt được khóa đúng trong điều kiện có nhiễu nhỏ.



Hình 1

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp sử dụng mạng nơ-ron tích chập trong phân tích kênh bên không lập mẫu trong đánh giá bảo mật đối với thuật toán mã mật AES 128 bit.

Tình trạng kỹ thuật của sáng chế

Trong thời đại kỹ thuật số ngày nay, các thiết bị mã hoá sẽ thực hiện biến bản tin rõ thành bản tin được mã hoá để bảo vệ dữ liệu. Tuy nhiên, trong quá trình thực thi mã hoá hoặc giải mã thì các thiết bị này luôn có sự rò rỉ thông tin nhất định và được gọi là thông tin kênh kề, ví dụ thời gian thực hiện, bức xạ điện từ trường, công suất tiêu thụ điện. Từ các nguyên lý kỹ thuật căn bản về cấu tạo của các vi mạch điện tử, chúng ta nhận định rằng công suất tiêu thụ của thiết bị mã hoá không phải là một tham số ngẫu nhiên mà nó phụ thuộc rất nhiều vào quá trình thiết bị xử lý dữ liệu, đặc biệt là thực hiện tính toán trong thời gian đủ lớn. Do đó, thông tin về công suất tiêu thụ không mong muốn này được chính kẻ tấn công lợi dụng nhằm phân tích và đánh cắp các khoá bí mật.

Tấn công kênh bên thường được chia thành hai hướng tiếp cận đó là tấn công lập mẫu (Profiled attack) và không lập mẫu (Non-profiled attack). Đối với tấn công lập mẫu, điển hình là Template attack, 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 phải thực hiện mã hóa và ghi lại một lượng lớn vết công suất tiêu thụ (power trace) của mỗi từ khóa dự đoán. Thứ hai là áp dụng mẫu đã có lên thiết bị mục tiêu. Người tấn công thực hiện mã hóa trên thiết bị mục tiêu, ghi lại một lượng nhỏ vết công suất, so sánh các khóa dự đoán với mẫu đã lập và xác định key với mẫu nào giống nhất. Về lý thuyết, tấn công lập mẫu được coi là tấn công hiệu quả khi chỉ cần một số lượng nhỏ các vết công suất. Tuy nhiên, tấn công lập mẫu thường khó áp dụng trên thực tế khi mà người tấn công chỉ có duy nhất thiết bị mục tiêu và số lượng vết công suất thu được bị giới hạn. Ngược lại, với tấn công không lập mẫu, ví dụ như tấn công phân tích vi sai công suất tiêu thụ (DPA: *Differential Power Analysis*), hoặc tấn công phân tích tương quan công suất (CPA: *Correlation Power Analysis*), người tấn công chỉ cần sử dụng thiết bị mục tiêu để ghi lại vết công suất và dùng các kỹ thuật thống kê để tìm ra khóa bí mật.

Trong những năm gần đây, dưới sự phát triển không ngừng của trí tuệ nhân tạo (AI: *Artificial Intelligence*), các ứng dụng của nó đang ngày càng trở nên thiết thực và quan trọng mà điển hình là kỹ thuật học máy (ML: *Machine Learning*) hay kỹ thuật

học sâu (DL: *Deep Learning*). Trong đó, mạng nơ-ron tích chập (CNN: *Convolutional Neural Network*) là một mô hình mạng thuộc kỹ thuật học sâu, được ứng dụng chính trong việc nhận dạng ảnh, phân loại đối tượng, nhận diện gương mặt và một số lĩnh vực khác. Từ các dữ liệu đầu vào, trải qua quá trình huấn luyện của mạng CNN, máy tính hoàn toàn có thể phân tích và phân loại các dữ liệu ở đầu ra một cách nhanh chóng và chính xác. Chính vì vậy, ứng dụng trí tuệ nhân tạo trong lĩnh vực bảo mật đang trở thành một xu hướng nghiên cứu trên thế giới.

Năm 2018, một số nghiên cứu ứng dụng mạng CNN để tìm ra khóa bí mật của mã hóa AES bằng phương pháp lập mẫu được công bố. Tuy nhiên, B. Timon năm 2019 đã chứng minh mạng CNN không chỉ áp dụng trong tấn công lập mẫu mà còn có thể áp dụng rất tốt với tấn công không lập mẫu. Một điểm chú ý là các công bố trên chỉ áp dụng với dữ liệu có số mẫu đầu vào nhỏ. Trên thực tế với dữ liệu đầu vào là các vết công suất, số lượng mẫu là rất lớn, thường là vài nghìn cho đến hàng chục nghìn mẫu. Trong khi đó, số lượng mẫu đầu vào cho mạng CNN ảnh hưởng rất nhiều đến quá trình xử lý và tính toán của mạng. Vì vậy, sáng chế này đề xuất xây dựng tập dữ liệu và sử dụng mô hình mạng CNN nhằm thực hiện phân tích mã mật AES 128 bit trong điều kiện không lập mẫu. Ứng dụng trong đánh giá bảo mật của một số vi mạch thực thi thuật toán mã hóa AES như vi điều khiển, FPGA.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là khắc phục những hạn chế của các giải pháp trước đây, xây dựng dữ liệu học máy, ứng dụng kỹ thuật học máy vào trong đánh giá bảo mật phần cứng. Do đó, sáng chế đề xuất kỹ thuật và qui trình xây dựng tập dữ liệu học máy của một vi mạch và xây dựng kiến trúc mạng CNN sử dụng trong phân tích kênh bên thuật toán AES 128 bit, từ đó đánh giá mức độ bảo mật cho vi mạch. Trên cơ sở đó, sáng chế đề xuất phương pháp thực hiện phân tích kênh bên không lập mẫu sử dụng mạng nơ-ron tích chập đối với thuật toán mã mật AES 128 bit.

Về xây dựng tập dữ liệu học máy của một vi mạch, sáng chế đề xuất sử dụng kỹ thuật phân tích tương quan công suất tiêu thụ để thực hiện. Kỹ thuật này khai thác sự tương quan giữa công suất tiêu thụ thực tế và một mô hình công suất giả thiết. Để xác định được sự tương quan nói trên, nhóm tác giả sử dụng hệ số tương quan Pearson được tính theo công thức (1).

$$\rho_{k,i} = \frac{\sum_{n=1}^{N}\left[\left(h_{n,k} - \overline{h_k}\right)\left(t_{n,i} - \overline{t_i}\right)\right]}{\sqrt{\sum_{n=1}^{N}\left(h_{n,k} - \overline{h_k}\right)^2 \sum_{n=1}^{N}\left(t_{n,i} - \overline{t_i}\right)^2}} \quad (1)$$

Trong đó $\bar{h}_k$ và $\bar{t}_i$ lần lượt là trung bình của mô hình công suất dự đoán và công suất tiêu thụ thực tế tại mẫu *i*. Các mô hình công suất thường được sử dụng là mô hình trọng số Hamming và khoảng cách Hamming, sáng chế sử dụng mô hình trọng số Hamming (HW) và tập trung khai thác vòng mã hóa số 1 của thuật toán. Điểm đặc nổi bật trong kỹ thuật này là việc sử dụng ít hơn giá trị trọng số Hamming, chỉ sử dụng 3 giá trị (HW=[3;5]) so với 9 giá trị (HW=[0;8]) của các giải pháp được đề xuất trước đây. Việc sử dụng ít giá trị trọng số Hamming dẫn đến số lượng nhãn (label) của mô hình phân loại trở nên ít hơn, dẫn đến mô hình mạng huấn luyện đơn giản, dễ thực hiện, đồng thời giảm số lượng vệt công suất khoảng 30% so với các đề xuất tương tự. Kết quả cuối cùng của kỹ thuật này là một tập dữ liệu (dataset) có số chiều nhỏ hơn gấp nhiều lần (sáng chế giảm chiều dữ liệu từ 10000 xuống còn 50 đối với dữ liệu được lấy theo minh họa trên Hình 2).

