XÂY DỰNG MÔ HÌNH PHÁT HIỆN VÀ CẢNH BÁO CÁC GIAO DỊCH BẤT THƯỜNG TRÊN VÍ ĐIỆN TỬ

Nguyễn Quốc Chí, Lương Trần Ngọc Khiết, Lương Trần Hy Hiến, Trần Hoàng Đạt, Hoàng Tấn Dũng, Đoàn Quang Thiệu, Lê Minh Triết

Khoa Công nghệ thông tin, Trường Đại học Sư Phạm TP.HCM

4901104014@student.hcmue.edu.vn, khietltn@hcmue.edu.vn, hienlth@hcmue.edu.vn, 4901104030@student.hcmue.edu.vn, 4901104018@student.hcmue.edu.vn, 4901104142@student.hcmue.edu.vn, trietlm@hcmue.edu.vn

TÓM TẮT— Với sự phát triển mạnh mẽ của ví điện tử, các rủi ro về an toàn giao dịch trực tuyến cũng gia tăng nhanh chóng, đặc biệt là các mối đe dọa kép: từ nguy cơ chiếm quyền kiểm soát thiết bị đến gian lận tài chính. Do đó nghiên cứu này đề xuất một kiến trúc bảo mật hai lớp toàn diện nhằm cung cấp giải pháp phát hiện và cảnh báo giao dịch bất thường có tính đa tầng. Lớp thứ nhất (phát hiện chiếm quyền thiết bị) thực hiện các kiểm tra tĩnh (phát hiện thiết bị đã bị bẻ khóa - rooting, ứng dụng độc hại) và phân tích ngưỡng động học thao tác người dùng (ví dụ: tốc độ vuốt, độ rung) để đảm bảo tính toàn vẹn của môi trường giao dịch và ngăn chặn chiếm quyền kiểm soát. Lớp thứ hai (phát hiện giao dịch bất thường) sử dụng thuật toán Isolation Forest không giám sát. Lớp này được huấn luyện trên bộ dữ liệu mô phỏng uy tín Paysim (6.3 triệu giao dịch) đã được công bố tại "Hội nghị chuyên đề về mô hình hóa và mô phỏng châu Âu lần thứ 28", trong đó nghiên cứu đã áp dụng kỹ thuật xây dựng đặc trưng chuyên sâu để tối ưu hóa khả năng nhận diện bất thường. Kết quả thực nghiệm cho thấy: Lớp thứ nhất đã thành công trong việc phát hiện và ngăn chặn đăng nhập khi thiết bị đã bị chiếm quyền kiểm soát trong môi trường mô phỏng. Lớp thứ hai đạt hiệu quả vượt trội với mô hình Isolation Forest đạt chỉ số ROC-AUC là 0.9160, chứng minh khả năng phân loại mạnh mẽ trong việc cô lập các giao dịch bất thường. Mặc dù hiệu suất được chứng minh trên dữ liệu mô phỏng, nhưng kiến trúc kết hợp này đã khẳng định tính khả thi của một giải pháp bảo mật đa tầng, thích ứng linh hoạt với các mối đe dọa từ cả thiết bị và hành vi gian lận, lừa đảo.

Từ khóa— Học máy, Thuật toán học máy không giám sát, Phát hiện giao dịch bất thường, Bộ dữ liệu Paysim, Giao dịch ví điện tử, Phát hiện giao dịch gian lận, Mô hình Isolation Forest

I. GIỚI THIỆU

Trong bối cảnh của cuộc Cách mạng công nghiệp 4.0, các dịch vụ tài chính số đặc biệt là ví điện tử, đã trở thành một phần không thể thiếu trong đời sống kinh tế - xã hội trên toàn thế giới. Tuy nhiên, kèm theo sự phát triển rực rỡ ấy là sự gia tăng ngày càng nhanh chóng của các hoạt động phi pháp liên quan đến giao dịch trực tuyến với quy mô ngày càng tinh vi, đa dạng hình thức và mang tính xuyên biên giới, gây ra hàng loạt các tổn thất tài chính khổng lồ, kéo theo sự mất niềm tin vào nền kinh tế, dẫn đến bất ổn và hậu quả tài chính trực tiếp cho các bên liên quan [1]. Theo tài liệu [2], gian lận được định nghĩa là "Sự lạm dụng hệ thống của một tổ chức lợi nhuận mà không nhất thiết dẫn đến hậu quả pháp lý trực tiếp" (nguyên văn tiếng Anh: fraud is defined as – *the abuse of profit organization system without necessarily leading to direct legal consequences*). Một trong những nguyên nhân chính dẫn đến các tình trạng trên là do các kỹ thuật bảo mật truyền thống đã ngày càng bộc lộ ra nhiều hạn chế trước sự tinh vi của các cuộc lừa đảo hoặc tấn công mạng [3], đồng thời các yếu tố như mất cân bằng dữ liệu, quyền riêng tư và chi phí tính toán cũng tiếp tục cản trở việc phát hiện giao dịch gian lận kịp thời [4, 5].

Dù có nhiều mô hình học máy được đề xuất [6], nghiên cứu về ví điện tử vẫn tồn tại những lỗ hổng cần giải quyết. Các nghiên cứu hiện nay chủ yếu tập trung vào thẻ tín dụng hoặc ngân hàng truyền thống, chưa đi sâu vào đặc thù của ví điện tử. Bên cạnh đó, việc phân tách giữa hành vi gian lận cố ý và các giao dịch bất thường do thói quen cá nhân chưa được làm rõ. Đặc biệt, rủi ro từ việc thiết bị người dùng bị chiếm quyền kiểm soát thường bị bỏ qua, khiến các lớp bảo mật phía sau mất đi hiệu quả tối đa [7].

Trong khi đó, ví điện tử vốn có sự khác biệt hơn về môi trường thiết bị và cơ chế xác thực, lại chưa được nghiên cứu chuyên sâu. Còn ở Việt Nam, ví điện tử ngày càng phổ biến với hàng chục triệu tài khoản đang hoạt động [8]. Mặc dù hiện nay vẫn chưa có thống kê nào được chính phủ Việt Nam chính thức công bố về số vụ lừa đảo liên quan trực tiếp đến ví điện tử, mà thường chỉ gọi chung là lừa đảo trực tuyến như giả mạo, lừa đảo đầu tư [9, 10].

Xuất phát từ thực trạng trên, mục tiêu cốt lõi của nghiên cứu này là xây dựng một kiến trúc hai lớp bảo mật toàn diện cho giao dịch ví điện tử. Kiến trúc đề xuất bao gồm: Lớp 1 (phát hiện chiếm quyền kiểm soát thiết bị), lớp này sẽ phát hiện và ngăn chặn nguy cơ chiếm quyền kiểm soát thiết bị; và Lớp 2 (phát hiện giao dịch bất thường), hoạt động ngay sau khi Lớp 1 đã đảm bảo môi trường giao dịch an toàn, lớp này sẽ tích hợp mô hình học máy để phát hiện ra bất thường trong hành vi giao dịch của người dùng để đưa ra cảnh báo.

Điểm mới của kiến trúc đề xuất nằm ở cơ chế phối hợp có điều kiện: Lớp 1 đóng vai trò kiểm soát bối cảnh , đảm bảo tính xác thực của chủ thể thao tác trước khi dữ liệu hành vi được đưa vào mô hình học máy ở Lớp 2. Sự kết hợp này giúp giảm thiểu tỷ lệ báo động giả và ngăn chặn các cuộc tấn công tinh vi mà nếu chỉ sử dụng dữ liệu giao dịch đơn thuần sẽ không thể phát hiện được.

Do đó, đóng góp chính của nghiên cứu này là thiết kế và kiểm chứng tính khả thi của mô hình bảo mật đa tầng dành riêng cho môi trường ví điện tử, giải quyết đồng thời rủi ro từ thiết bị (Lớp 1) và nguy cơ trong hành vi giao dịch (Lớp 2). Các phần tiếp theo sẽ trình bày chi tiết về phương pháp xây dựng hai lớp, quá trình xử lý bộ dữ liệu Paysim, kết quả thực nghiệm và hướng phát triển.

II. TỔNG QUAN TÀI LIỆU

A. CÁC KHÁI NIỆM XUNG QUANH VÍ ĐIỆN TỬ

1. KHÁI NIỆM VÍ ĐIỆN TỬ

Ví điện tử (*e-wallet*) là một nền tảng thanh toán di động giúp thực hiện các giao dịch không tiền mặt, tại chỗ hoặc từ xa, giữa người tiêu dùng và nhà cung cấp dịch vụ/hàng hóa [11, 12, 13]. Sự ra đời và phổ biến của ví điện tử xuất phát từ nhu cầu giao dịch tiện lợi và bảo mật trong kỷ nguyên số. Mục đích cốt lõi của giao dịch ví điện tử là cung cấp cho một một giải pháp thanh toán không tiền mặt nhanh chóng, hạn chế sự cồng kềnh của tiền mặt và thẻ vật lý. Hiểu rộng hơn, ví điện tử là một "ví tiền ảo" cho phép người dùng nạp sẵn một số tiền vào tài khoản đăng ký và sử dụng để thanh toán trực tuyến hoặc trực tiếp cho hàng hóa, dịch vụ [14]. Bên cạnh đó, ví điện tử còn cung cấp các tính năng linh hoạt như chia sẻ chi phí và thanh toán độc lập trên cùng một hóa đơn, giúp nâng cao trải nghiệm người dùng [15].

2. CÁC LOẠI GIAO DỊCH TRÊN VÍ ĐIỆN TỬ

Ví điện tử có một số loại giao dịch chính như sau:

- Nạp tiền (*CASH-IN*) và Rút tiền (*CASH-OUT*): Hai chức năng cơ bản quyết định sự luân chuyển tiền giữa hệ thống ngân hàng và ví.
- Chuyển tiền (*TRANSFER*): Cho phép người dùng gửi tiền cho nhau nhanh chóng, không cần số tài khoản ngân hàng.
- Thanh toán hóa đơn/dịch vụ (*PAYMENT*): Loại giao dịch đa dạng nhất (điện, nước, vé, mua sắm).
- Ví trả sau/ghi nợ (*DEBIT*): Hình thức tín dụng vi mô, mở rộng khả năng tiếp cận tài chính.
- Sử dụng đa tiện ích hệ sinh thái: Bao gồm các dịch vụ giá trị gia tăng như tích điểm, vay tiêu dùng nhỏ, gây quỹ từ thiện (ví dụ như Momo có "heo đất" tiết kiệm, gây quỹ từ thiện, …).

B. ĐẶC ĐIỂM DỮ LIỆU VÀ KHÁI NIỆM BẤT THƯỜNG

1. ĐẶC ĐIỂM PHÁT SINH DỮ LIỆU

Dữ liệu giao dịch được phát sinh tức thì, trở thành nguồn thông tin đầu vào quan trọng cho việc theo dõi và phát hiện bất thường. Đặc điểm của dữ liệu này bao gồm:

- Tính thời gian thực: Đòi hỏi mô hình phát hiện phải phản hồi nhanh chóng để ngăn chặn thiệt hại tài chính, như bài báo [3] đã nêu trên có đề cập.
- Tính đa dạng: Bao gồm nhiều thông tin như số tiền, loại giao dịch, thời gian, và thông tin người gửi/nhận.
- Tính liên kết: Các giao dịch luôn tạo thành một chuỗi hành vi, cho phép phân tích luân chuyển tiền.

