ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

Ngày nhận hồ sơ	
(Do CQ quản lý ghi)	

THUYẾT MINH
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ CẤP SINH VIÊN 2024

A. THÔNG TIN CHUNG

A1. Tên đề tài

- Tên tiếng Việt (IN HOA): PHÁT HIỆN LỖ HỔNG TRONG HỢP ĐỒNG THÔNG MINH TRÊN MẠNG LIÊN CHUỖI KHỐI BẰNG PHƯƠNG PHÁP HỌC MÁY VÀ HỌC SÂU
- Tên tiếng Anh (IN HOA): SMART CONTRACT VULNERABILITIES AUTOMATIC DETECTION ON THE CROSS-CHAIN NETWORK USING MACHINE LEARNING AND DEEP LEARNING

A2. Loại hình nghiên cứu

(Tham khảo tiêu chuẩn đề tài đối với từng loại hình NC, chọn 01 trong 03 loại hình)

☒ Nghiên cứu cơ bản

□ Nghiên cứu ứng dụng

□ Nghiên cứu triển khai

A3. Thời gian thực hiện

..06.. tháng (kể từ khi được duyệt).

A4. Tổng kinh phí

(Lưu ý tính nhất quán giữa mục này và mục B8. Tổng hợp kinh phí đề nghị cấp)

Tổng kinh phí: …**6**.. triệu đồng, gồm

- Kinh phí từ Trường Đại học Công nghệ Thông tin: ..**6**.. triệu đồng

A5. Chủ nhiệm

Họ và tên: **VÕ ANH KIỆT** .

Ngày, tháng, năm sinh: 27/12/2002 . Giới tính (Nam/Nữ): Nam .

Số CMND: 079202029779; Ngày cấp: 21/12/2021 ; Nơi cấp: Cục cảnh sát .

Mã số sinh viên: 20520605 .

Số điện thoại liên lạc: 0365642317 .

Đơn vị (Khoa): Khoa Mạng máy tính và Truyền thông .