Đối với mô hình mạng CNN, sáng chế thực hiện xây dựng trên cơ sở cấu trúc mạng cơ bản nhất của một mạng CNN, nghĩa là mạng bao gồm các lớp cơ bản như: Lớp đầu vào (input), Lớp tích chập (Convolutional Layer), xen giữa các lớp tích chập là lớp Pooling, cuối cùng là lớp kết nối đầy đủ (Fully connected) như minh họa trong Hình 5. Mục đích chính của sử dụng mô hình CNN trong điều kiện không lập mẫu là có thể học cách phân loại các dữ liệu đầu vào là các vệt công suất, đầu ra là các giá trị trọng số Hamming tương ứng. Đặc biệt, trường hợp ứng dụng CNN ở đây hoàn toàn khác với tấn công có lập mẫu, thay vì mô hình mạng phải đưa ra kết quả cuối cùng là từ khóa đúng, mô hình đề xuất của nhóm tác giả sử dụng chính các tham số trong quá trình huấn luyện để kết luận khóa bí mật. Điều này dẫn đến có thể loại bỏ khóa giả thiết sai ngay từ trong quá trình huấn luyện.

Mô tả vắn tắt các hình vẽ

Hình 1 là hình vẽ minh họa quá trình đánh giá bảo mật của sáng chế.

Hình 2 là hình vẽ minh họa sơ đồ kết nối thu thập vệt công suất (power trace).

Hình 3 là hình vẽ minh họa kỹ thuật tách 50 mẫu tín hiệu có giá trị tương quan cao của sáng chế.

Hình 4 là hình vẽ minh họa cấu trúc của tập dữ liệu (dataset) đề xuất của sáng chế.

Hình 5 là hình vẽ minh họa sơ đồ mô tả kiến trúc mạng nơ-ron tích chập (CNN: Convolutional Neural Network) đề xuất.

Hình 6 là hình vẽ minh họa biểu đồ thể hiện khả năng phân biệt giữa khóa đúng và khóa sai của sáng chế.

Hình 7 là hình vẽ minh họa đồ thị đánh giá sự ảnh hưởng bởi nhiễu lên tham số accuracy cho phân tích hoàn toàn 1 byte khóa bí mật.

Mô tả chi tiết sáng chế

Hình 1 minh họa quá trình thực hiện đánh giá bảo mật của sáng chế từ sử dụng tập dữ liệu huấn luyện, gán nhãn cho đến tấn công và cuối cùng là phân tích kết quả.

Hình 2 thể hiện quy trình thu thập các dữ liệu về tiêu thụ công suất. Dữ liệu tiêu thụ công suất này được ghi lại từ vi điều khiển Atmel Xmega gắn trên board mạch chuyên dụng ChipWhisperer (CW). Quá trình thực hiện như sau:

i) Thực hiện lập trình thực thi thuật toán mã hóa AES 128 bit trên vi điều khiển thông qua máy tính cá nhân (PC).

ii) Kết nối board mạch chuyên dụng CW với máy tính bằng cáp giao tiếp USB.

iii) Thực hiện gửi bản tin cần mã hóa (plaintext) và chuỗi khóa 128 bit cố định xuống vi điều khiển Xmega. CW tiến hành mã hóa, ghi lại giá trị tỉ lệ với công suất tiêu thụ trong quá trình mã hóa, gửi về PC bản tin đã được mã hóa (CipherText) cùng dữ liệu công suất tiêu thụ.

iv) PC thực hiện nhận và lưu các dữ liệu về PlainText, CipherText, Power trace dưới dạng file *.npy.

Hình 3 minh họa kỹ thuật tách các mẫu công suất có giá trị tương quan cao nhất trong một vệt công suất. Trước hết, nhóm tác giả xác định vị trí của các điểm có giá trị tương quan cao. Phương pháp này khai thác sự tương quan giữa công suất tiêu thụ thực tế và mô hình công suất giả thiết. Sáng chế sử dụng mô hình trọng số Hamming và được xác định bằng công thức (2)

$$h = HW(SubByte(PlainText \oplus Key)) \quad (2)$$

trong đó h là giá trị HW thu được, SubByte là phép thế Byte của thuật toán AES-128 với đầu vào là kết quả của phép tính XOR giữa bản tin cần mã hóa (Plaintext) và khóa bí mật (key).

Từ giá trị của trọng số Hamming, giá trị tương quan được xác định theo công thức sau:

$$\rho_{k,i} = \frac{\sum_{n=1}^{N}\left[\left(h_{n,k} - \overline{h_k}\right)\left(t_{n,i} - \overline{t_i}\right)\right]}{\sqrt{\sum_{n=1}^{N}\left(h_{n,k} - \overline{h_k}\right)^2 \sum_{n=1}^{N}\left(t_{n,i} - \overline{t_i}\right)^2}} \quad (1)$$

Trong đó $\overline{h_k}$ và $\overline{t_i}$ lần lượt là trung bình của mô hình công suất dự đoán và công suất tiêu thụ thực tế tại mẫu i. Tuy nhiên, so với các phương pháp truyền thống, giá trị $\rho_{k,i}$ phải được tính trên toàn bộ các mẫu của tất cả N power trace nhằm tìm ra giá trị lớn nhất. Tuy nhiên qua nghiên cứu, nhóm tác giả đề xuất việc phân đoạn tính giá trị

tương quan, nhằm giảm số lượng phép tính từ việc tính toán với số lượng mẫu lớn. Bằng phương pháp trên, sáng chế có số phép tính cần thực hiện $16\times K\times\alpha\times N$ so với phương pháp thông thường là $16\times K\times L\times N$ (Với N là số lượng power trace cần phân tích, L là số mẫu có trong một power trace và K là số lượng khóa giả thiết trong một Byte, a rất nhỏ so với L). Giá trị α nói trên chính là tổng số mẫu sẽ được sử dụng để tạo ra tập dữ liệu mới với số chiều giảm rất nhiều so với ban đầu. Sau khi đã xác định được tổng số mẫu α, đồng thời vị trí của từng mẫu có giá trị tương quan từ cao xuống thấp được xác định. Đảm bảo mỗi vị trí có giá trị tương quan là cao nhất so với mẫu còn lại, không bị trùng lặp. Nhóm tác giả thực hiện tách tất cả các mẫu tại vị trí đó của toàn bộ power trace, tạo nên tập power trace mới với α điểm. Chi tiết được thể hiện bằng các vị trí đánh dấu màu đỏ trong hình 3. Kỹ thuật phân đoạn đã được nhóm tác giả công bố trong báo cáo "*An Efficient Side Channel Attack Technique with Improved Correlation Power Analysis" Industrial Networks and Intelligent Systems. INISCOM 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 334. Springer, Cham.* https://doi.org/10.1007/978-3-030-63083-6_22.