2. KHÁI NIỆM VÀ PHÂN LOẠI BẤT THƯỜNG

Giao dịch tài chính bất thường là các hành vi giao dịch khác biệt đáng kể so với mô hình giao dịch thông thường. Sự khác biệt này có thể do nhiều nguyên nhân gây ra, chẳng hạn như gian lận, lỗi hệ thống và biến động bất thường của thị trường. Ví dụ, các giao dịch lớn đột ngột, tần suất giao dịch tăng hoặc giảm bất thường, và sự khác biệt đáng kể giữa mô hình giao dịch và hành vi lịch sử đều có thể được coi là giao dịch bất thường [16]. Tuy nhiên cũng cần phải nhấn mạnh rằng bất thường không đồng nghĩa với gian lận (vì đôi khi chỉ là sự cố đặc biệt hy hữu, ví dụ như bị bệnh đến mức phải nhập viện thì buộc phải thanh toán viện phí với lượng tiền giao dịch khá cao). Giao dịch bất thường được phân thành bốn loại sau, đòi hỏi các cơ chế xử lý khác nhau:

- Bất thường do chiếm quyền thiết bị: Là một dạng đặc biệt của giao dịch bất thường, xảy ra khi tài khoản của người dùng bị truy cập và kiểm soát bởi người khác. Các bất thường này có thể có đặc điểm giống với gian lận, nguyên nhân là do điện thoại hoặc tài khoản của người dùng bị chiếm quyền kiểm soát. Đây là loại bất thường mà Lớp 1 sẽ được triển khai để ngăn chặn.
- Bất thường do gian lận/ lừa đảo: Đây là các giao dịch bất thường được thực hiện với mục đích chiếm đoạt tài sản. Đặc điểm nổi bật là sự bất thường về hành vi và tính chất (qua thời gian, giá trị giao dịch, địa điểm, tần suất, mối quan hệ của bên gửi và nhận). Đây là loại bất thường mà Lớp 2 sẽ phải được triển khai để cảnh báo và ngăn chặn.
- Bất thường do lỗi hệ thống: Loại bất thường này không liên quan đến hành vi cố ý mà là kết quả của các sự cố kỹ thuật. Chúng có thể xảy ra trong quá trình xử lý giao dịch hoặc từ các vấn đề về cơ sở hạ tầng. Thường là do giao dịch bị lặp lại, giao dịch thất bại nhưng vẫn lưu lại, hoặc giao dịch không chính xác về thông tin. Do đó loại bất thường này cần phải được khoanh vùng riêng, sẽ được xử lý bằng các công cụ giám sát trong hệ thống quản trị phía sau (*backend*).

- Bất thường do sự cố: Do sự thay đổi đột ngột so với thói quen (du lịch, khám bệnh, …) khiến cho tính chất giao dịch sẽ trở nên khác thường, chứ không thật sự là do lừa đảo, gian lận. Cần cảnh báo và yêu cầu xác thực bằng Lớp 2, phát hiện giao dịch bất thường.

C. TỔNG QUAN VỀ PHÁT HIỆN BẤT THƯỜNG TRONG GIAO DỊCH TÀI CHÍNH

1. TỔNG QUAN VỀ CÁC PHƯƠNG PHÁP PHÁT HIỆN TRUYỀN THỐNG

Các phương pháp phát hiện bất thường truyền thống (phân tích thống kê và dựa trên quy tắc) đã thất bại về mặt chiến lược trước tốc độ thay đổi của thị trường ví điện tử và tội phạm công nghệ, khiến chúng trở nên cứng nhắc và thụ động, cụ thể:

- Thiếu khả năng thích ứng với dữ liệu biến động: Các phương pháp thống kê dựa trên giả định lỗi thời rằng dữ liệu giao dịch tuân theo phân phối ổn định. Sự thay đổi hành vi người dùng (như sự cố) ngay lập tức làm giảm mạnh độ chính xác vì các ngưỡng thống kê không còn đại diện cho thực tế [17].
- Tính chủ quan và độ trễ dựa trên quy tắc: Các mô hình dựa trên quy tắc (do chuyên gia đặt ra) mang tính chủ quan và có độ trễ cao. Quy trình xây dựng/cập nhật thủ công luôn tụt hậu so với tốc độ tiến hóa của các kịch bản gian lận mới. Hơn nữa, quy tắc không thể bao phủ hết mọi tình huống phức tạp, khiến chúng khó có hiệu quả trước các hành vi giao dịch chưa từng được biết đến [18, 19].

2. TỔNG QUAN CÁC PHƯƠNG PHÁP PHÁT HIỆN BẤT THƯỜNG TRONG GIAO DỊCH THÔNG QUA HỌC MÁY

a) Mô hình phát hiện bằng thuật toán giám sát và phân tích quy tắc

Nghiên cứu [20] đề xuất một hệ thống phát hiện lai, kết hợp giữa học máy và phân tích dựa trên quy tắc để tăng cường khả năng nhận dạng bất thường. Mô hình học máy Random Forest được các tác giả lựa chọn vì khả năng xử lý các bộ dữ liệu phức tạp, chống quá khớp (*overfitting*), và khả năng phân tích tầm quan trọng của tính năng. Ngoài ra, hệ thống của họ cũng nhấn mạnh việc xây dựng đặc trưng kỹ thuật (*feature engineering*) tinh vi và việc kết hợp các quy tắc bất thường được xác định trước cùng với dự đoán của học máy để đưa ra cảnh báo.

b) Mô hình phát hiện dựa trên chuỗi

Nghiên cứu [21] chứng minh rằng mô hình Transformer, thông qua cơ chế tự chú ý (*self-attention*), có thể nắm bắt hiệu quả các mối quan hệ phụ thuộc tầm xa trong chuỗi dữ liệu giao dịch. Mô hình này đã đạt hiệu suất rất tốt (F1-Score đạt 97.7%), vượt trội hơn so với hai mô hình Autoencoder và Isolation Forest. Tuy nhiên ở nghiên cứu này các tác giả đã không cung cấp chỉ số AUC-ROC, nên nhóm chúng tôi sẽ không dùng bài báo này để so sánh ở các phần sau.

c) Lý do nhóm lựa chọn mô hình Isolation Forest

Mặc dù các mô hình phức tạp như Transformer đạt hiệu suất tốt hơn, nhưng chúng thường đòi hỏi tài nguyên tính toán lớn và độ phức tạp triển khai cao (đặc biệt trong môi trường ví điện tử cần phản hồi nhanh). Do đó, nghiên cứu này chọn Isolation Forest làm nền tảng cho Lớp 2 (phát hiện giao dịch bất thường) vì đây là một mô hình học máy không giám sát hiệu quả, đã được chứng minh đạt AUC cao nhất (0.9168) so với các mô hình phát hiện bất thường khác (OCSVM, LOF, K-Means) trên bộ dữ liệu "Credit Card Fraud Detection" (sẽ gọi tắt là bộ dữ liệu CCFD) cung cấp bởi European Cardholders, đã được đề cập bởi nghiên cứu [22]. Ngoài ra, hiệu quả tính toán cũng là một lợi thế, mô hình Isolation Forest hoạt động nhanh và tiết kiệm tài nguyên, là giải pháp cân bằng giữa hiệu suất và chi phí trong các hệ thống theo dõi thời gian thực.

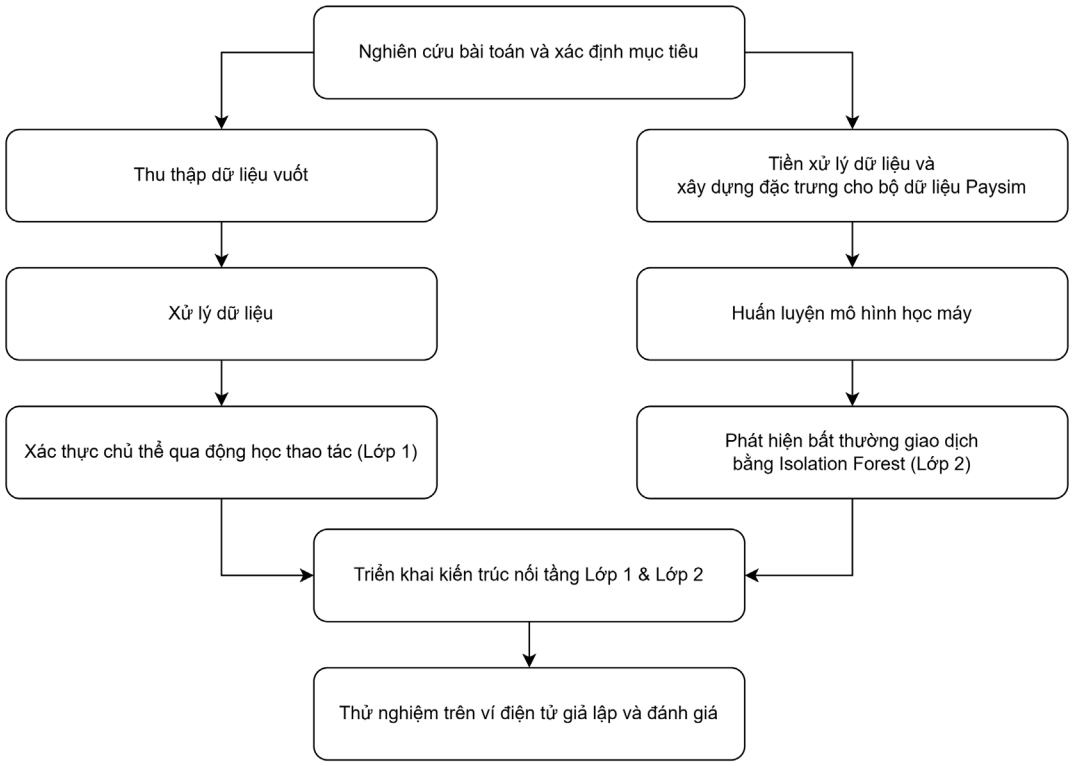
3. HẠN CHẾ CHUNG CỦA CÁC MÔ HÌNH HỌC MÁY VÀ SỰ CẦN THIẾT CỦA KIẾN TRÚC BẢO MẬT 2 LỚP

Tuy nhiên các mô hình (Isolation Forest hay Random Forest/Transformer) nêu trên đều có chung một lỗ hổng: chúng chỉ phát hiện bất thường dựa trên dữ liệu giao dịch đã được ghi nhận và không thể tự bảo vệ khỏi rủi ro chiếm quyền kiểm soát thiết bị, vốn là một lỗ hổng bảo mật cấp thiết (như [7] đã đề cập trước đó).

Do vậy, sự cần thiết phải giải quyết đồng thời rủi ro từ thiết bị và rủi ro hành vi giao dịch đã chứng minh tính cấp thiết cho kiến trúc bảo mật hai lớp đã đề xuất. Kết hợp Lớp 1 để ngăn chặn chiếm quyền thiết bị, tạo ra môi trường tương đối an toàn cho việc thực hiện giao dịch. Tiếp đó Lớp 2 sẽ được triển khai, sử dụng Isolation Forest được tối ưu hóa để phát hiện giao dịch bất thường.

III. PHƯƠNG PHÁP ĐỀ XUẤT VÀ TRIỂN KHAI THỰC NGHIỆM

A. SƠ LƯỢC VỀ QUY TRÌNH NGHIÊN CỨU



Hình 1. *Sơ đồ quy trình nghiên cứu*

Quy trình nghiên cứu được triển khai theo lộ trình có cấu trúc như *Hình 1* biểu diễn. Giai đoạn khởi đầu tập trung vào việc nghiên cứu lý thuyết, xác định mục tiêu và phạm vi bài toán phát hiện giao dịch bất thường trên ví điện tử. Đây là nền tảng cốt lõi giúp định hướng tính khoa học và khả năng ứng dụng thực tiễn của đề tài. Tiếp theo, quá trình thực hiện được phân tách thành hai luồng xử lý dữ liệu độc lập trước khi hội tụ tại bước tích hợp hệ thống:

Luồng xử lý Lớp 1 (Phát hiện chiếm quyền thiết bị): Tập trung vào việc thu thập và phân tích dữ liệu động học hành vi của người dùng (tọa độ, vận tốc, gia tốc vuốt màn hình). Thông qua việc xử lý dữ liệu và thiết lập các ngưỡng an toàn, nghiên cứu xây dựng lớp bảo mật thứ nhất nhằm ngăn chặn các nguy cơ chiếm quyền điều khiển thiết bị từ sớm.

Luồng xử lý Lớp 2 (Phát hiện giao dịch bất thường): Do tính bảo mật của dữ liệu tài chính thực tế, nghiên cứu sử dụng bộ dữ liệu mô phỏng Paysim [23], một bộ dữ liệu uy tín được công nhận tại "Hội nghị chuyên đề về mô hình hóa và mô phỏng châu Âu lần thứ 28", với các thông số thống kê trong giao dịch tương tự với dữ liệu thực tế ở một nước ở Châu Phi [24]. Dữ liệu sau khi được tiền xử lý và xây dựng đặc trưng sẽ được dùng để huấn luyện mô hình học máy không giám sát Isolation Forest. Hiệu suất của mô hình được đánh giá qua chỉ số ROC-AUC để đảm bảo khả năng cô lập các giao dịch bất thường hiệu quả nhất.