Số tài khoản: 1410478279 Ngân hàng: BIDV CHI NHANH CHO LON PGD QUAN 6

A6. Nhân lực nghiên cứu

TT	Họ tên	MSSV	Khoa/Bộ Môn
1	Võ Anh Kiệt	20520605	MMT&TT/ATTT

B. MÔ TẢ NGHIÊN CỨU

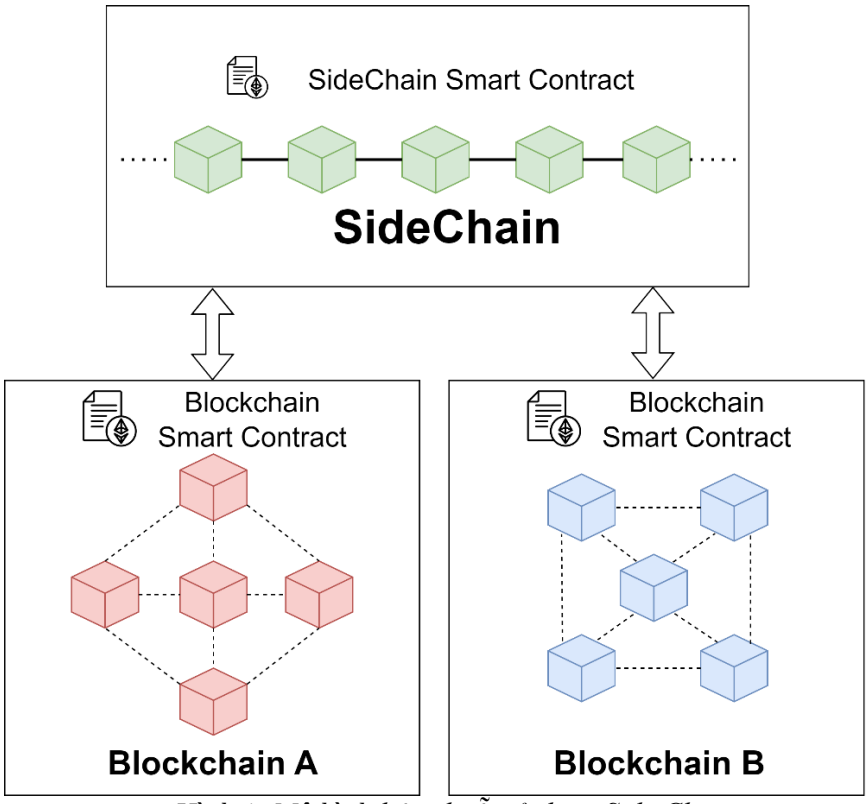
B1. Giới thiệu về đề tài

Sự phát triển của công nghệ blockchain đã tạo ra bước tiến lớn trong việc hình thành các mạng không tập trung, mang lại khả năng lưu trữ và trao đổi thông tin một cách an toàn và chính xác. Điều này được thực hiện thông qua cơ chế đồng thuận phân tán giữa các nút mạng [1, 2]. Tuy nhiên, một hạn chế của hệ thống blockchain là việc chúng thường được xây dựng một cách độc lập với quy tắc và giao thức riêng, dẫn đến khó khăn trong việc tương tác và trao đổi dữ liệu giữa các blockchain khác nhau.

Đáp ứng vấn đề này, công nghệ liên chuỗi (cross-chain) đã được phát triển như một giải pháp cho phép các mạng blockchain tương tác với nhau một cách an toàn và hiệu quả [3]. Công nghệ này hỗ trợ việc chuyển giao tài sản số giữa các chuỗi khác nhau, từ đó thúc đẩy khả năng tích hợp và phát triển của hệ sinh thái blockchain [4]. Các giải pháp như Polkadot, chẳng hạn, đã cung cấp một khung làm việc cho phép các blockchain độc lập, với cấu trúc và chức năng khác nhau, giao tiếp và tương tác với nhau, mở ra một bước tiến mới trong lĩnh vực công nghệ blockchain và các ứng dụng phi tập trung [5].

Công nghệ liên chuỗi nhằm kết nối các hệ sinh thái blockchain bị cô lập, cho phép tài sản và dữ liệu được chuyển giao và chia sẻ một cách thuận tiện giữa các blockchain khác nhau [6]. Đóng góp này tiến hành giải quyết một vấn đề rất quan trọng trong việc xử lý các điểm hạn chế về khả năng xử lý cũng như khả năng về mở rộng và đồng thời là chức năng mà các blockchain riêng lẻ thường gặp phải. Hiện tại tồn tại ba cách tiếp cận chính để kết nối và cho phép chuyển tài sản giữa các chuỗi khối khác nhau: các giải pháp công chứng, khóa băm, và các relays/sidechains [7–9]. Cơ chế công chứng là phương pháp đồng thuận trong đó các bên thứ ba tin cậy (công chứng viên) xác minh giao dịch bằng chữ ký số trước khi chúng được thêm vào blockchain, nhằm ngăn giao dịch kép. Giải pháp khóa băm có thể triển khai thông qua cổng để truy cập các hợp đồng khóa thời gian khóa băm (Hash Time Lock Contract - HTLC) trên blockchain từ xa, đảm bảo việc nhận thanh toán trước khi phục dựng tài sản trên blockchain đích.

Sidechains là các blockchain phụ được gắn nối với blockchain chính thông qua thanh gài 2 chiều, cho phép chuyển tài sản giữa các chuỗi, thêm tính năng mới cho blockchain chính mà không cần sửa đổi giao thức, giải quyết những vấn đề thách thức như việc mở rộng hệ thống và đảm bảo vấn đề về sự riêng tư của các thông tin. Tuy nhiên, giao thức tương tác chuỗi khối này vẫn tồn tại các vấn đề về tính bí mật, tính riêng tư và sự sẵn sàng của hệ thống tư cần giải quyết.