Hình 4 mô tả cấu trúc dữ liệu được tạo ra từ kỹ thuật đề xuất của sáng chế. Áp dụng kỹ thuật phân đoạn, sáng chế đề xuất thuật toán 1 cho việc xây dựng tập dữ liệu huấn luyện. Với N là số lượng power trace cần phân tích, L là số mẫu có trong một power trace, a rất nhỏ so với L. N được chia thành N_1 và N_2 thỏa mãn $N_1 \le \frac{1}{2}N$, $d_{i,B}$ là bản tin mã hóa thứ i, btye B, $t_{i,j}$ là giá trị mẫu thứ j của power trace thứ i.

Thuật toán 1: Tạo tập dữ liệu huấn luyện từ dữ liệu power trace ban đầu

Input: $d_{i,B}$, $t_{i,j\ (1\le i\le N,\ 1\le j\le L)}$, L, N_1, N ($N_1 \le \frac{1}{2}N$)

Khởi tạo các folder chứa data: HW3, HW4, HW5

for $B \in (1;16)$ **do**

//Bước 1: xác định vị trí của a điểm có giá trị tương quan cao nhất

for $KEY \in (0;255)$ **do**

$$h_{1<i<N} = HW(SubByte(d_{i,B} \oplus KEY))$$

$$mean_h = \sum_{i=1}^{N_1} h_i, mean_{t_j} = \sum_{i=1}^{N_1} t_{i,j}$$

for $i \in (1;N_1)$ **do**

$hdif1 = h_i - mean_h$

$tdif1 = t_{i,j} - mean_t$

$sum1\_1 = sum1\_1 + hdif1 * tdif1$

$sum1\_2 = sum1\_2 + hdif1 * hdif1$

$sum1\_3 = sum1\_3 + tdif1.*tdif1$

end for

$\rho_{KEY,\bar{L}} = \text{abs}(sum1\_1 ./ sqrt(sum1\_2.*sum1\_3));'$

end for

for $i \in (1;\alpha)$ **do**

$\rho_{\bar{L}'} = \arg\max(\rho_{KEY,\bar{L}} - \rho_{\bar{L}'-1})$

$\bar{N}_{i,\bar{L}'} = N_{i,\bar{L}'}$

end for

//Bước 2: Sử dụng a điểm đã xác định tại bước 1, tạo N power trace mới từ N power trace ban đầu, mỗi power trace có a mẫu.

for $KEY \in (0;255)$ **do**

for $i \in (1:N)$ **do**

$h_{1<i<N} = HW(SubByte(d_{i,B} \oplus KEY))$

switch h

case 3 save file // 50 điểm của power trace tương ứng vào foler HW3

case 4 save file // 50 điểm của power trace tương ứng vào foler HW4

case 5 save file // 50 điểm của power trace tương ứng vào foler HW5

end

end for

end for

end for

So với dữ liệu gốc ban đầu, số chiều của dữ liệu đã giảm từ L xuống còn a. Điều này có nghĩa dữ liệu đầu vào cho mô hình huấn luyện CNN là dữ liệu 1 chiều dạng vetor a điểm.

Hình 5 thể hiện kiến trúc mạng CNN đề xuất, nhóm tác giả đã xây dựng mô hình mạng CNN bao gồm có thành phần lớp nhận giá trị đầu vào, các lớp tích chập xem giữa là các lớp phi tuyến (pooling) và cuối cùng là lớp đầu ra. Một cách tổng quát, kết quả đầu ra y của dữ liệu đầu vào x được tính theo công thức sau:

$$F(x) = s \circ \left[\lambda\right]^{n_{dense}} \circ \left[\delta \circ \left[\alpha \circ \gamma\right]^{n_{conv}}\right]^{n_{blocks}} (x) = y. \quad (3)$$

Trong đó, s ký hiệu cho lớp Softmax, thực hiện chuyển đổi kết quả đầu ra thể hiện dưới dạng xác suất để đầu vào x_i rơi vào lớp i, tổng tất cả các giá trị của x_i bằng 1 và được tính theo công thức (4), α là hàm kích hoạt, thực hiện hàm phi tuyến được tính theo công thức (5). Kí hiệu δ đại diện cho lớp pooling.

$$x_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_i}}, \forall i = 1,2,..C \tag{4}$$

trong đó, $z_i = \mathrm{w}^T x$ (w là trọng số cần hiệu chỉnh khi huấn luyện, x là giá trị đầu vào)

$$ELU : R_{(z)} = \begin{cases} z & z > 0 \\ \alpha \cdot (e^z - 1) & z <= 0 \end{cases} \tag{5}$$

trong đó, α là giá trị mà ELU sẽ tiệm cận khi z là một số âm lớn, thường có giá trị là 1.

Ứng với tập dữ liệu đã đề xuất, mô hình CNN sử dụng trong sáng chế có lớp đầu vào chứa 50 nút mạng tương ứng với dữ liệu vào là 50 điểm của tập dữ liệu. Ngay sau đó là lớp tích chập sử dụng 16 bộ lọc, kích thước [1 3], số lượng dịch chuyển trên ma trận đầu vào (stride) là 1 vì vậy kích thước dữ liệu đầu vào sẽ bằng với dữ liệu đầu ra. Ngay sau lớp tích chập, kỹ thuật Batch Normalization được sử dụng để chuẩn hóa các đầu ra của lớp tích chập về trạng thái zero-mean với độ lệch chuẩn là 1. Như minh họa tại hình 5, tất cả lớp tích chập đều được đi kèm với lớp chuẩn hóa này. Xen giữa các lớp tích chập là các lớp giảm kích thước (pooling) với kích thước [1 2] và stride [1 2]. Tuy nhiên để việc lấy đặc trưng của dữ liệu huấn luyện được đa dạng, nhóm tác giả sử dụng 2 dạng khác nhau là Max Pooling và Average Pooling.

Nguyên lý chính của áp dụng mạng CNN trong phân tích khóa bí mật AES là sử dụng các tham số của mô hình trong quá trình huấn luyện. Bằng việc xây dựng tập dữ liệu dựa trên giá trị tương quan cao nhất giữa power trace và giá trị trung gian HW. Nhóm tác giả đã tạo ra được mối liên hệ giữa đầu vào (50 mẫu) và giá trị đầu ra (HW). Vì vậy, khi áp dụng mạng CNN, ứng với khóa dự đoán là chính xác, mạng CNN có thể dễ dàng huấn luyện và phân loại đơn giản với 3 nhãn. Điều này dẫn đến các tham số của mạng CNN trong quá trình huấn luyện như độ chính xác (accuracy) sẽ tăng dần hoặc giá trị của hàm mất mát (loss function) sẽ giảm dần. Ngược lại, với khóa dự đoán là sai, giá trị trung gian cũng sẽ bị tính toán sai, dẫn đến mạng CNN không thể huấn luyện được. Điều này dẫn đến các tham số nêu trên sẽ giữ nguyên hoặc thay đổi không đáng kể qua mỗi epoch huấn luyện. Đây chính là yếu tố chính để ứng dụng mạng CNN nói riêng hay các kỹ thuật học sâu khác vào phân tích khóa AES không cần lập mẫu. Tuy nhiên, với mỗi mô hình chỉ thực hiện huấn luyện dự đoán cho một giá trị khóa, vì

vậy với 256 giá trị khóa dự đoán, chúng ta cần chạy 256 mô hình mạng CNN để tìm ra khóa nào đúng.