Giai đoạn quan trọng nhất là việc triển khai kiến trúc nối tầng, tích hợp đồng thời cả hai lớp bảo mật vào ứng dụng ví điện tử giả lập. Sự liên kết này cho phép hệ thống vận hành theo quy trình tuần tự: Lớp 1 đảm bảo tính toàn vẹn của chủ thể thao tác, tạo môi trường an toàn để Lớp 2 phân tích nội dung giao dịch. Cuối cùng, hệ thống được thử nghiệm tổng thể trong môi trường mô phỏng để kiểm chứng tính khả thi và điều chỉnh các tham số thực vận hành.

B. XÂY DỰNG LỚP THỨ NHẤT (LỚP PHÒNG THỦ TRƯỚC GIAO DỊCH)

1. KIỂM TRA TÍNH TOÀN VẸN CỦA THIẾT BỊ

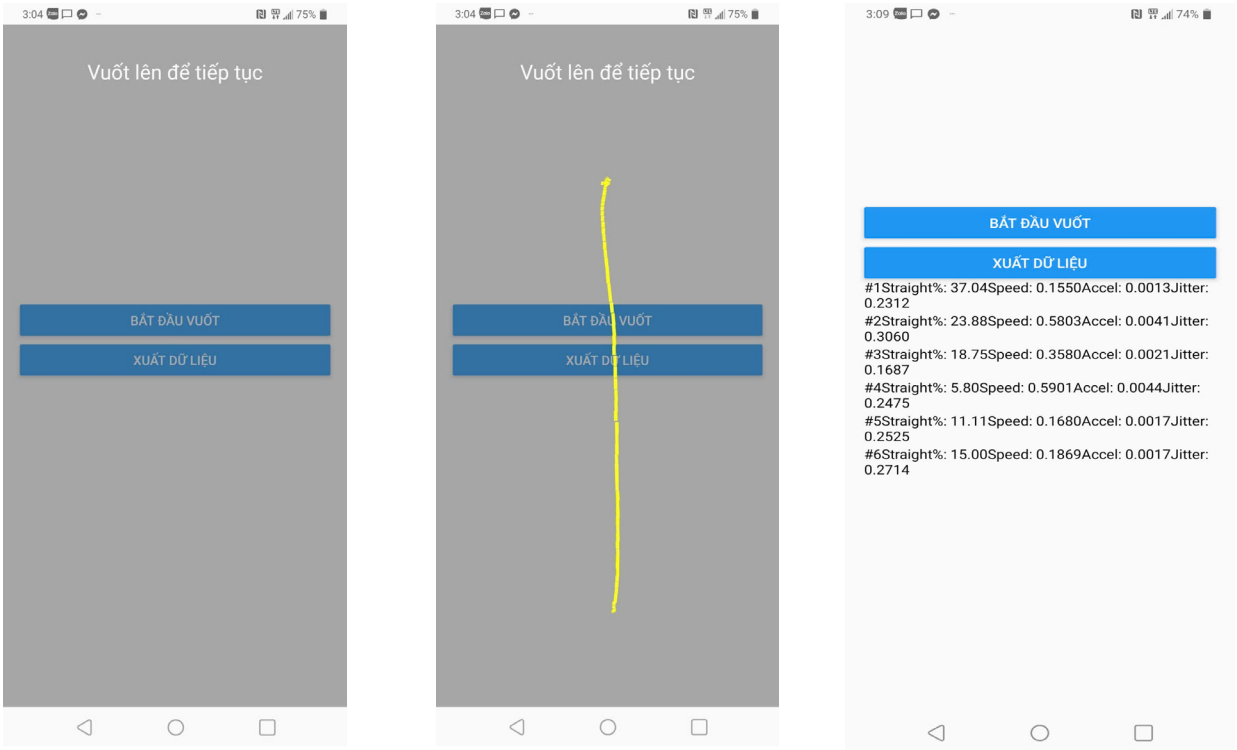
Lớp này được thiết kế để phát hiện các rủi ro bảo mật cơ bản nhưng nguy hiểm, có khả năng vô hiệu hóa hệ điều hành và các lớp bảo mật phía sau:

- Phát hiện thiết bị đã bị bẻ khóa (*root*) và mã độc: Cơ chế kiểm tra sự tồn tại của các tập tin và thư mục đặc trưng, cảnh báo về nguy cơ bị bẻ khóa hoặc can thiệp vào hệ điều hành.
- Phát hiện ứng dụng bất thường: Sử dụng cơ chế truy vấn nguồn cài đặt (*installer package*) để xác định các ứng dụng cài đặt từ nguồn không chính thống (ngoài Google Play), đồng thời duy trì một cơ sở dữ liệu (*whitelist*) để tăng độ chính xác.

- Phát hiện công cụ điều khiển từ xa: Xây dựng danh sách các gói ứng dụng (*package name*) phổ biến (như AnyDesk, TeamViewer) và phân tích trạng thái hoạt động của chúng trong thời gian thực, ngăn chặn đăng nhập hoặc thực hiện giao dịch nếu phát hiện có nguy cơ bị kiểm soát thiết bị.

2. THU THẬP VÀ KIỂM TRA THAO TÁC VUỐT MÀN HÌNH

Để phân biệt hành vi người dùng thật với thao tác của bot hoặc kẻ tấn công bằng cách điều khiển từ xa (vốn có đặc điểm tuyến tính và ổn định) với ý tưởng dựa trên các bài toán Captcha [25], do đó nhóm đã phát triển một bộ ngưỡng động học dựa trên phân tích hành vi vuốt màn hình của người dùng thực nghiệm.



(a) Ứng dụng thu thập dữ liệu vuốt (b) Thao tác vuốt để lại các điểm vàng (c) Dữ liệu thu thập được từ các điểm vàng

Hình 2. *Quá trình thu thập dữ liệu vuốt tay phục vụ cho Lớp 1*

Cụ thể, nhóm đã thu thập dữ vuốt màn hình từ 50 người dùng bằng cách viết ra một ứng dụng đơn giản để lưu lại các thông số chi tiết về tọa độ (như *Hình 2* đã trình bày), sử dụng biểu thức tập hợp điểm $\{(x_i, y_i, t_i)\}_{i=1}^{n}$ để tính toán. Phương pháp này đã được chứng minh hiệu quả trong các nghiên cứu về nhận diện viết tay điện tử [26, 27], và được nhóm cho là phù hợp để ghi nhận đặc tính động học của thao tác vuốt màn hình. Trong đó, x_i và y_i là tọa độ của ngón tay, còn t_i là dấu thời gian (*timestamp*), rồi từ đó tính ra các chỉ số như Vận tốc trung bình, Gia tốc trung bình, Tỉ lệ thẳng và Tốc độ rung lắc, như *Bảng 1* trình bày:

Bảng 1. *Tính toán các giá trị cho toàn bộ tập mẫu*

	Trung bình	**Độ lệch chuẩn**	**Lớn nhất**	**Nhỏ nhất**	**Trung vị**
Tỉ lệ thẳng (%)	23.1118	10.2093	66.6666	5.4054	21.4285
Tốc độ trung bình (px/ms)	1.4726	0.4409	3.2248	0.5102	1.4538
Gia tốc trung bình (px/ms^2)	0.0273	0.0124	0.0736	0.0068	0.0255
Tốc độ rung lắc (px/ms)	0.7515	0.2938	1.7387	0.1419	0.7383

Dựa trên các chỉ số đó, nghiên cứu áp dụng phương pháp thống kê *Mean*$\pm 2\sigma$, để thiết lập một bộ ngưỡng động học nhằm phân loại thao tác vuốt hợp lệ và giả mạo, kết quả thiết lập ngưỡng hợp lệ sẽ được tính bằng pixel trên mili giây, trình bày bởi *Bảng 2* sau:

Bảng 2. *Bảng xác định khoảng ngưỡng hợp lệ cho lớp bảo mật thứ nhất*

Các chỉ số	Khoảng ngưỡng hợp lệ
Vận tốc trung bình	[0.59; 2.35] px/ms
Gia tốc trung bình	[0.003; 0.051] px/ms^2
Tỷ lệ thẳng	[2.71; 43.51] %
Độ rung lắc	[0.17; 1.33] px/ms

Các thao tác vuốt nằm ngoài các khoảng trên, hoặc có sự kết hợp bất thường (ví dụ: vận tốc quá đều, tỷ lệ thẳng gần như tuyệt đối), sẽ được coi là giả mạo. Đây là cơ sở quan trọng để phân biệt thao tác thật của người dùng và thao tác được tạo ra bởi bot hoặc phần mềm điều khiển từ xa.

Việc hoàn tất xây dựng Lớp bảo mật 1 (Phát hiện chiếm quyền thiết bị) đã tạo ra một nền tảng môi trường tương đối an toàn cho giao dịch. Tiếp theo, nhóm nghiên cứu sẽ chuyển sang phân tích cấu trúc, đặc trưng và mối quan hệ của bộ dữ liệu Paysim, làm cơ sở để xây dựng Lớp 2: Mô hình học máy phát hiện giao dịch bất thường.

C. BỘ DỮ LIỆU ĐỂ SỬ DỤNG CHO LỚP PHÁT HIỆN GIAO DỊCH BẤT THƯỜNG

1. CẤU TRÚC CỦA BỘ DỮ LIỆU MÔ PHỎNG UY TÍN PAYSIM

Bộ dữ liệu Paysim là một tập dữ liệu tổng hợp được tạo ra nhằm mô phỏng các giao dịch tài chính của dịch vụ tiền di động. Bộ dữ liệu này được phát triển để phục vụ nghiên cứu trong lĩnh vực phát hiện gian lận tài chính, vốn xuất phát từ thực tiễn các nhà nghiên cứu hay gặp khó khăn trong việc tiếp cận dữ liệu thật do tính riêng tư và bảo mật cao của các giao dịch tiền tệ. Mặc dù bài báo gốc của nhóm tác giả Paysim báo cáo bộ dữ liệu mô phỏng khoảng 23 triệu bản ghi, nhưng bộ dữ liệu mà nhóm chúng tôi sử dụng của họ trên Kaggle chỉ chứa 6.362.621 bản ghi dữ liệu. Sự khác biệt này xuất phát từ mục tiêu và phiên bản mô phỏng khác nhau: bài báo dùng một bộ dữ liệu lớn để đánh giá hiệu quả mô phỏng và thuật toán phát hiện gian lận, trong khi phiên bản Kaggle là dữ liệu công khai được chia sẻ cho cộng đồng nghiên cứu, với kích thước nhỏ hơn nhưng vẫn giữ nguyên đặc trưng hành vi giao dịch và gian lận.

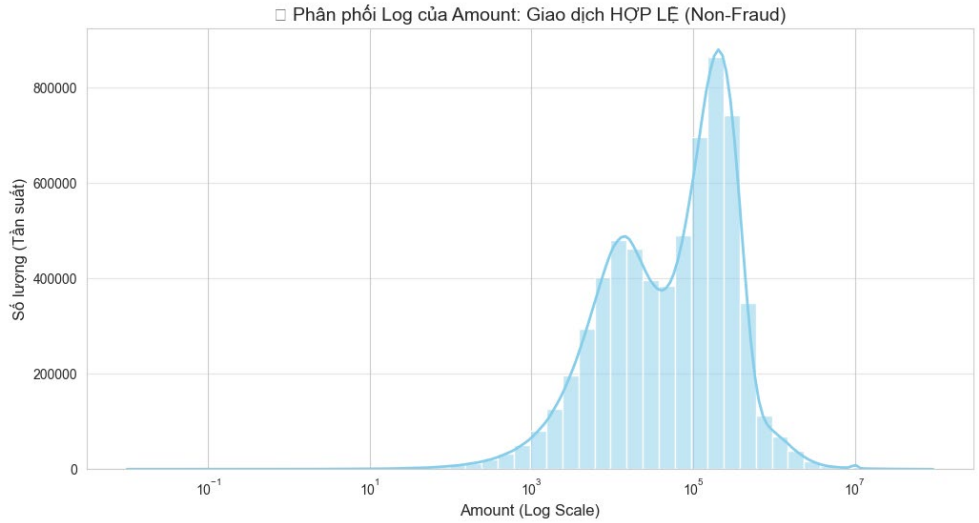
Bộ dữ liệu này được nhóm lựa chọn cho nghiên cứu ví điện tử vì các lý do như sau: Thứ nhất, việc tiếp cận dữ liệu giao dịch thật của ví điện tử là rất hạn chế do tính nhạy cảm, bảo mật cao và các quy định pháp lý nghiêm ngặt, nhóm chúng tôi vẫn chưa thật sự đủ khả năng để tiếp cận được. Thứ hai, Paysim đã cung cấp các loại giao dịch cơ bản (nạp tiền, rút tiền, chuyển tiền, thanh toán, ghi nợ) và đặc trưng hành vi bất thường về tần suất, khối lượng, cũng như mô hình chuyển tiền, có tính tương đồng cao với hệ sinh thái ví điện tử, cho phép nghiên cứu và huấn luyện các mô hình phát hiện giao dịch bất thường mà không vi phạm quyền riêng tư của người dùng. Tuy vẫn còn một hạn chế là thiếu kiểu giao dịch *đa tiện ích* mà nhóm đã phân tích ở phần tổng quan tài liệu.