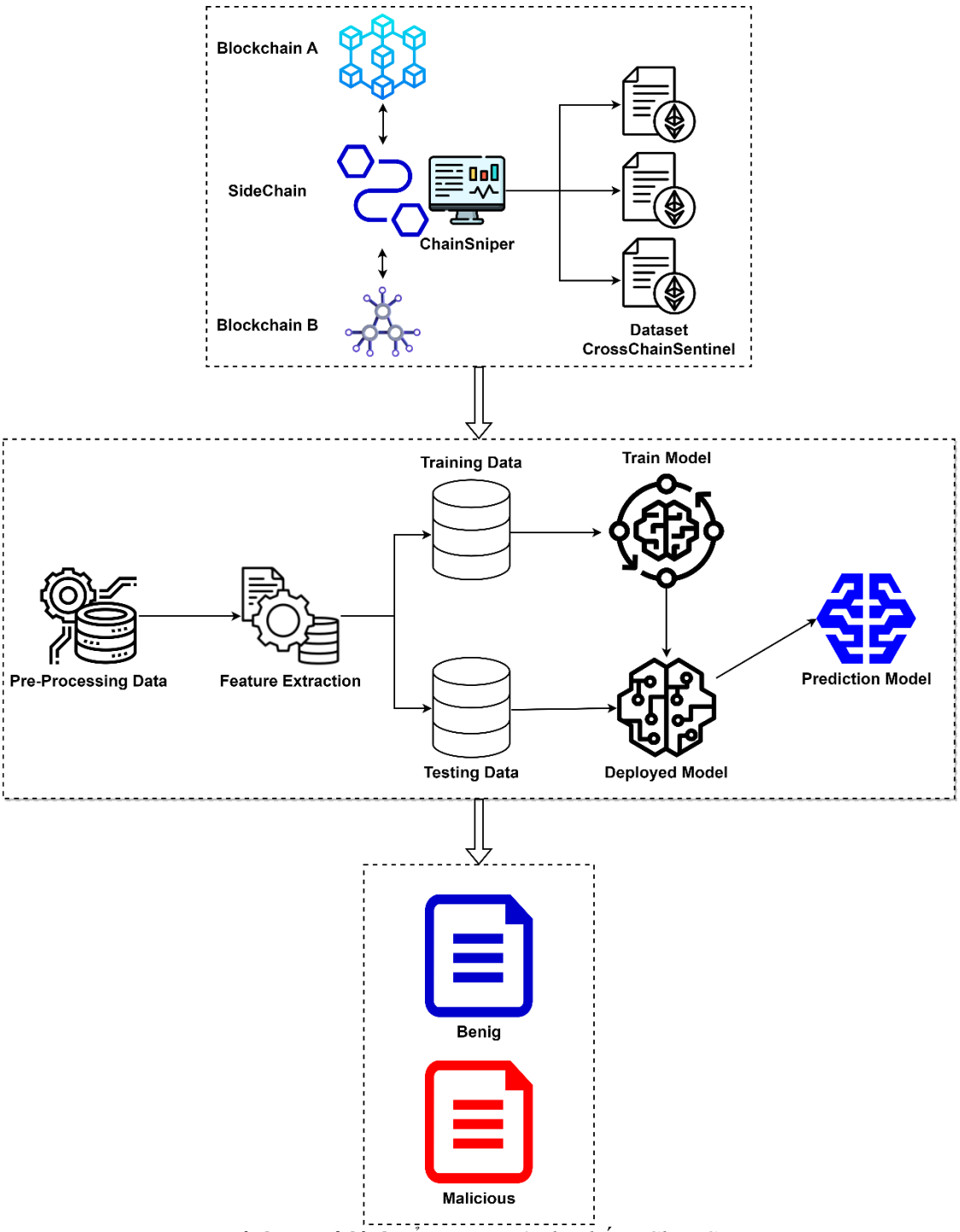
Hình 1: Mô hình liên chuỗi sử dụng Side Chain