Hình 6 minh họa kết quả của huấn luyện cũng như phương pháp để lấy được khóa đúng với kỹ thuật phân tích không lập mẫu sử dụng CNN. Đối với khóa sai sự phân bố của các điểm thực sự thuộc về một nhãn sẽ tập trung vào HW4, ngược lại, với khóa đúng, các điểm thực sự thuộc vào các nhãn có tỉ lệ phần trăm cao, thể hiện mô hình được huấn luyện và tiến hành phân loại tốt. Các giá trị này sẽ phân bố đều ra cả ba nhãn HW tạo thành một đường chéo. Từ đây, thông qua trực quan hoặc sử dụng lập bảng và tìm giá trị lớn nhất của tham số độ chính xác, có thể tìm được toàn bộ khóa bí mật AES như minh họa trong Hình 6b,c. Tuy nhiên để xác định toàn bộ khóa đúng, không thể thực hiện in ra toàn bộ hình ảnh phân bố HW cho 256 giá trị khóa giả thiết, vì vậy sáng chế đề xuất sử dụng giá trị xác thực độ chính xác (Validation accuracy), với mục đích phân biệt khóa dự đoán đúng ứng với giá trị tham số accuracy cao nhất như minh họa trong Hình 6a.

Hình 7 thể hiện kết quả phân tích khóa bí mật của sáng chế dưới sự tác động của nhiễu. Nhóm tác giả đặt giả thiết quá trình thu power trace bị ảnh hưởng bởi nhiễu ngẫu nhiên tác động lên giá trị toàn bộ mẫu của power trace. Mức độ của nhiễu trong khảo sát được xác định bởi độ lệch chuẩn có giá trị lần lượt là 0.025, 0.05 và 0.075. Ứng với 3 mức độ tác động bởi nhiễu, ở mức nhiễu nhỏ, mô hình đề xuất sử dụng tập dữ liệu của sáng chế cho kết quả tin cậy cao, khi tham số accuracy của khóa đúng là khoảng 55, trong khi tất cả các khóa sai còn lại tham số accuracy có giá trị trong khoảng từ 30 đến 48. Quan sát trên đồ thị minh họa tại Hình 7a cho thấy, có thể phân biệt rõ ràng khóa đúng. Đặc biệt, có thể phân biệt khóa đúng chỉ sau 5 epoch huấn luyện bằng mô hình. Kết quả thực nghiệm được báo cáo trong *"Performance Analysis of Non-Profiled Side Channel Attacks Based on Convolutional Neural Networks," 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Ha Long, Vietnam, 2020, pp. 66-69, doi: 10.1109/APCCAS50809.2020.9301673.*

Hiệu quả đạt được bởi sáng chế

- Nhờ việc đề xuất kỹ thuật xây dựng tập dữ liệu cho huấn luyện, sáng chế có thể áp dụng cho nhiều loại vi điều khiển khác nhau hoặc các loại chip FPGA. Không phụ thuộc vào các tập dữ liệu mở.
- Nhờ sử dụng ít số lượng nhãn cho huấn luyện, số lượng power trace cần thiết để thực hiện phân tích nhỏ hơn 30% so với số lượng sử dụng trong phương pháp thông thường.
- Nhờ kiến trúc mạng nơ-ron có áp dụng một số lớp và hàm đặc trưng của mô hình mạng tích chập, kết quả phân tích khóa bí mật AES 128 bit trở nên tin cậy và hiệu quả.

YÊU CẦU BẢO HỘ

1. Phương pháp thực hiện phân tích kênh bên không lập mẫu sử dụng mạng nơ-ron tích chập đối với thuật toán mã mật AES 128 bit, phương pháp này bao gồm các bước:

xây dựng tập dữ liệu tiêu thụ công suất của một vi mạch (thuật toán 1) bao gồm các bước theo trình tự:

(i) xác định vị trí của a điểm có giá trị tương quan cao nhất sử dụng $N_1 \le \frac{1}{2}N$ (N là tổng số power trace),

(ii) sử dụng a điểm đã xác định tại bước 1, tạo N power trace mới từ N power trace ban đầu, mỗi power trace có a mẫu, trong đó:

để tính giá trị tương quan, phương pháp này sử dụng mô hình trọng số Hamming và được xác định bằng công thức (2)

$$h = HW(SubByte(PlainText \oplus Key)) \quad (2)$$

trong đó h là giá trị HW thu được, SubByte là phép thế Byte của thuật toán AES-128 với đầu vào là kết quả của phép tính XOR giữa bản tin cần mã hóa (Plaintext) và khóa bí mật (key),

từ giá trị của trọng số Hamming, giá trị tương quan được xác định theo công thức sau:

$$\rho_{k,i} = \frac{\sum_{n=1}^{N}\left[\left(h_{n,k} - \overline{h_k}\right)\left(t_{n,i} - \overline{t_i}\right)\right]}{\sqrt{\sum_{n=1}^{N}\left(h_{n,k} - \overline{h_k}\right)^2 \sum_{n=1}^{N}\left(t_{n,i} - \overline{t_i}\right)^2}} \quad (1)$$

trong đó $\overline{h_k}$ và $\overline{t_i}$ lần lượt là trung bình của mô hình công suất dự đoán và công suất tiêu thụ thực tế tại mẫu i, khác biệt ở chỗ, chỉ tính giá trị $\rho_{k,i}$ trên một phần trong số toàn bộ các mẫu của tất cả N power trace nhằm tìm ra giá trị lớn nhất, số phép tính cần thực hiện là $16 \times K \times \alpha \times N$ so với phương pháp thông thường là $16 \times K \times L \times N$ (Với N là số lượng power trace cần phân tích, L là số mẫu có trong một power trace và K là số lượng khóa giả thiết trong một Byte, a rất nhỏ so với L), giá trị α nói trên chính là tổng số mẫu sẽ được sử dụng để tạo ra tập dữ liệu mới với số chiều giảm rất nhiều so với ban đầu, sau khi đã xác định được tổng số mẫu α, đồng thời vị trí của từng mẫu có giá trị tương quan từ cao xuống thấp được xác định, đảm bảo mỗi vị trí có giá trị tương quan là cao nhất so với mẫu còn lại, không bị trùng lặp, thực hiện tách tất cả các mẫu tại vị trí đó của toàn bộ power trace, tạo nên tập power trace mới với α điểm;

sử dụng mạng nơ-ron tích chập (CNN) để phân tích tập dữ liệu đã xây dựng, trong đó kiến trúc mạng CNN bao gồm:

lớp Input: 50 nút,

lớp tích chập 1: 16 bộ lọc, kích thước [1 3], stride =[1 1], đi kèm Batch Nomarlization và hàm kích hoạt ELU,

lớp Max Pooling: kích thước [1 2], stride = [1 2],

lớp tích chập 2: 24 bộ lọc, kích thước [1 3], stride =[1 1], đi kèm Batch Nomarlization và hàm kích hoạt ELU

lớp Max Pooling: kích thước [1 2], stride = [1 2]

lớp tích chập 3: 32 bộ lọc, kích thước [1 3], stride =[1 1], đi kèm Batch Nomarlization và hàm kích hoạt ELU

lớp Average Pooling: kích thước [1 2], stride = [1 2]

lớp kết nối đầy đủ, hàm SoftMax,

trong đó:

kết quả đầu ra y của dữ liệu đầu vào x được tính theo công thức sau:

$$F(x) = s \circ \left[\lambda\right]^{n_{dense}} \circ \left[\delta \circ \left[\alpha \circ \gamma\right]^{n_{conv}}\right]^{n_{blocks}} (x) = y. \quad (3)$$

trong đó, s ký hiệu cho lớp Softmax, thực hiện chuyển đổi kết quả đầu ra thể hiện dưới dạng xác suất để đầu vào x_i rơi vào lớp i, tổng tất cả các giá trị của x_i bằng 1 và được tính theo công thức (4), α là hàm kích hoạt, thực hiện hàm phi tuyến được tính theo công thức (5), kí hiệu δ đại diện cho lớp pooling,