Về cấu trúc dữ liệu, bộ dữ liệu Paysim gồm các thông tin chính như:

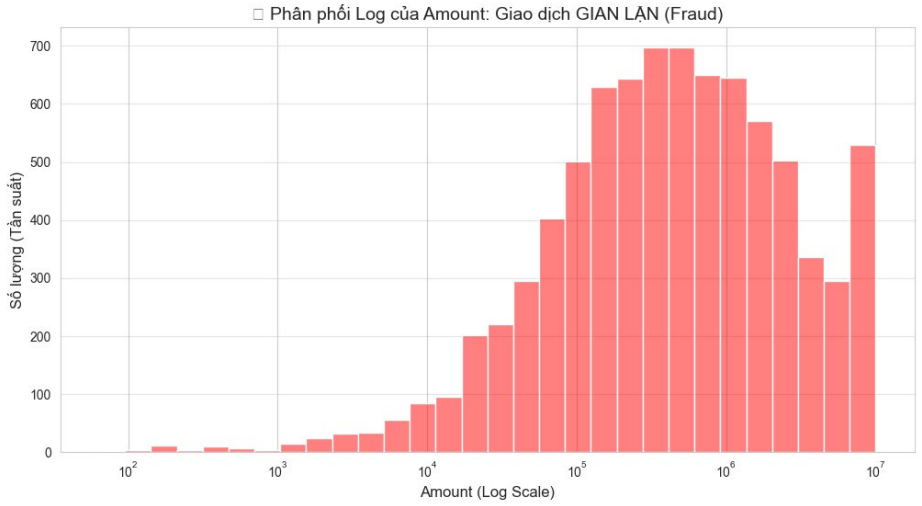
Bảng 3. *Bảng mô tả cấu trúc cơ bản của bộ dữ liệu*

Cột dữ liệu	Kiểu dữ liệu	Ý nghĩa
step	Số	Đơn vị thời gian mô phỏng (1 step = 1 giờ)
type	Loại	Loại giao dịch: CASH-OUT, CASH-IN, PAYMENT, TRANSFER, DEBIT. Tương ứng lần lượt là: Rút tiền, Nạp tiền, Thanh toán, Chuyển tiền, Ghi nợ.
amount	Số	Lượng tiền giao dịch
nameOrig/Dest	Định danh	Tài khoản người gửi/nhận
oldbalanceOrg/Dest	Số	Số dư trước của người gửi/nhận
newbalanceOrig/Dest	Số	Số dư sau của người gửi/nhận
isFraud	Nhị phân	Đánh dấu giao dịch gian lận
isFlaggedFraud	Nhị phân	Đánh dấu giao dịch vượt hạn mức cảnh báo

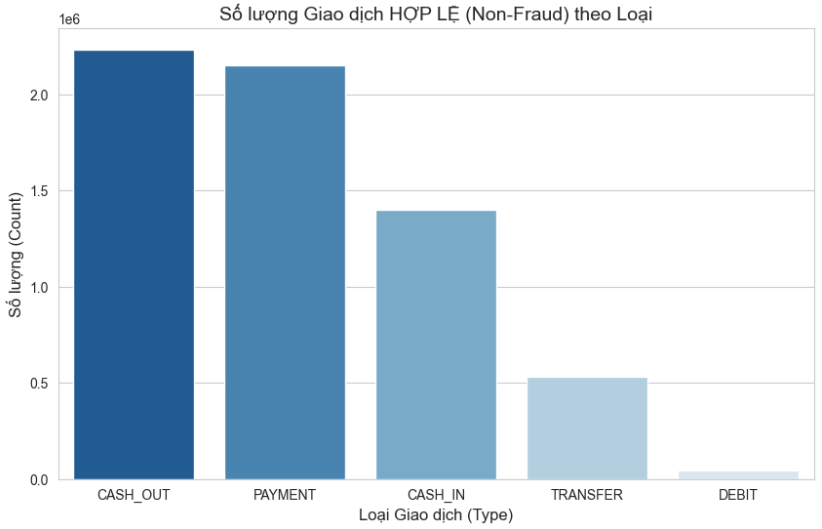
Nhờ cấu trúc ở *Bảng 3* trên, dữ liệu Paysim tương đồng với hoạt động thực tế của ví điện tử mà nhóm đã phân tích chương tổng quan tài liệu, thể hiện được mối quan hệ giữa các tài khoản, loại hình giao dịch và thời gian. Điều này giúp mô hình học máy học được các chuỗi hành vi tài chính và phát hiện các giao dịch bất thường một cách hiệu quả, trong khi vẫn giữ nguyên tính bảo mật và khả năng nghiên cứu công khai.



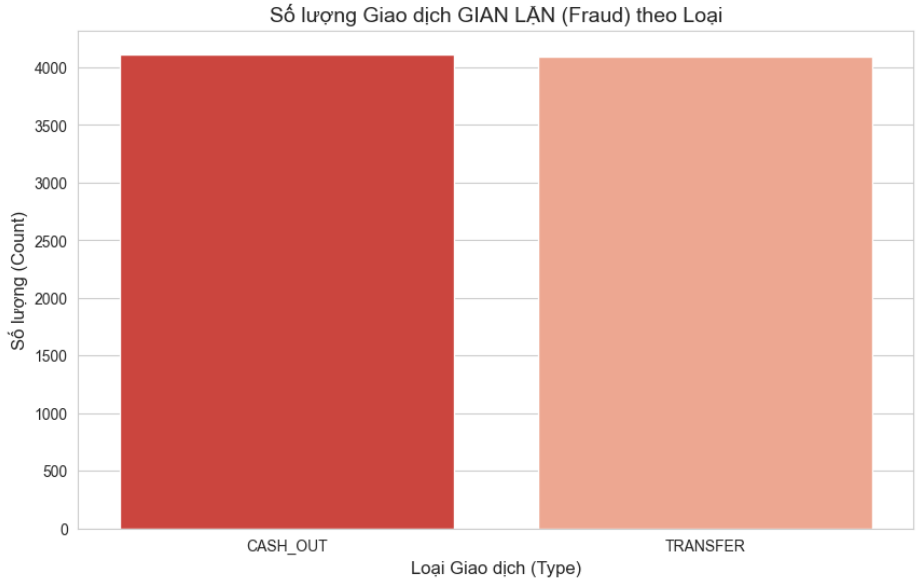
Hình 3. *Biểu đồ phân phối hàm log của số tiền giao dịch hợp lệ*



Hình 4. *Biểu đồ phân phối hàm log của số tiền giao dịch gian lận*



Hình 5. *Biểu đồ phân phối loại giao dịch hợp lệ*



Hình 6. *Biểu đồ phân phối loại giao dịch gian lận*

Sau khi trực quan hóa bộ dữ liệu Paysim *ở Hình 3, Hình 4, Hình 5, Hình 6*, chúng ta đã xác định được một số thông tin quan trọng trong bộ dữ liệu và xu hướng gian lận. Trong tổng số hơn 6 triệu giao dịch chỉ có 0.1291% là gian lận, một bộ dữ liệu mất cân bằng nghiêm trọng, với các loại giao dịch hợp lệ chủ yếu là Rút tiền (CASH-OUT) và Thanh toán (PAYMENT).

Tuy nhiên, khi chuyển sang phân tích các dữ liệu gian lận, bộ dữ liệu chỉ ra rằng các hoạt động bất thường tập trung duy nhất vào hai loại: Rút tiền (CASH-OUT) và Chuyển tiền (TRANSFER). Điều đáng chú ý là dù Chuyển tiền (TRANSFER) chỉ chiếm một phần nhỏ trong tổng số giao dịch hợp lệ (nhỏ thứ nhì chỉ sau Ghi nợ - DEBIT), nhưng lại có số lượng gian lận cao gần ngang bằng Rút tiền (CASH-OUT), cho thấy nguy cơ tiềm ẩn cao trong loại giao dịch này. Việc tập trung vào hai loại giao dịch Thanh toán (TRANSFER) và Rút tiền (CASH-OUT) này là cần thiết vì chúng mô hình hóa hành vi rửa tiền, gian lận hoặc lừa đảo để trục lợi bất chính, phù hợp với động cơ của tội phạm tài chính [28].

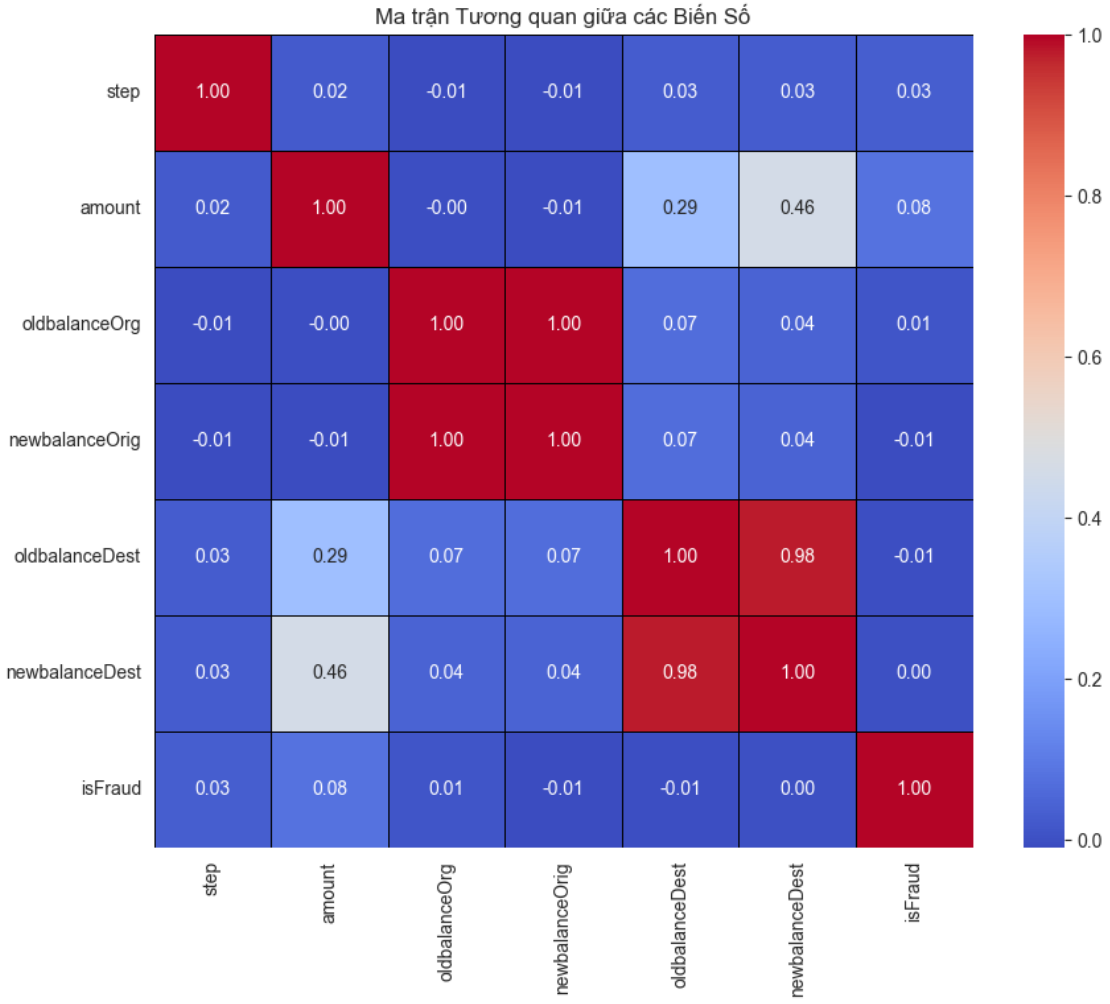
Nhưng chúng ta cần nhìn nhận một cách khách quan những đặc điểm và hạn chế vốn có của bộ dữ liệu mô phỏng này. Việc gian lận chỉ tập trung vào hai loại giao dịch Rút tiền (CASH-OUT) và Thanh toán (TRANSFER) là một sự đơn giản hóa, chưa phản ánh đầy đủ tính đa dạng và tinh vi của các chiến thuật gian lận trong môi trường tài chính toàn cầu. Mặc dù vậy, chúng ta cũng cần thông cảm rằng các tác giả của Paysim đã nỗ lực mô phỏng dựa trên dữ liệu thực tế tại một quốc gia Châu Phi, điều này tạo ra một mức độ thực tế nhất định trong giới hạn của bộ dữ liệu mô phỏng.