Đặc biệt, các cuộc tấn công liên chuỗi nhằm vào việc lợi dụng các điểm yếu trong hợp đồng thông minh tiến hành thực thi trên nhiều mạng blockchain riêng biệt, có thể gây ra tổn thất tài chính nghiệm trọng. Một ví dụ điển hình là vụ việc tấn công gần đây, nơi mà lỗ hổng trong một hợp đồng thông minh liên chuỗi đã bị khai thác, dẫn đến việc mất mát tài sản trị giá hơn 600 triệu đô la [11]. Một trường hợp khác, vào năm 2016, dự án gây quỹ The DAO trên Ethereum cũng đã trở thành nạn nhân của một cuộc tấn công, mất đi hơn 50 triệu đô la do lỗ hổng trong hợp đồng thông minh của họ. Các sự cố tấn công hợp đồng thông minh tiếp tục xảy ra, ví dụ như vụ tấn công vào ví Parity tháng 7/2017 khiến mất mát 30 triệu đô la tiền điện tử Ether. Hay vụ đánh cắp gần 300.000 đô la từ nền tảng KingDice tháng 8/2017 cũng do lợi dụng lỗ hổng trong mã hợp đồng. Hơn thế nữa, trong gian đoạn gần đây, thị trường blockchain đã tiếp nhận 1 cuộc tấn công là loạt vụ tấn công vào các hợp đồng thông minh trên Binance Smart Chain năm 2021, trong đó có vụ đánh cắp hơn 200 triệu đô la thông qua hợp đồng của Venus Protocol [12, 13]. Như vậy, việc phân tích và bảo mật hợp đồng thông minh trước khi tiến hành thực thi và sử dụng mang tính cấp thiết lớn trong việc hạn chế rủi ro mất mát tài sản. Hợp đồng thông minh, là các giao thức số được thiết kế để làm đơn giản hóa, kiểm chứng hoặc thực thi các quy trình đàm phán và thực hiện hợp đồng. Chúng được ứng dụng rộng rãi, từ dịch vụ tài chính đến thị trường dự đoán và trong lĩnh vực Internet vạn vật [14]. Những hợp đồng này hoạt động hiệu quả trên các nền tảng blockchain, tự động hoá các hành động theo các điều kiện đã đặt ra trước, giảm bớt nhu cầu cho các bên trung gian. Hợp đồng thông minh, do đó, tạo điều kiện cho việc giao dịch không dựa vào sự tin tưởng và tự động hóa thực hiện các quy trình trong hệ thống blockchain. Với sự phát triển nhanh chóng của công nghệ blockchain và việc ứng dụng nó trong nhiều ngành nghề khác nhau, việc phân tích và kiểm tra bảo mật cho hợp đồng thông minh trở nên cực kỳ quan trọng trước khi chúng được triển khai [15]. Do chúng hoạt động dựa trên mã tự thực thi, bất kỳ lỗ hổng nào cũng có thể gây ra hậu quả lớn. Mặc dù việc kiểm tra và phân tích mã bằng phương pháp thủ công là thiết yếu, nhưng quá trình này lại mất nhiều thời gian và công sức, và cũng dễ phát sinh lỗi do con người. Trong các hệ thống blockchain có liên kết chéo như Polkadot, việc bảo mật và kiểm thử càng trở nên phức tạp và thách thức [16].