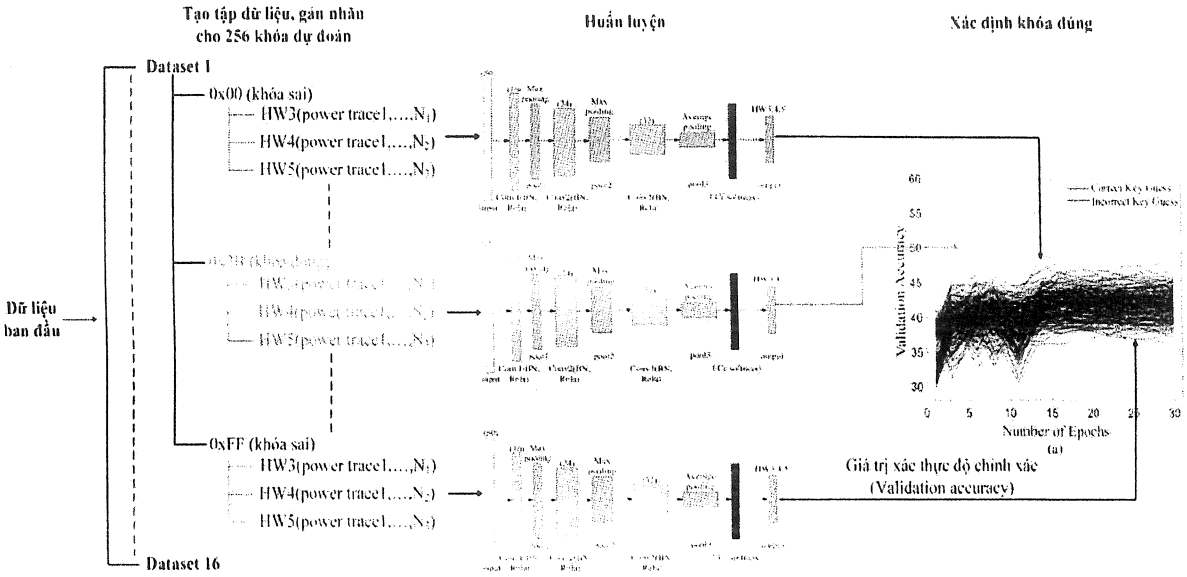
$$x_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_i}}, \forall i = 1,2,..C \quad (4)$$

trong đó, $z_i = \mathrm{w}^T x$ (w là trọng số cần hiệu chỉnh khi huấn luyện, x là giá trị đầu vào)

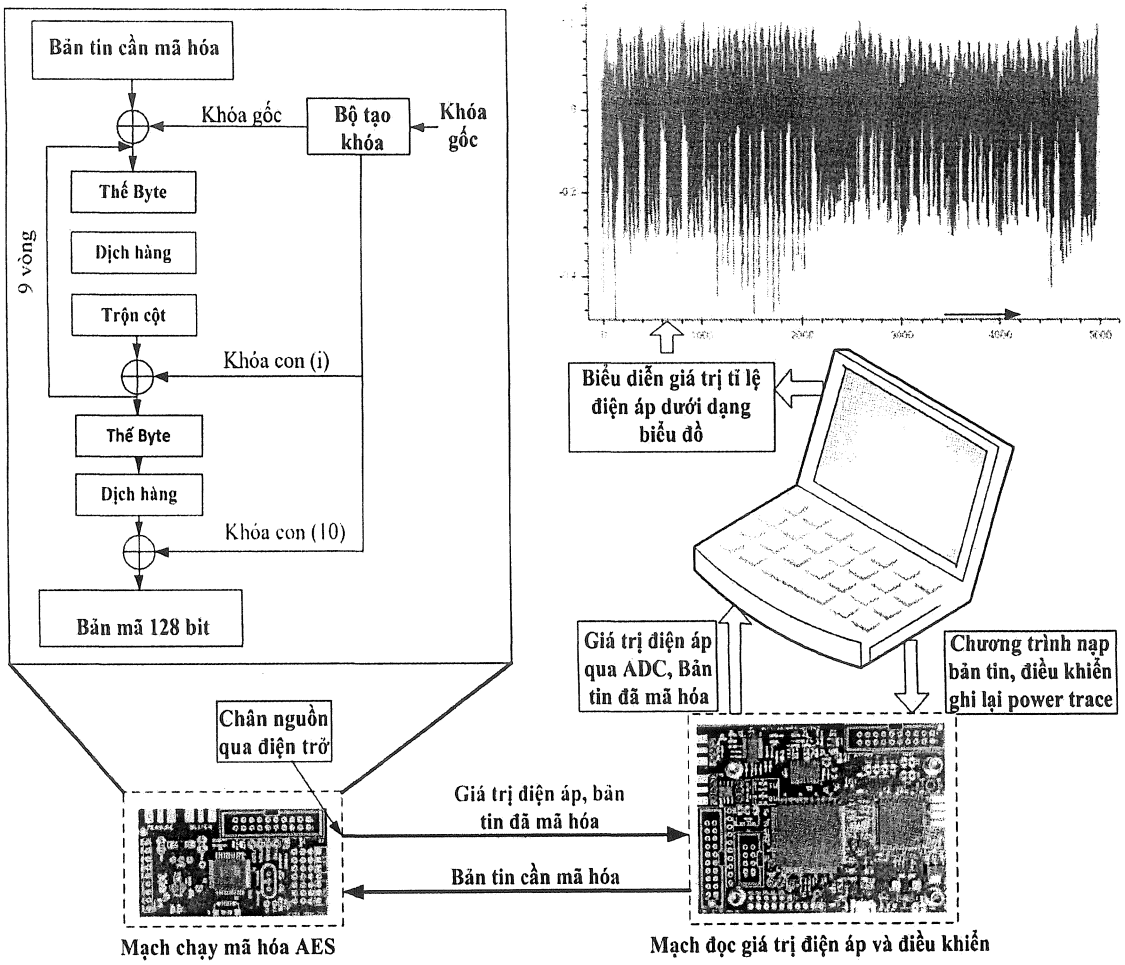
$$ELU: R_{(z)} = \begin{cases} z & z > 0 \\ \alpha \cdot (e^z - 1) & z <= 0 \end{cases} \quad (5)$$

trong đó, α là giá trị mà ELU sẽ tiệm cận khi z là một số âm lớn, thường có giá trị là 1,