Thêm vào đó, với tỷ lệ gian lận cực kỳ thấp (chỉ khoảng 0.1291%), bộ dữ liệu đặt ra một thử thách đáng kể về mất cân bằng lớp, điều này đòi hỏi chúng ta phải áp dụng các kỹ thuật cân bằng lớp chuyên sâu trong các bước tiếp theo. Dù còn những điểm cần khắc phục, nhóm chúng tôi vẫn coi Paysim là một nền tảng học tập thiết yếu và một điểm khởi đầu đáng tin cậy, những đặc điểm này không làm mất đi giá trị của bộ dữ liệu trong việc giúp chúng ta xác định các biến số quan trọng như *lượng tiền giao dịch* và *số dư trong tài khoản*. Từ đó xây dựng nền tảng và chuẩn bị kỹ lưỡng cho giai đoạn tiền xử lý dữ liệu chuyên sâu hơn.

2. XỬ LÝ DỮ LIỆU

Quy trình xử lý dữ liệu được thực hiện có hệ thống, bắt đầu bằng việc sắp xếp theo trình tự thời gian dựa trên biến thời gian mô phỏng theo mỗi giờ (*step*). Qua kiểm tra sơ bộ, bộ dữ liệu không ghi nhận giá trị khuyết thiếu (*null*), đảm bảo tính toàn vẹn cho các bước phân tích tiếp theo. Thách thức lớn nhất là sự mất cân bằng lớp cực đoan (0,1291% gian lận), do đó kỹ thuật trọng số lớp (*class_weight*) được áp dụng để tăng cường độ nhạy của mô hình đối với các mẫu hành vi hiếm.

Ma trận tương quan giữa các biến số chính bao gồm: lượng tiền giao dịch (*amount*), số dư tài khoản gửi trước và sau giao dịch (*oldbalanceOrg*, *newbalanceOrig*), cùng số dư tài khoản nhận trước và sau giao dịch (*oldbalanceDest*, *newbalanceDest*) được phân tích thông qua biểu đồ nhiệt tương quan (*Heatmap*) tại *Hình 7* sau đây. Kết quả này định hướng cho việc lựa chọn các đặc trưng có giá trị phân loại cao, đồng thời loại bỏ các biến có độ tương quan thừa, tối ưu hóa không gian đầu vào cho mô hình Isolation Forest.



Hình 7. *Biểu đồ ma trận nhiệt (Heatmap) để kiểm tra độ tương quan của các đặc trưng*

Tiếp đó, nhóm tiến hành kỹ thuật xây dựng đặc trưng để tạo ra các biến mới có ý nghĩa hơn: đáng chú ý nhất là đặc trưng lỗi số dư như lỗi số dư tài khoản gửi (*errorBalanceOrig*) và lỗi số dư tài khoản nhận (*errorBalanceDest*), được tính toán từ sự chênh lệch giữa lượng thay đổi số dư thực tế và giá trị giao dịch, một chỉ số mạnh mẽ cho thấy sự bất thường của giao dịch. Ngoài ra, các đặc trưng nhị phân như tài khoản gốc có bị rút cạn sau giao dịch hay không (*isOrigEmptyAfterTx*), và người nhận có phải là khách hàng hay không (*isDestCustomer*) cũng được xây dựng để nắm bắt thêm các hành vi gian lận tiềm ẩn. Đồng thời, cột phân loại loại giao dịch đã được chuyển đổi thành các cột nhị phân thông qua kỹ thuật mã hóa nhị phân (còn được gọi là kỹ thuật One-Hot Encoding), nâng cấu trúc chi tiết của toàn bộ 21 đặc trưng cùng định nghĩa kỹ thuật cụ thể cho từng cột dữ liệu được trình bày chi tiết tại *Phụ lục A* của bài báo này, và được chia thành 6 nhóm chính ở *Bảng 4* tóm tắt sau đây:

Bảng 4. *Bảng mô tả cấu trúc của bộ dữ liệu sau khi tiến hành kỹ thuật xây dựng đặc trưng*

Nhóm đặc trưng	**Các nhóm cột dữ liệu**	**Ý nghĩa**
Giá trị giao dịch	amount, step	Giá trị giao dịch, là một trong những chỉ báo chính cho hành vi bất thường.
Độ lệch số dư của kênh gửi	deltaBalanceOrig, errorBalanceOrig	Xác định sự không khớp giữa số tiền lệnh và số dư thực tế bị trừ tại nguồn.
Độ lệch số dư của kênh nhận	deltaBalanceDest, errorBalanceDest	Kiểm tra tính nhất quán giữa số tiền nhận và biến động số dư tại tài khoản đích.
Trạng thái tài khoản	oldbalanceOrg, newbalanceOrig, oldbalanceDest,	Theo dõi biến động số dư tổng quát và phát hiện hành vi rút cạn tài khoản, đặc biệt cần lưu ý: *isOrigEmptyAfterTx*.

	newbalanceDest, isOrigEmptyAfterTx	
Loại hình giao dịch	type_TRANSFER, type_CASH_OUT, type_CASH_IN, type_PAYMENT, type_DEBIT	Phân loại các hành vi giao dịch; tập trung vào TRANSFER và CASH_OUT (tức Chuyển tiền và Rút tiền, là hai loại có tỷ lệ gian lận cao nhất).
Định danh và nhãn	nameOrig, nameDest, isDestCustomer, isFraud, isFlaggedFraud	Các thông tin định danh hệ thống và nhãn phục vụ quá trình huấn luyện/đánh giá.

Cuối cùng, dữ liệu được chuẩn bị cho mô hình hóa bằng cách chọn lọc các đặc trưng cuối cùng và loại bỏ các cột định danh không cần thiết (*nameOrig*, *nameDest*, …) cùng với các cột số dư ban đầu, trước khi tiến hành chia tập dữ liệu thành tập huấn luyện và tập kiểm tra, sử dụng phương pháp chia tầng trên biến mục tiêu *isFraud* (có gian lận hay không) để bảo toàn tỷ lệ mất cân bằng lớp trong cả hai tập. Các đặc trưng mang tính định danh hoặc không đóng góp vào quy luật hành vi gian lận như mã định danh người gửi/nhận, cờ cảnh báo hệ thống cũ và các giá trị số dư ban đầu được loại bỏ để giảm nhiễu và tối ưu hóa không gian đặc trưng cho mô hình.

Bước tiền xử lý cuối cùng là chuẩn hóa dữ liệu bằng phương pháp chỉnh biên độ tối thiểu - tối đa (được gọi là MinMaxScaler). Phương pháp này đưa toàn bộ giá trị đặc trưng về khoảng đồng nhất [0, 1] theo công thức:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Trong đó, *x* là giá trị gốc của đặc trưng cần chuẩn hóa; x_{min} là giá trị nhỏ nhất của đặc trưng đó trong tập dữ liệu huấn luyện; x_{max} là giá trị lớn nhất của đặc trưng đó trong tập dữ liệu huấn luyện; Và cuối cùng, x_{norm} là giá trị mới của đặc trưng sau khi đã đưa về thang đo [0, 1].

Việc chuẩn hóa giúp cân bằng tầm ảnh hưởng của các đặc trưng có biên độ khác biệt lớn (ví dụ: số dư tài khoản có thể lên tới hàng tỷ, trong khi các biến nhị phân chỉ là 0 hoặc 1), từ đó giúp thuật toán Isolation Forest hội tụ nhanh và chính xác hơn.

D. XÂY DỰNG LỚP PHÁT HIỆN GIAO DỊCH BẤT THƯỜNG

1. HUẤN LUYỆN THUẬT TOÁN

a) Kiến trúc mô hình học máy

Bảng 5. *Bảng mô tả kiến trúc mô hình và các siêu tham số được sử dụng*

Siêu tham số	**Giá trị cấu hình**	**Mục đích và lý do lựa chọn**
n_estimators	50	Thiết lập số lượng cây quyết định trong rừng. Con số này được chọn để đạt được sự cân bằng giữa hiệu suất và chi phí tính toán.
max_samples	0.3	Tỷ lệ mẫu ngẫu nhiên từ tập huấn luyện được sử dụng để xây dựng mỗi cây, nhằm tăng tính mạnh mẽ, chống quá khớp (*overfitting*) và tăng hiệu quả xử lý dữ liệu lớn.
contamination	auto	Cho phép thuật toán tự động ước tính tỷ lệ ngoại lai (gian lận/bất thường) trong tập huấn luyện. Dù tỷ lệ thực tế thấp, mô hình vẫn tập trung vào các mẫu dễ bị cô lập.
random_state	42	Cố định giá trị này để đảm bảo khả năng tái lập lại kết quả chính xác trong mọi lần thực thi.
n_jobs	-1	Tận dụng tất cả các lõi CPU sẵn có để tối đa hóa tốc độ và hiệu suất huấn luyện mô hình.

Mô hình phát hiện giao dịch bất thường được xây dựng dựa trên thuật toán Isolation Forest, được cấu hình với các siêu tham số cụ thể như *Bảng 5* trên nhằm tối ưu hóa hiệu suất và khả năng tổng quát hóa.

b) Quy trình huấn luyện

Dữ liệu đầu vào đã được tiền xử lý và chuẩn hóa bằng phương pháp MinMaxScaler trên các đặc trưng được xây dựng ở *Bảng 4* (chi tiết hơn ở *Phụ Lục A*). Trong đó, dữ liệu được chia thành tập huấn luyện (70% - 4.453.834 mẫu) và tập kiểm tra (30% - 1.908.786 mẫu) bằng phương pháp chia tầng trên biến mục tiêu (*isFraud* – có gian lận, là biến nhị nguyên) để bảo toàn tỷ lệ lớp. Mô hình Isolation Forest được huấn luyện chỉ trên tập dữ liệu đã được chuẩn hóa.

c) Xác định bất thường

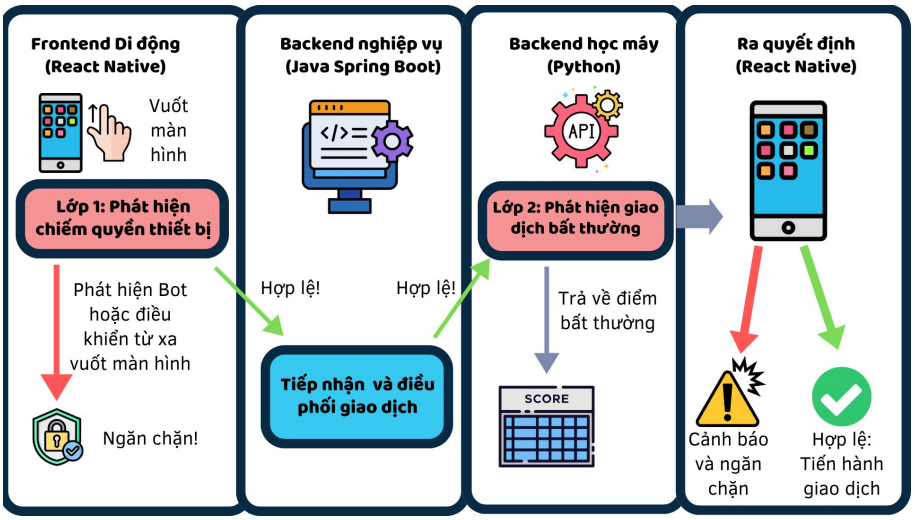
Sau khi mô hình được huấn luyện, nhóm sẽ tiến hành sử dụng hàm *decision_function* để tính toán điểm số bất thường (*anomaly_scores*) cho mỗi giao dịch trong tập kiểm tra. Trong đó, Isolation Forest (IF) gán nhãn 1 cho giao dịch bình thường (ít cô lập) và -1 cho giao dịch ngoại lai (dễ bị cô lập). Tiếp theo là chuyển đổi nhãn để phù hợp với bài toán phát hiện giao dịch bất thường (0=Bình thường, 1=Bất thường), nhãn dự đoán được chuyển đổi như sau:

- Mẫu có nhãn IF là -1 (Ngoại lai) được gán thành 1 (Bất thường).
- Mẫu có nhãn IF là 1 (Bình thường) được gán thành 0 (Bình thường).