Hiện nay, hệ sinh thái Ethereum đang phải đối mặt với nhiều lỗ hổng bảo mật nghiêm trọng như các cuộc tấn công lặp lại (Reentrancy) [16], vấn đề tràn số (Overflow/Underflow) [17] và các vấn đề liên quan đến việc rút token không an toàn trong các hợp đồng thông minh [18]. Những vấn đề này đều là những rủi ro lớn đối với các ứng dụng phi tập trung (dApps) trên Ethereum. Do đó, cần có các giải pháp và công cụ kiểm thử bảo mật hiệu quả hơn để đáp ứng với tốc độ phát triển của các ứng dụng blockchain hiện đại. Mạng chuỗi liên kết chéo (cross-chain) mở ra cơ hội phát triển các ứng dụng phi tập trung phức tạp hơn bằng cách kết nối nhiều blockchain với nhau. Tuy nhiên, điều này cũng tạo ra nhiều lỗ hổng bảo mật hơn, mà kẻ tấn công có thể khai thác để gây ra các vụ tấn công [19]. Một trong những cuộc tấn công đáng được chú ý đó là cuộc tấn công lặp lại (reentrancy), cho phép đối tượng tấn công lặp đi lặp lại lời gọi hàm của hợp đồng thông minh trước khi hoàn thành yêu cầu trước đó, dẫn đến các hậu quả khó lường [20]. Bên cạnh đó, các lỗ hổng tràn số và underflow cũng rất nguy hiểm, xảy ra khi giá trị vượt ngưỡng trên hoặc ngưỡng dưới cho phép, sẽ gây ra hậu quả khôn lường [21]. Đặc biệt, lỗ hổng rút tiền điện tử mà không được xác thực (unprotected ether withdrawal) cũng đang nhận được sự quan tâm, khi một hợp đồng thông minh không xác minh chính xác yêu cầu rút tiền, cho phép hacker rút Ether một cách bất hợp pháp [22]. Những hậu quả từ các lỗ hổng bảo mật nói trên đã và đang lan rộng trong hệ sinh thái Ethereum, gây thiệt hại tài chính lớn cho nhiều bên liên quan [12]. Tình hình nghiêm trọng hiện nay nhấn mạnh sự cần thiết của việc phát triển và tuân thủ nghiêm ngặt các chính sách và quy trình bảo mật trong quá trình tạo ra các ứng dụng blockchain. Điều này đòi hỏi các doanh nghiệp và tổ chức phải tập trung đầu tư vào các hoạt động kiểm tra, phát hiện lỗ hổng và cải tiến mã nguồn của sản phẩm trước khi chúng được triển khai. Bằng cách này, có thể giảm thiểu rủi ro và thiệt hại do các lỗi bảo mật trong ứng dụng blockchain gây ra.

Học máy, một lĩnh vực của trí tuệ nhân tạo, cho phép máy tính học hỏi và cải thiện từ kinh nghiệm mà không cần sự lập trình chi tiết. Sự tiến bộ trong học máy, đặc biệt là phát triển của học sâu, đã mở ra khả năng xử lý dữ liệu lớn, dự đoán và ra quyết định trong nhiều lĩnh vực vượt trội so với con người [23, 24]. Trong lĩnh vực hợp đồng thông minh, các mô hình học máy có thể được sử dụng để phân tích mã và nhận diện các vấn đề như lỗi lặp, tấn công tràn số, và tấn công từ chối dịch vụ (DoS) [25]. Các mô hình này, được huấn luyện từ cả các ví dụ về hợp đồng thông minh dễ tổn thương và an toàn, có thể thực hiện xác minh hợp đồng một cách hiệu quả và tự động ở quy mô lớn.

Như vậy, có nhiều tiến triển trong quá trình phát triển và triển khai hợp đồng thông minh trên mạng chuỗi chéo, việc này giúp tạo điều kiện cho giao tiếp và trao đổi dữ liệu giữa các blockchain khác nhau được diễn ra một cách hiệu quả. Tuy nhiên, việc triển khai các ứng dụng này mang theo rủi ro về an toàn thông tin, đặc biệt là trong việc phát hiện lỗ hổng trong các hợp đồng thông minh, có thể gây nguy hiểm đến tính bảo mật. Trong các nghiên cứu trước đây đã tập trung vào việc xác định và phát hiện lỗ hổng trong hợp đồng thông minh bằng các phương pháp kiểm tra ký tự và thực thi, tuy nhiên, các phương pháp hiện nay vẫn chưa đạt được khả năng phân tích toàn diện. Do đó, trong nghiên cứu này, nhóm đề xuất sử dụng các phương pháp học máy và học sâu để phân tích các lỗ hổng này một cách hiệu quả hơn. Ở công trình nghiên cứu này, em giới thiệu các phương pháp học máy và học sâu dựa trên ChainSniper - một khung phân tích tích hợp học máy dựa trên sidechain để tự động đánh giá lỗ hổng hợp đồng thông minh chéo chuỗi. Phương pháp mà nhóm đề xuất là xây dựng một tập dữ liệu quy mô lớn gồm 300 đoạn mã có gán nhãn thủ công, được gọi là CrossChainSentinel, đã được biên soạn để huấn luyện các mô hình phân biệt mã dễ bị tấn công và mã an toàn. Các đoạn mã này bao gồm các lỗ hổng: Reentrancy, Interger Overflow/Underflow và Unprotected Ether Withdrawl. Kết quả thực nghiệm đã chứng minh được tính hiệu quả của việc ứng dụng học máy và học sâu giúp tăng sự hiệu quả của việc kiểm tra hợp đồng thông minh cho các ứng dụng phi tập trung phân tán trên nhiều blockchain. Độ chính xác phát hiện đạt mức đáng kể, khẳng định tiềm năng của ChainSniper trong việc tăng cường an ninh thông qua đánh giá tự động và toàn diện mã hợp đồng.