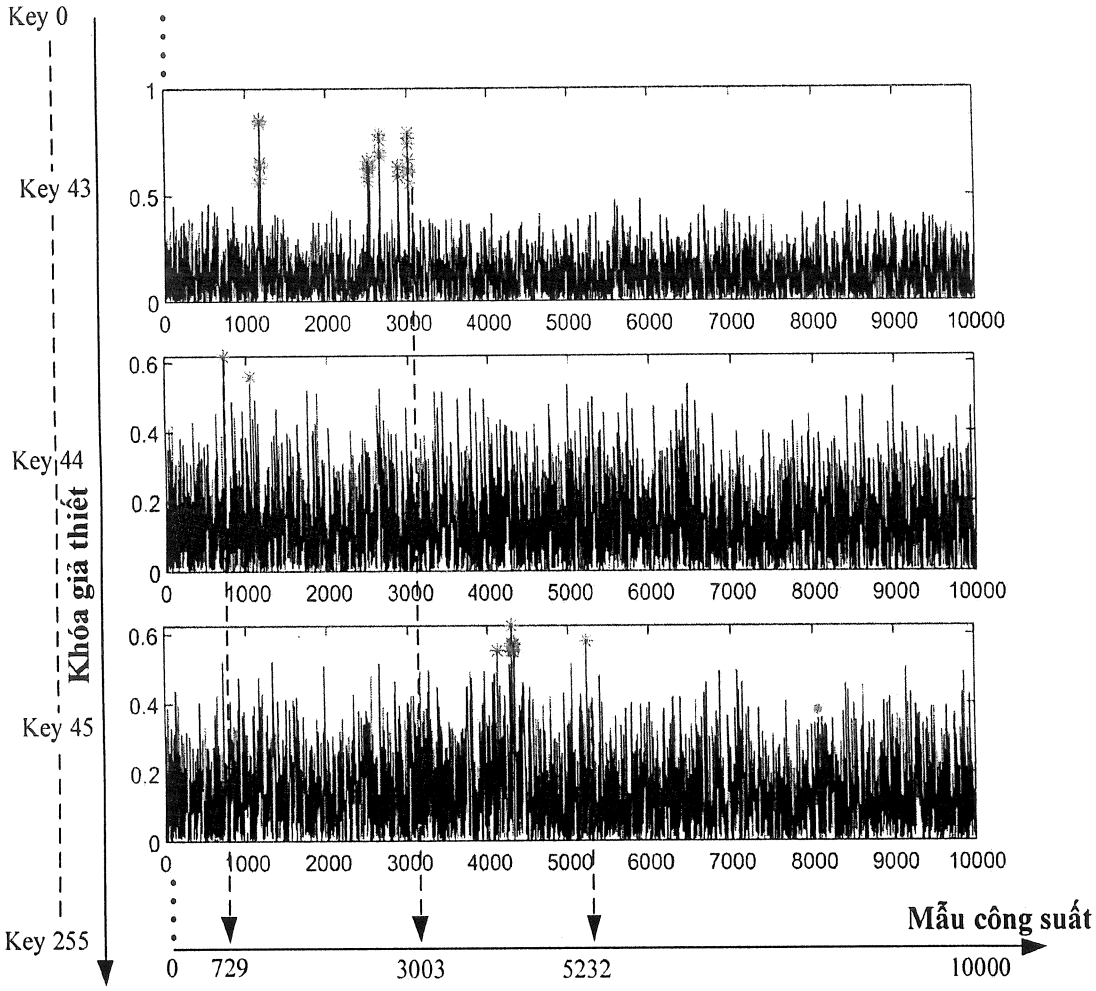
tạo ra được mối liên hệ giữa đầu vào (50 mẫu) và giá trị đầu ra (HW), bằng cách xây dựng tập dữ liệu dựa trên giá trị tương quan cao nhất giữa power trace và giá trị trung gian HW, vì vậy, khi áp dụng mạng CNN, ứng với khóa dự đoán là chính xác, mạng CNN có thể huấn luyện và phân loại đơn giản với 3 nhãn.


Tạo tập dữ liệu, gán nhãn cho 256 khóa dự đoán
Huấn luyện
Xác định khóa đúng
Dữ liệu ban đầu
Dataset 1
0x00 (khóa sai)
HW3(power trace1,...,N1)
HW4(power trace1,...,N2)
HW5(power trace1,...,N3)
0xFF (khóa sai)
HW3(power trace1,...,N1)
HW4(power trace1,...,N2)
HW5(power trace1,...,N3)
Dataset 16
Correct Key Guess
Incorrect Key Guess
Validation Accuracy
Number of Epochs
(a)
Giá trị xác thực độ chính xác
(Validation accuracy)

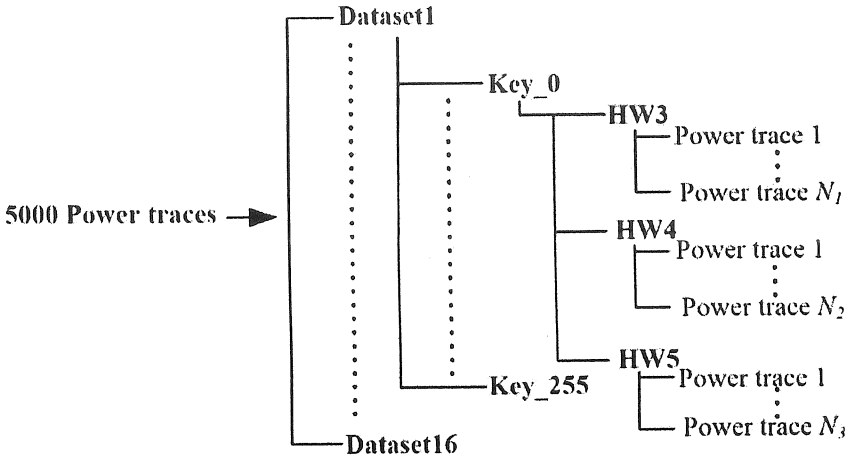
Hình 1


Bản tin cần mã hóa
Khóa gốc
Bộ tạo khóa
Khóa gốc
Thế Byte
Dịch hàng
Trộn cột
9 vòng
Khóa con (i)
Thế Byte
Dịch hàng
Khóa con (10)
Bản mã 128 bit
Biểu diễn giá trị tỉ lệ điện áp dưới dạng biểu đồ
Giá trị điện áp qua ADC, Bản tin đã mã hóa
Chương trình nạp bản tin, điều khiển ghi lại power trace
Chân nguồn qua điện trở
Giá trị điện áp, bản tin đã mã hóa
Bản tin cần mã hóa
Mạch chạy mã hóa AES
Mạch đọc giá trị điện áp và điều khiển

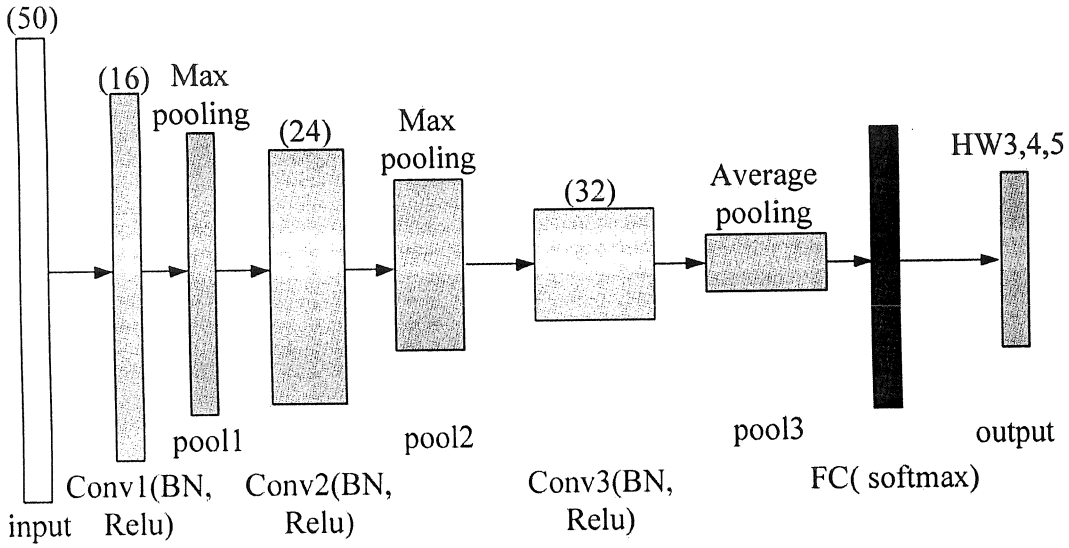
Hình 2


Key 0
Key 43
Key 44
Key 45
Key 255
Khóa giả thiết
Mẫu công suất
0
729
3003
5232
10000