Do đó, giao dịch được xem là bất thường khi mô hình Isolation Forest dự đoán nó là ngoại lai (nhãn -1) và được chuyển đổi thành nhãn 1 (Bất thường).

E. TÍCH HỢP KIẾN TRÚC HAI LỚP VÀO ỨNG DỤNG VÍ ĐIỆN TỬ GIẢ LẬP

Trong quá trình hiện thực hóa mô hình sử dụng kiến trúc nối tầng nghiêm ngặt thiết lập một cơ chế ngắt mạch logic nhằm đảm bảo tính toàn vẹn của dữ liệu đầu vào. Hệ thống không chỉ đơn thuần là sự kết hợp của các thành phần kỹ thuật như máy chủ nghiệp vụ (Java Spring Boot), giao diện (React Native) và máy chủ phân tích (Python), mà còn là một chỉnh thể thống nhất về mặt logic bảo mật, kiến trúc này sẽ được *Hình 8* trình bày cụ thể như sau:



Hình 8. *Kiến trúc mô hình sau khi tích hợp Giao diện & Lớp 1 & Lớp 2*

Lớp 1 (phát hiện chiếm quyền): Được triển khai trực tiếp tại thiết bị đầu cuối để tối ưu hóa thời gian phản hồi. Lớp này thực hiện kiểm tra tính toàn vẹn hệ thống (phát hiện bẻ khóa, ứng dụng xâm nhập) và trích xuất các đặc trưng động học từ thao tác vuốt (vận tốc, áp lực, độ rung). Dựa trên cảnh báo của nghiên cứu [7], nếu thiết bị đã bị chiếm quyền hoặc phát hiện hành vi giả mạo từ các ứng dụng xâm nhập, hệ thống sẽ thực thi ngắt mạch logic của ứng dụng ví điện tử ngay tại thiết bị. Việc ngăn chặn giao dịch ngay lập tức tại Lớp 1 đóng vai trò là "điều kiện cần", giúp bảo vệ mô hình học máy ở Lớp 2 khỏi dữ liệu giả mạo và giảm tải tài nguyên cho máy chủ.

Lớp 2 (phát hiện giao dịch bất thường): Chỉ khi vượt qua được bộ lọc ở Lớp 1, dữ liệu giao dịch mới được chuyển tiếp đến máy chủ phân tích bằng ngôn ngữ Python. Tại đây, thuật toán Isolation Forest sẽ thực hiện tính toán điểm rủi ro dựa trên 21 đặc trưng đã được kỹ thuật hóa (như lỗi số dư, loại giao dịch). Việc giữ sự độc lập về đặc trưng dữ liệu giữa hai lớp là chiến lược chủ động nhằm thực hiện phòng thủ chiều sâu, ngay cả khi kẻ tấn công vượt qua được lớp nhận diện hành vi, chúng vẫn phải đối mặt với một hàng rào bảo mật hoàn toàn khác biệt ở Lớp 2.

Do đó, mối liên hệ giữa hai lớp không chỉ dừng lại ở việc hoạt động tuần tự mà còn nằm ở sự hội tụ quyết định. Kết quả từ Lớp 1 cung cấp ngữ cảnh về tính chính chủ của phiên làm việc, tạo ra môi trường an toàn cho Lớp 2 đánh giá mức độ bất thường của dòng tiền. Nếu điểm rủi ro vượt ngưỡng cho phép, hệ thống sẽ kích hoạt quy trình xác thực bổ sung hoặc tạm dừng giao dịch dựa trên logic bảo mật phi tập trung. Kiến trúc này cho phép tối ưu hóa hiệu

suất thực thi đồng thời đảm bảo tính linh hoạt khi nâng cấp các mô hình học máy mà không làm gián đoạn luồng nghiệp vụ cốt lõi.

IV. KẾT QUẢ THỰC NGHIỆM

A. ĐÁNH GIÁ HIỆU QUẢ CỦA LỚP PHÁT HIỆN CHIẾM QUYỀN KIỂM SOÁT THIẾT BỊ

Sau khi hoàn thành việc triển khai Lớp 1 (phát hiện chiếm quyền kiểm soát thiết bị), nhóm đã tiến hành kiểm thử các tính năng của lớp, bằng cách thử bẻ khóa (*root*) điện thoại, cài đặt các ứng dụng không trong quyền kiểm soát của GooglePlay hoặc các phần mềm điều khiển từ xa như AnyDesk, TeamViewer, AirDroid hay Chrome Remote. Kết quả cho thấy lớp bảo vệ này đã thành công trong việc kiểm tra, phát hiện nguy cơ và ngăn chặn đăng nhập vào tài khoản ví điện tử khi thiết bị đã bị chiếm quyền kiểm soát.

Ngoài ra, sau khi đã đăng nhập thành công và tiến hành giao dịch, mô hình sẽ yêu cầu vuốt màn hình lần nữa theo hướng thẳng đứng từ dưới lên. Lớp bảo mật này nhanh chóng dựa vào tọa độ của các điểm vuốt để tính toán các chỉ số hành vi (như Vận tốc trung bình, Gia tốc trung bình, Tỷ lệ thẳng, Độ rung lắc), từ đó quyết định cho phép giao dịch tiếp tục hay không. Tuy nhiên, cơ chế quyết định hiện tại vẫn đang sử dụng các ngưỡng giá trị cứng để phân biệt thao tác của người dùng thật với thiết bị điều khiển từ xa hoặc bot.

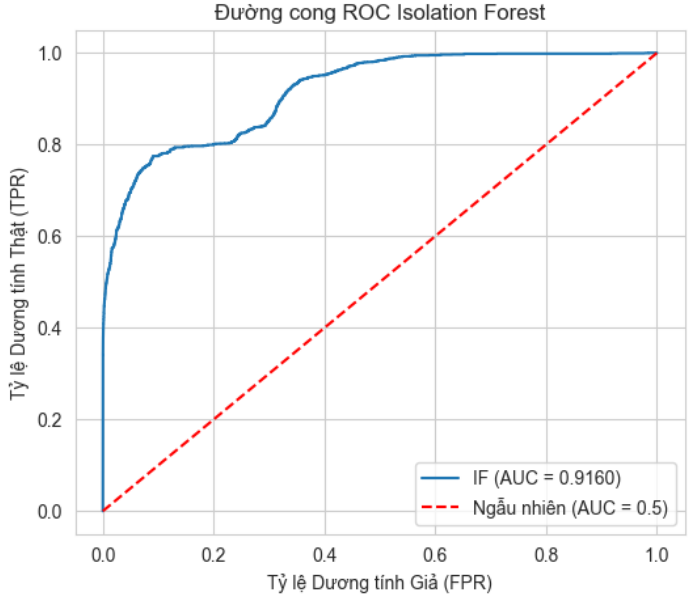
Kết quả thực nghiệm cho thấy Lớp 1 có khả năng hoạt động ổn định trong môi trường mô phỏng, giúp giảm đáng kể nguy cơ bị chiếm quyền điều khiển thiết bị. Tuy nhiên, do dữ liệu huấn luyện chủ yếu được giả lập từ mẫu hành vi người dùng hạn chế nên hiệu suất có thể chưa phản ánh đầy đủ được các tình huống thực tế. Hạn chế lớn nhất là sự kém linh hoạt của ngưỡng cứng.

Trong các bước tiếp theo, nhóm sẽ mở rộng nguồn dữ liệu, bổ sung thêm nhiều mẫu thao tác từ thiết bị thật, và chuyển sang áp dụng phương pháp học máy tiên tiến để tự động và thông minh hóa việc đo lường các chỉ số ngưỡng, tương tự như cách các mô hình học máy đã được dùng để giải quyết bài toán Captcha phức tạp với phương pháp sẽ tham khảo từ nghiên cứu [29].

Nhìn chung, Lớp 1 đã hoàn thành tốt vai trò tiền phong, tạo môi trường an toàn cho việc phát hiện giao dịch bất thường ở Lớp 2.

B. ĐÁNH GIÁ HIỆU QUẢ PHÁT HIỆN BẤT THƯỜNG TRONG GIAO DỊCH CỦA LỚP HAI

Sau khi Lớp 1 thành công trong việc tạo ra một môi trường giao dịch an toàn và đáng tin cậy ở cấp độ thiết bị, chúng ta chuyển sang đánh giá hiệu suất của Lớp 2 (phát hiện giao dịch bất thường) trên bộ dữ liệu Paysim. Lớp này sử dụng mô hình Isolation Forest và được đánh giá bằng chỉ số ROC-AUC, một tiêu chí quan trọng để đo lường khả năng phân loại và cô lập hiệu quả các giao dịch bất thường khỏi hành vi hợp lệ.



Hình 9. *Biểu đồ AUC-ROC trực quan cho khả năng phát hiện và phân loại giao dịch gian lận của Isolation Forest*

Việc đánh giá mô hình phát hiện giao dịch bất thường Isolation Forest (IF) trên bộ dữ liệu mô phỏng PaySim đã cho thấy kết quả khả quan. Chỉ số AUC-ROC đạt 0.9160 (như *Hình 9* đã trình bày) chứng minh khả năng phân biệt vượt trội của mô hình, đặc biệt quan trọng trong môi trường mất cân bằng lớp cực đoan của dữ liệu tài chính.

Bảng 6. *Bảng so sánh kết quả của nghiên cứu khác với phương pháp đề xuất*

	Cấu hình 1 của tác giả Gillespie	**Cấu hình 2 của tác giả Gillespie**	**Cấu hình 3 của tác giả Gillespie**	**Cấu hình đề xuất của chúng tôi**	**Cấu hình của tác giả S. Ounacer**	**Cấu hình của tác giả S. Ounacer**	**Cấu hình của tác giả S. Ounacer**
Bộ dữ liệu	Paysim	Paysim	Paysim	Paysim	CCFD	CCFD	CCFD
Số dòng dữ liệu và tỉ lệ gian lận	6.362.621 (0.1291%)	6.362.621 (0.1291%)	6.362.621 (0.1291%)	6.362.621 (0.1291%)	284,807 (0.172%)	284,807 (0.172%)	284,807 (0.172%)
Mô hình	Isolation Forest	Isolation Forest	Isolation Forest	Isolation Forest	Isolation Forest	OCSVM	K-Means
Tập thử nghiệm	20%	20%	20%	30%	30%	30%	30%
Chỉ số AUC-ROC	0.830 (Với 1000 giao dịch bất thường nhất)	0.845 (Với 1000 giao dịch bất thường nhất)	0.784 (Với 1000 giao dịch bất thường nhất)	0.9160 (Trên toàn bộ giao dịch)	0.9168 (Trên toàn bộ giao dịch)	0.5154 (Trên toàn bộ giao dịch)	0.5191 (Trên toàn bộ giao dịch)

Dựa trên kết quả tổng hợp tại *Bảng 6*, hiệu năng của mô hình đề xuất thể hiện ưu thế rõ rệt khi đặt trong tương quan với các nghiên cứu sử dụng cùng bộ dữ liệu PaySim và các mô hình cơ sở khác.

So với nghiên cứu của Gillespie [30] sử dụng cùng tập PaySim (6,3 triệu dòng), mô hình của chúng tôi đạt chỉ số AUC-ROC là 0.9160, vượt trội đáng kể so với các mốc cấu hình của tác giả này (0.830; 0.845; 0.784). Đáng chú ý, kết quả này đạt được dù chúng tôi chỉ sử dụng 70% dữ liệu để huấn luyện (so với 80% của Gillespie), việc đạt được AUC cao hơn với lượng dữ liệu huấn luyện ít hơn là minh chứng cho sự tối ưu hóa thành công các siêu tham số được nhóm đề xuất ở *Bảng 5*, giúp mô hình có khả năng tổng quát hóa tốt hơn trên tập dữ liệu lớn.

Đồng thời, kết quả AUC 0.9160 của nhóm cũng tiệm cận rất gần với mốc 0.9168 của tác giả S. Ounacer [22] trên bộ dữ liệu CCFD. Ngoài ra, khi đối chiếu với mô hình OCSVM và K-Means của S. Ounacer vốn chỉ đạt lần lượt là AUC-ROC 0.5154 và 0.5191 (tương đương phân loại ngẫu nhiên), cho thấy các thuật toán cơ sở truyền thống thường thất bại trong việc nhận diện gian lận trên dữ liệu mất cân bằng lớp cực đoan.

Tuy nhiên, cũng cần phải thừa nhận rằng tập CCFD là bộ dữ liệu thực tế dù có quy mô nhỏ (284.807 dòng), còn PaySim là bộ dữ liệu mô phỏng với các thông số thống kê tương tự thực tế, có quy mô lớn gấp hơn 22 lần (6.362.621 dòng). Việc duy trì độ chính xác cao trên một tập dữ liệu khổng lồ chứng minh khả năng mở rộng vượt trội của các tham số đề xuất cho mô hình Isolation Forest.

Do đó, kết quả thực nghiệm xác nhận mô hình đề xuất phản ứng nhạy bén với các giao dịch lệch pha (như đột biến giá trị giao dịch vượt ngưỡng trung bình), từ đó kích hoạt chính xác quy trình bảo mật bổ sung trên ứng dụng, giúp bảo vệ người dùng trước các rủi ro gian lận thực tế.

V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

A. KẾT LUẬN

Nghiên cứu này trình bày một nỗ lực thử nghiệm tính khả thi nhằm thiết kế và tích hợp một mô hình có kiến trúc bảo mật hai lớp toàn diện cho ứng dụng ví điện tử, tập trung vào các chức năng cốt lõi là phát hiện chiếm quyền kiểm soát thiết bị và phát hiện giao dịch bất thường thông qua ứng dụng mô hình học máy Isolation Forest không giám sát. Dù được triển khai trong điều kiện thách thức về mặt tiếp cận nguồn dữ liệu thực tế và hạn chế về khảo sát, các kết quả đã xác lập tiềm năng của kiến trúc đề xuất, đặt nền móng vững chắc cho các nghiên cứu và triển khai mở rộng sau này.

Thành tựu chính là việc thiết lập thành công mô hình bảo mật hai lớp trên một ứng dụng mô phỏng, kết hợp chặt chẽ giữa các giải pháp bảo mật tĩnh và thuật toán học máy động. Lớp thứ nhất tập trung vào bảo vệ thụ động, bao gồm các cơ chế tĩnh như phát hiện bẻ khóa (*root*) bằng cách kiểm tra tập tin đặc trưng và cảnh báo ứng dụng ngoài Google Play (sử dụng cơ chế truy vấn nguồn cài đặt và danh sách trắng thủ công) để ngăn chặn rủi ro vô hiệu hóa hệ điều hành và nguy cơ mã độc gián điệp. Đặc biệt, nghiên cứu đã tiến hành khảo sát thao tác vuốt màn hình của

hơn 50 người dùng để xây dựng bộ ngưỡng động học lên các chỉ số như Vận tốc trung bình, Gia tốc trung bình, Tỷ lệ thẳng, Độ rung lắc. Bộ ngưỡng này đóng vai trò quan trọng trong việc phân biệt hành vi vuốt màn hình tự nhiên của con người với thao tác giả mạo có tính tuyến tính, ổn định cao từ bot hoặc phần mềm điều khiển từ xa, tạo ra một tập dữ liệu thử nghiệm có giá trị và cung cấp ngưỡng tham khảo ban đầu cho việc phát hiện giả mạo.

Lớp thứ hai giải quyết bài toán phát hiện bất thường trong giao dịch bằng mô hình Isolation Forest. Việc đánh giá mô hình trên bộ dữ liệu mô phỏng uy tín Paysim đã cho thấy kết quả khả quan với chỉ số AUC-ROC đạt 0.9160. Chỉ số này chứng minh khả năng phân biệt vượt trội của mô hình, đặc biệt trong môi trường mất cân bằng lớp cực đoan của dữ liệu tài chính. Kết quả này không chỉ thể hiện tính cạnh tranh cao so với các nghiên cứu tương đồng mà còn xác nhận khả năng của mô hình Isolation Forest trong việc nhận diện nhạy bén các giao dịch lệch pha (ví dụ: giao dịch có giá trị lớn vượt xa ngưỡng hành vi chi tiêu trung bình của người dùng), từ đó kích hoạt cảnh báo bảo mật trên ứng dụng giả lập. Toàn bộ module xử lý được tích hợp vào một kiến trúc đa tầng hoạt động gần như thời gian thực, khẳng định tính thực tiễn của giải pháp này.

Tuy nhiên, nghiên cứu này vẫn còn tồn tại những hạn chế đáng kể, nhất là hạn chế về bộ dữ liệu và tính đại diện. Cả hai lớp đều dựa trên dữ liệu tổng hợp, phỏng vấn quy mô nhỏ và mô phỏng. Cụ thể là dữ liệu khảo sát hành vi vuốt màn hình chỉ gồm 50 người, không đủ đại diện về mặt thống kê; Còn bộ dữ liệu Paysim thì đã hơi cũ (công bố vào năm 2016) nên vẫn còn khá đơn giản so với thực tế ở các môi trường hiện đại và tiên tiến hơn, do đó bộ dữ liệu này vẫn chưa thể phản ánh đầy đủ tính phức tạp và các mẫu bất thường ẩn sâu trong dữ liệu giao dịch thực tế trên ví điện tử hiện nay trên toàn thế giới.

B. HƯỚNG PHÁT TRIỂN

Để chuyển đổi nghiên cứu thử nghiệm này thành một giải pháp bảo mật hoàn chỉnh và hiệu quả trong môi trường sản phẩm thực tế, nhóm nghiên cứu đề xuất các hướng phát triển chính sau: Ưu tiên tuyệt đối là việc thu thập và mở rộng bộ dữ liệu, cụ thể là cần tìm kiếm cơ hội hợp tác với các tổ chức tài chính để xin cấp quyền truy cập vào dữ liệu giao dịch thực tế đã được ẩn danh và gán nhãn, bao gồm các trường hợp gian lận đã xác minh, đồng thời cam kết tuân thủ nghiêm ngặt các quy định về bảo mật và quyền riêng tư. Song song đó, cần mở rộng quy mô khảo sát/thu thập dữ liệu động học lên ít nhất 5.000 mẫu thông qua phương pháp lấy mẫu ngẫu nhiên có phân tầng (theo độ tuổi, vùng miền, thiết bị) để đảm bảo tính đại diện thống kê. Mục tiêu là xây dựng một bộ ngưỡng động học đáng tin cậy cho "Lớp phát hiện chiếm quyền thiết bị", có thể tùy chỉnh theo nhóm người dùng/thiết bị.

Khi có dữ liệu thực tế được gán nhãn, cần chuyển đổi sang mô hình học có giám sát hoặc bán giám sát (ví dụ: Random Forest, Gradient Boosting hoặc RNN/Transformer) để dự đoán nhãn gian lận, thay vì chỉ dựa vào mô hình Isolation Forest không giám sát. Cần tối ưu hóa bộ đặc trưng bằng cách khám phá và trích xuất thêm các đặc trưng kỹ thuật và bối cảnh phức tạp từ dữ liệu thực tế (ví dụ: mô hình kết nối mạng, địa điểm chi tiêu cụ thể) để làm giàu bộ dữ liệu huấn luyện, giúp mô hình Isolation Forest (hoặc các mô hình kế tiếp) có thể hoạt động hiệu quả hơn. Ngoài ra, việc sử dụng tham số tỷ lệ ngoại lai cố định cần được thay thế bằng ngưỡng động. Ngưỡng này sẽ được tính toán và hiệu chỉnh lại theo thời gian (dựa trên điểm bất thường của N giao dịch bình thường gần nhất) nhằm thích ứng với sự thay đổi hành vi người dùng, giảm thiểu tỷ lệ cảnh báo giả.

Tiếp đó là trong giai đoạn triển khai và vận hành, cần sử dụng các thuật toán học máy có giám sát để tăng cường đánh giá hiệu quả bằng các tiêu chí thực tế như *Precision*, *Recall*, và đặc biệt là tỷ lệ *False Positive*, hướng tới mục tiêu thấp nhất có thể để kiểm soát chi phí rà soát thủ công. Việc đạt được hiệu quả cao trong các tiêu chí này cũng là chìa khóa để phân biệt được thật sự là giao dịch bất thường cần được xác minh lẫn chặn hoàn toàn giao dịch gian lận thật sự, thay vì là chỉ cảnh báo chung vài phút rồi mới yêu cầu xác minh giao dịch. Ngoài ra, nhóm cũng có định hướng là sẽ xây dựng một cơ chế phản hồi vòng lặp, là một cơ chế phản hồi cần được xây dựng để thu thập thông tin từ đội ngũ vận hành (ví dụ: kết quả rà soát thủ công các cảnh báo), sử dụng dữ liệu này để gán nhãn lại và huấn luyện lại mô hình Isolation Forest hoặc mô hình giám sát mới một cách định kỳ, đảm bảo mô hình luôn cập nhật với thực tế gian lận.

Tóm lại, công trình đã chứng minh tính khả thi của mô hình sử dụng kiến trúc bảo mật hai lớp, trọng tâm phát triển tiếp theo phải là thu hẹp khoảng cách dữ liệu giữa mô phỏng và thực tế thông qua các nỗ lực mở rộng quy mô thu thập, khảo sát, và hợp tác chiến lược với đối tác tài chính có uy tín để có được dữ liệu thực.

VI. LỜI CẢM ƠN

Nghiên cứu này đã được tài trợ bởi Nguồn ngân sách khoa học và công nghệ của Trường Đại học Sư phạm Thành phố Hồ Chí Minh trong đề tài sinh viên nghiên cứu khoa học năm học 2025-2026. Đồng thời, nhóm chúng tôi cũng xin chân thành cảm ơn sự hướng dẫn của các giảng viên là Lương Trần Ngọc Khiết, Lương Trần Hy Hiến, Lê Minh Triết đã hỗ trợ và đồng hành cùng với nhóm để phát triển đề tài này.

VII. TÀI LIỆU THAM KHẢO

[1] A. Reurink (2018), "Financial Fraud: A Literature Review," *Journal of Economic Surveys*, pp. 1292-1325.

[2] C. Phua, V. Lee, K. Smith, R. Gayler (2005), "A comprehensive survey of data mining based fraud detection research," *Artificial Intelligence Review*, pp. 1-14.

[3] Z. Liu (2024), "A comparative Study of Machine Learning Methods in Financial Fraud Detection," trong *Proceedings of the 2024 2nd International Conference on Finance, Trade and Business Management (FTBM 2024)*.

[4] Dantas RM, Firdaus R, Jaleel F, Neves Mata P, Mata MN, Li G (2022), "Systemic acquired critique of credit card deception exposure through machine learning," *Journal of Open Innovation: Technology, Market, and Complexity (JOItmC)*, p. 8(4):192.

[5] Mongwe W, Malan K (2020), "A survey of automated financial statement fraud detection with relevance to the South African context," *South African Computer Journal (SACJ)*, p. 32(1): 75.

[6] Elena-Adriana Minastireanu, Gabriela Mesnita (2024), "An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection," trong *Proceedings of the 2024 2nd International Conference on Finance, Trade and Business Management (FTBM 2024)*, Iași.

[7] Ibrahim Y. Hafez, Ahmed Y. Hafez, Ahmed Saleh, Amr A. Abd El-Mageed, Amr A. Abohany (2025), "A symenatic review of AI-enhanced techniques in credit card fraud detection," *Journal of Big Data*, pp. 3-23.

[8] M. Ngọc (2025), "Xu hướng công nghệ định hình ví điện tử tại Việt Nam,"[Trực tuyến]. Available: https://thitruongtaichinhtiente.vn/xu-huong-cong-nghe-dinh-hinh-vi-dien-tu-tai-viet-nam-68019.html.

[9] M. Thiện (2024), "Thực trạng và xu hướng lừa đảo trực tuyến tại Việt Nam," [Trực tuyến]. Available: https://vjst.vn/thuc-trang-va-xu-huong-lua-dao-truc-tuyen-tai-viet-nam-67940.html.

[10] P. Anh (2025), "Cảnh giác để không mất tiền oan khi sử dụng ví điện tử," [Trực tuyến]. Available: https://cand.com.vn/Phong-su/canh-giac-de-khong-mat-tien-oan-khi-su-dung-vi-dien-tu-i780310/.