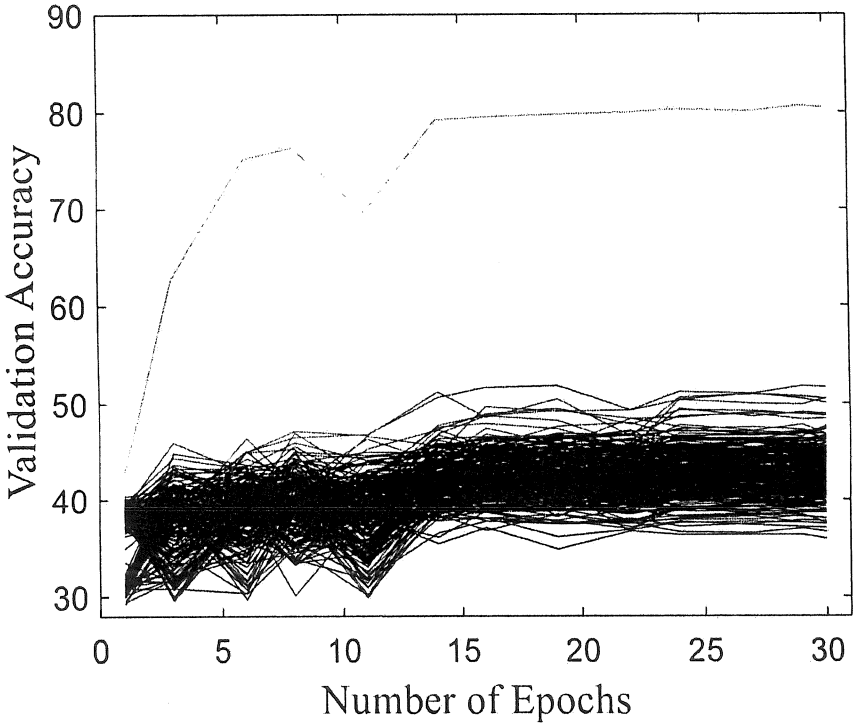
Hình 3


Dataset1
Key_0
HW3
Power trace 1
Power trace N_1
5000 Power traces
HW4
Power trace 1
Power trace N_2
HW5
Power trace 1
Power trace N_3
Key_255
Dataset16

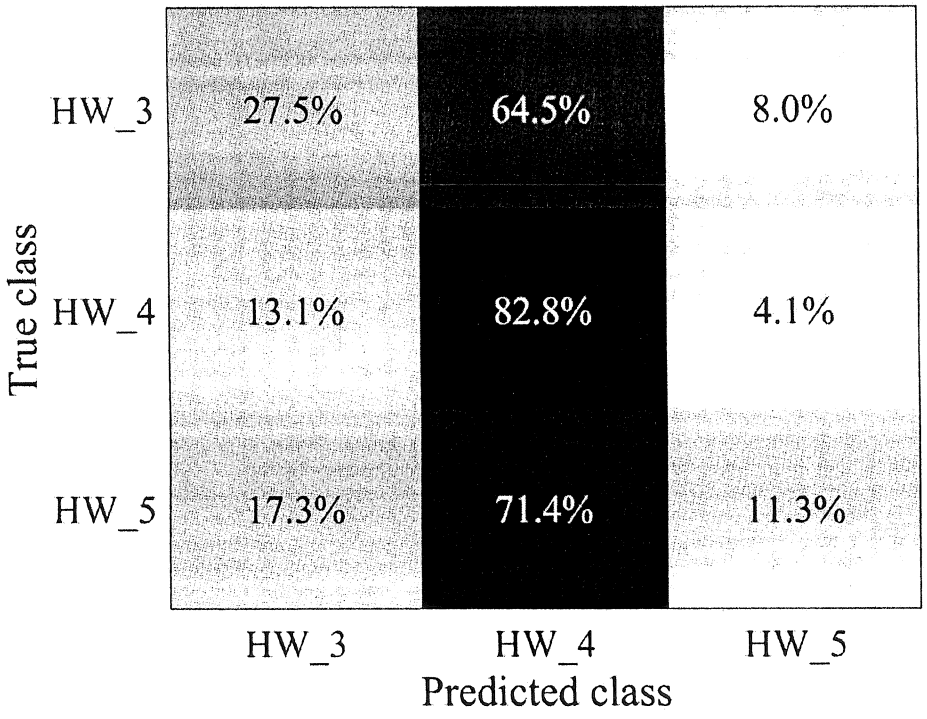
Hình 4


(50)
(16)
Max pooling
(24)
Max pooling
(32)
Average pooling
HW3,4,5
pool1
pool2
pool3
output
input
Conv1(BN, Relu)
Conv2(BN, Relu)
Conv3(BN, Relu)
FC(softmax)

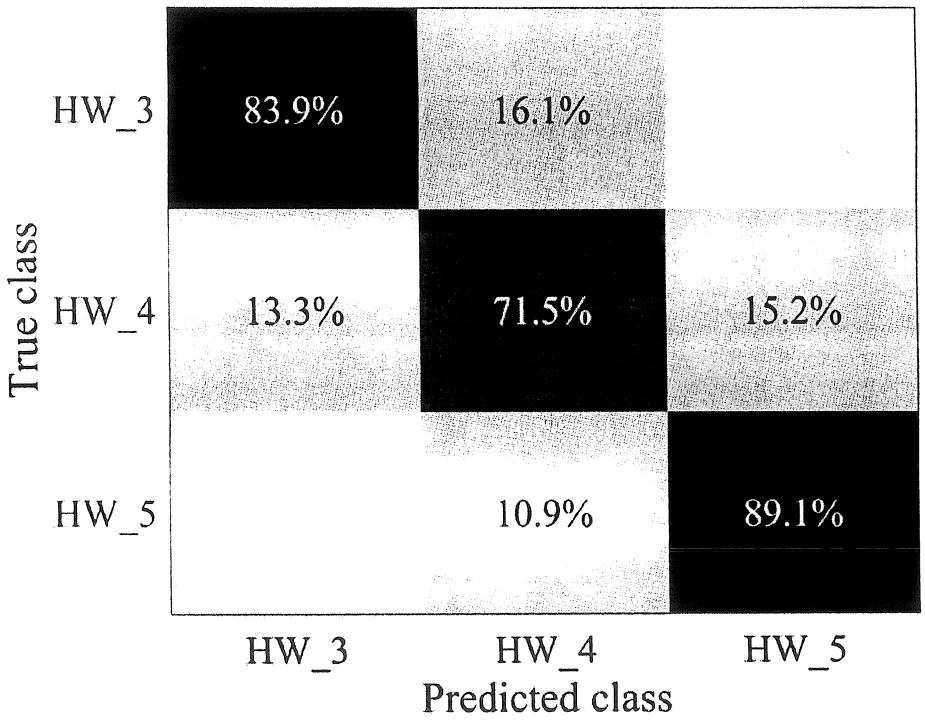
Hình 5


Validation Accuracy
90
80
70
60
50
40
30
0
5
10
15
20
25
30
Number of Epochs

a)