[11] Ramli, F.A.A. and Hamzah, M.I. (2021), "Mobile payment and e-wallet adoption in emerging economies: a systematic literature review," *Journal of Emerging Economies and Islamic Research*, tập 9, số 2, pp. 1-39.

[12] Elok, C.S., Kom, S. and Hidayati, A. (2021), "Customer Loyalty in Digital Wallet Industry: the Role of Satisfaction, Effort Expectancy, Performance Expectancy, and Habit," trong *International Conference on Emerging Challenges: Business Transformation and Circular Economy*.

[13] Lee, W.I., Fu, H.P., Mendoza, N. and Liu, T.Y. (2021), "Determinants impacting user behavior towards emergency use intentions of m-health services in Taiwan," *Healthcare*, tập 9, số 5, p. 535.

[14] Phuong, N.N.D., Luan, L.T., Dong, V. and van and Nguyen, L.N.K. (2020), "Examining customers' continuance intentions towards e-wallet usage: the emergence of mobile payment acceptance in Vietnam," *The Journal of Asian Finance, Economics and Business*, tập 7, số 9, pp. 505-516.

[15] Syifa, N. and Tohang, V. (2020), "The use of e-wallet system," trong *International Conference on Information Management and Technology, ICIMTech 2020*, Bandung, Indonesia.

[16] Zhao Peng, Wang Wenjian, Wu Di (2025), "Abnormal transaction detection in blockchain based on XGBoost and random forest," *Journal of Nanjing University of Posts and Telecommunications*, tập 5, số 1, pp. 115-122.

[17] Wang Dong, Li Da, Wang Hejian (2024), "Blockchain anomaly transaction detection based on deep PCA and Bayesian optimization," *Southern Power Grid Technology*, tập 18, số 9, pp. 2-14.

[18] Liao Qian, Gu Yijun (2022), "Bitcoin network abnormal transaction detection based on LSCP algorithm," *Computer Engineering and Applications*, tập 58, số 15, pp. 117-123.

[19] Song Yuhan, Zhu Yuefei, Wei Fushan (2024), "A blockchain abnormal transaction detection scheme based on AdaBoost model," *Information Network Sercurity*, tập 1, pp. 24-35.

[20] M. Bharath Maneela, M. Sri SaiHarshaa, S. Rahul Saia, CH. Viveka, M Kavithaa, Dharmaiah Devarapallia (2025), "Anomaly Detection in Transactions Using Machine Learning," *Edu - Tech Enterprise*, tập 3, pp. 2-7.

[21] Zhengkun Xiu (2023), "Financial Transaction Anomaly Detection Based on Transformer Model," trong *The 5th International Conference on Multi-modal Information Analytics (MMIA)*.

[22] Soumaya Ounacer, Hicham Ait El Bour, Younes Oubrahim, Mohamed Yassine Ghoumari, Mohamed Azzouazi (2018), "Using Isolation Forest in anomaly detection: the case of credit card transactions," *Periodicals of Engineering and Natural Sciences*, tập 6, số 2, pp. 394-400.

[23] E. Lopez-Rojas (2016), "Synthetic Financial Datasets For Fraud Detection," [Trực tuyến]. Available: https://www.kaggle.com/datasets/ealaxi/paysim1.

[24] Edgar Alonso Lopez-Rojas, Ahmad Elmir, and Stefan Axelsson (2016), "PaySim: A Financial Mobile Money Simulator For Fraud Detection," trong *28th European Modeling and Simulation Symposium*, Rende, Italy.

[25] D. P. N. M. Prakash S. Bodkhe (2021), "CAPTCHA Techniques: An Overview," trong *2nd National Conference Recent Innovations in Science and Engineering*.

[26] Marcus Liwicki, Alex Graves, Horst Bunke Jürgen Schmidhuber (2007), "A Novel Approach to On-Line Handwriting Recognition Based on Bidirectional Long Short-Term Memory Networks," trong *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Toronto.

[27] Victor Carbune, Pedro Gonnet, Thomas Deselaers, Henry A. Rowley, Alexander Daryin, Marcos Calvo (2020), "Fast multi-language LSTM-based online handwriting recognition," *International Journal on Document Analysis and Recognition*, tập 23, p. 89.

[28] S. A. B. Edgar Alonso Lopez-Rojas (2018), "Analysis of Fraud Controls Using the PaySim Financial Simulator," *International Journal of Simulation and Process Modelling*, tập 13, số 4, pp. 4-6.

[29] A. M. J. F. R. V.-R. O. D.-M. Alejandro Acien (2021), "BeCAPTCHA: Behavioral Bot Detection using Touchscreen and Mobile Sensors benchmarked on HuMIdb," *Engineering Applications of Artificial Intelligence*, tập 98, số 1, pp. 2-18.

[30] R. Gillespie (2019), "Detecting Fraud and Other Anomalies Using Isolation Forests," *SAS Institute Inc*, pp. 3-5.

VIII. PHỤ LỤC

BẢNG CẤU TRÚC CHI TIẾT SAU KHI TIẾN HÀNH KỸ THUẬT XÂY DỰNG ĐẶC TRƯNG

STT	Cột dữ liệu	Kiểu dữ liệu	Ý nghĩa
1	amount	Số thực	Giá trị giao dịch, là một trong những chỉ báo chính cho hành vi bất thường.
2	deltaBalanceOrig	Số thực	Sự thay đổi số dư thực tế của người gửi (oldbalanceOrg - newbalanceOrig)
3	deltaBalanceDest	Số thực	Sự thay đổi số dư thực tế của người nhận (newbalanceDest - oldbalanceDest)
4	errorBalanceOrig	Số thực	Lỗi số dư của người gửi: deltaBalanceOrig - amount. Giá trị không gần 0 chỉ ra sự thiếu sót trong dữ liệu hoặc bất thường.
5	errorBalanceDest	Số thực	Lỗi số dư của người nhận: deltaBalanceDest - amount. Tương tự, chỉ ra sự không khớp/bất thường.
6	isOrigEmptyAfterTx	Nhị phân	1 nếu số dư người gửi là 0 sau giao dịch (tài khoản bị rút cạn), 0 nếu ngược lại.
7	isDestCustomer	Nhị phân	1 nếu người nhận là Khách hàng ('C'), 0 nếu là Thương gia ('M').
8	type_CASH_IN	Nhị phân	1 nếu giao dịch là nạp tiền mặt, 0 nếu ngược lại.
9	type_CASH_OUT	Nhị phân	1 nếu giao dịch là rút tiền mặt, 0 nếu ngược lại. Đây là loại giao dịch có tỷ lệ gian lận cao.
10	type_DEBIT	Nhị phân	1 nếu giao dịch là ghi nợ (thường là số lượng nhỏ), 0 nếu ngược lại.
11	type_PAYMENT	Nhị phân	1 nếu giao dịch là thanh toán (thường là tới thương gia), 0 nếu ngược lại.
12	type_TRANSFER	Nhị phân	1 nếu giao dịch là chuyển khoản giữa các tài khoản, 0 nếu ngược lại. Đây là loại giao dịch có tỷ lệ gian lận cao.

13	oldbalanceOrg	Số thực	Số dư ban đầu của người gửi.
14	newbalanceOrig	Số thực	Số dư mới của người gửi.
15	oldbalanceDest	Số thực	Số dư ban đầu của người nhận.
16	newbalanceDest	Số thực	Số dư mới của người nhận.
17	isFraud	Nhị phân	Nhãn dự đoán (0=Không gian lận, 1=Gian lận). Chỉ được sử dụng để tách tập và đánh giá.
18	isFlaggedFraud	Nhị phân	Cờ do mô hình ban đầu đặt ra. Rất hiếm khi là 1 và không phải là chỉ báo chính xác.
19	step	Số nguyên	Đơn vị thời gian trong mô phỏng (tương đương 1 giờ).
20	nameOrig	Định danh	ID của người gửi.
21	nameDest	Định danh	ID của người nhận.

BUILDING A MODEL TO DETECTION AND WARNING ANOMALIES TRANSACTIONS ON E-WALLET

Nguyen Quoc Chi*, Luong Tran Ngoc Khiet, Luong Tran Hy Hien, Tran Hoang Dat, Hoang Tan Dung, Doan Quang Thieu, Le Minh Triet

ABSTRACT— The rapid development of e-wallets has led to a fast increase in online transaction security risks, particularly the dual threats stemming from device takeover attempts and financial fraud. This research proposes a comprehensive two-layer security architecture designed to provide a multi-layered solution for anomaly detection and alerting. The First Layer (Device Takeover Detection) serves as the initial defense, performing static checks (e.g., rooting, malicious applications) and dynamic threshold analysis of user interaction kinematics (e.g., swipe velocity, jitter) to ensure environmental integrity and prevent device compromise. The Second Layer (Anomalies Transaction Detection) utilizes the unsupervised Isolation Forest (IF) algorithm to identify transactional anomalies. This layer was trained on the reputable simulated dataset Paysim (6.3 million transactions), originally published at the "28th European Modeling and Simulation Symposium," with specialized feature engineering applied to optimize anomaly recognition capability. Experimental results demonstrate that Layer 1 successfully detected and blocked login attempts when the device was compromised in simulated environments. Furthermore, Layer 2 achieved superior efficacy, with the Isolation Forest model attaining a ROC-AUC score of 0.9160, proving its robust classification ability in isolating anomalous transactions. Although performance was demonstrated using simulated data, this combined architecture affirms the feasibility of a multi-layered security solution, which flexibly adapts to threats originating from both the device level and fraudulent behaviors.

Keywords — Machine Learning, Unsupervised, Anomalies Transactions Detection, Paysim Dataset, E-wallet Transaction, Fraud Transactions Detection, Isolation Forest

Nguyễn Quốc Chí là sinh viên năm thứ ba của Trường Đại học Sư phạm TP.HCM vào thời điểm đăng bài báo này. Hiện đang quan tâm về các nghiên cứu trong lĩnh vực học máy ứng dụng, công nghệ giáo dục và phân tích dữ liệu.

Lương Trần Ngọc Khiết nhận bằng Cử nhân Công nghệ phần mềm (2016) và bằng Thạc sĩ ngành Khoa học máy tính (2019) tại Trường Đại học Sư phạm TP.HCM. Hiện ông là giảng viên của Trường Đại học Sư phạm TP.HCM. Hướng nghiên cứu chính tập trung vào Trí tuệ nhân tạo, phân tích dữ liệu và các ứng dụng công nghệ giáo dục. Đặc biệt là các ứng dụng tích hợp AI.

Lương Trần Hy Hiến nhận bằng thạc sĩ ngành Khoa học máy tính tại Trường Đại học Công nghệ thông tin, ĐHQG TP.HCM vào năm 2014. Hiện ông là nghiên cứu sinh ngành Hệ thống Thông tin tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, đồng thời là giảng viên Khoa Công nghệ thông tin tại Trường Đại học Sư phạm TP.HCM. Hướng nghiên cứu tập trung vào trí tuệ nhân tạo và các ứng dụng, đặc biệt là các bài toán tổ chức thông tin, tìm kiếm ảnh theo ngữ nghĩa, và các mô hình ngôn ngữ – thị giác máy tính.

Trần Hoàng Đạt là sinh viên năm thứ ba của Trường Đại học Sư phạm TP.HCM vào thời điểm đăng bài báo này. Hiện đang quan tâm về các nghiên cứu trong lĩnh vực học máy ứng dụng, công nghệ giáo dục và phân tích dữ liệu.

Hoàng Tấn Dũng là sinh viên năm thứ ba của Trường Đại học Sư phạm TP.HCM vào thời điểm đăng bài báo này. Hiện đang quan tâm về các nghiên cứu trong lĩnh vực học máy ứng dụng, công nghệ giáo dục và phân tích dữ liệu.

Đoàn Quang Thiệu là sinh viên năm thứ ba của Trường Đại học Sư phạm TP.HCM vào thời điểm đăng bài báo này. Hiện đang quan tâm về các nghiên cứu trong lĩnh vực học máy ứng dụng, công nghệ giáo dục và phân tích dữ liệu.

Lê Minh Triết nhận bằng thạc sĩ ngành Khoa học máy tính tại Trường Đại học Công nghệ Thông tin, ĐHQG TP.HCM vào năm 2015. Hiện nay, ông là giảng viên khoa Công nghệ thông tin, Trường Đại học Sư phạm TP.HCM. Hướng nghiên cứu của ông tập trung vào Khoa học máy tính, Trí tuệ nhân tạo và các ứng dụng, thị giác máy tính, mạng thông tin và truyền thông.