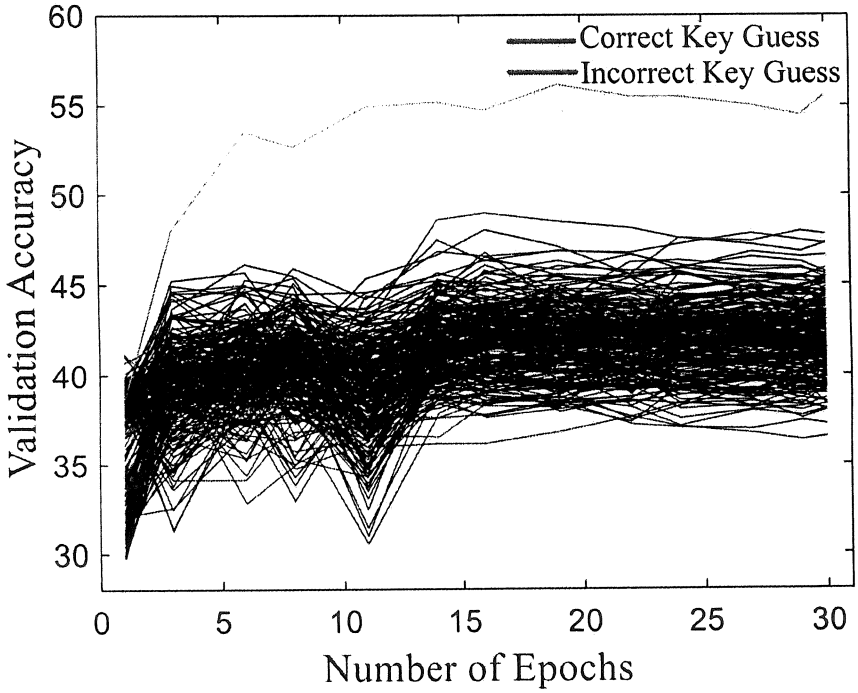

True class
HW_3
HW_4
HW_5
27.5%
64.5%
8.0%
13.1%
82.8%
4.1%
17.3%
71.4%
11.3%
HW_3
HW_4
HW_5
Predicted class

b)

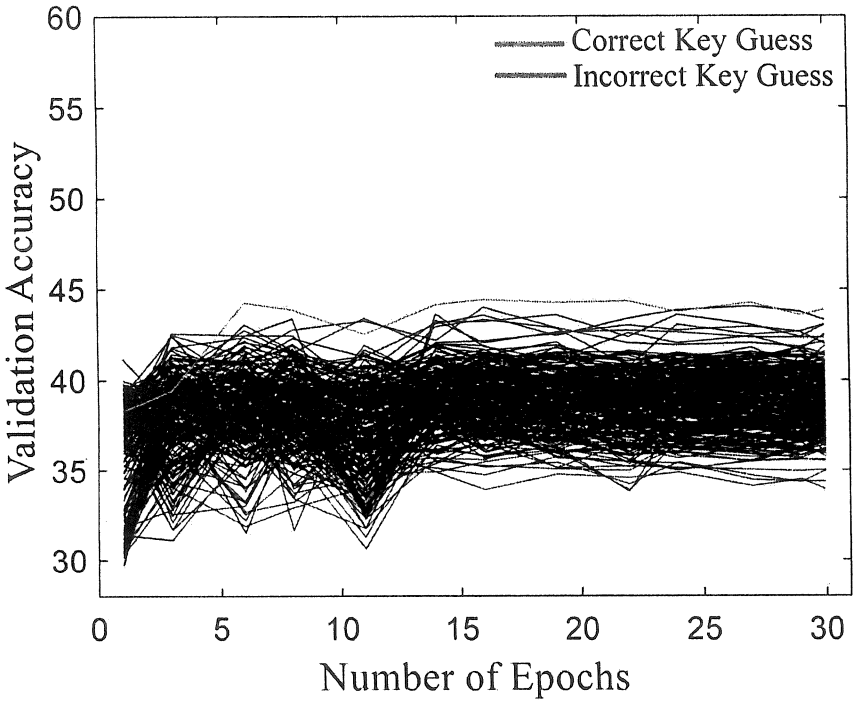

HW_3
HW_4
HW_5
True class
83.9%
16.1%
13.3%
71.5%
15.2%
10.9%
89.1%
HW_3
HW_4
HW_5
Predicted class

c)

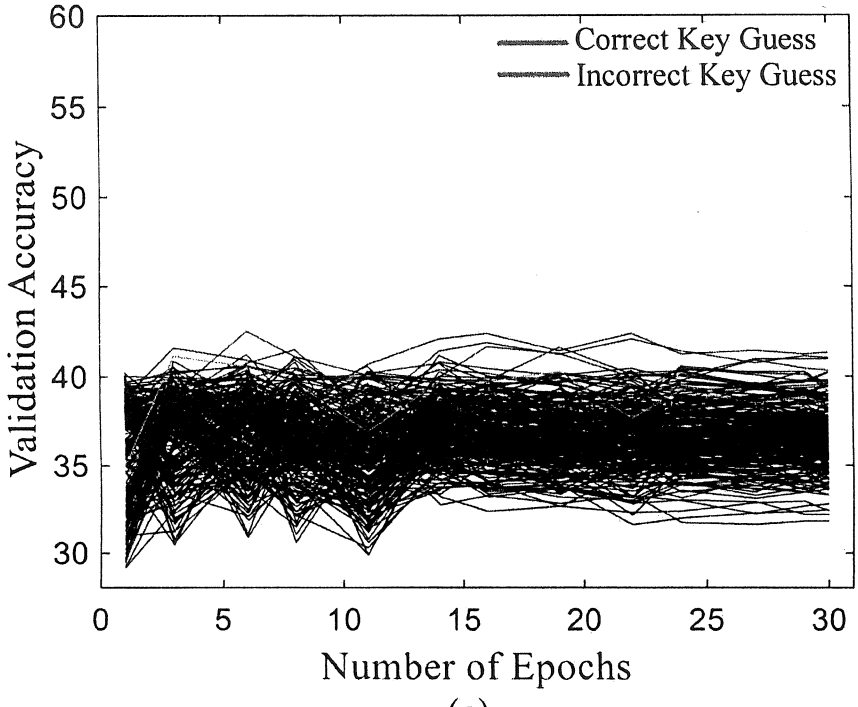
Hình 6


Correct Key Guess
Incorrect Key Guess
Validation Accuracy
60
55
50
45
40
35
30
0
5
10
15
20
25
30
Number of Epochs

(a)


Correct Key Guess
Incorrect Key Guess
Validation Accuracy
Number of Epochs
60
55
50
45
40
35
30
0
5
10
15
20
25
30

(b)


Correct Key Guess
Incorrect Key Guess
Validation Accuracy
Number of Epochs
60
55
50
45
40
35
30
0
5
10
15
20
25
30

(c)

Hình 7