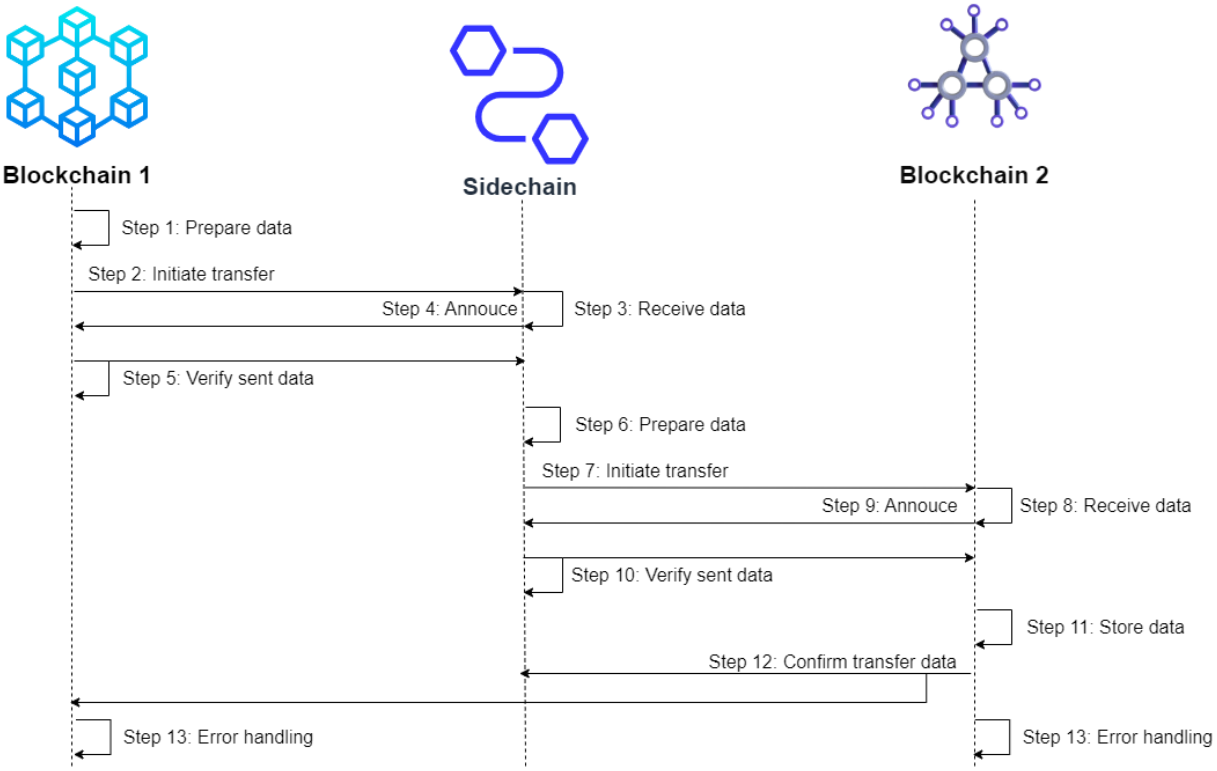
Hình 2: Mô hình tổng quan của hệ thống ChainSniper

Về mô hình cầu nối:

Hệ thống ChainSniper cho phép tương tác giữa các blockchain không đồng nhất qua cầu nối sidechain. Sidechain này đóng vai trò như một lớp trung gian, xử lý và chuyển dữ liệu giữa các mạng blockchain kết nối với nhau. Để có thể truyền tải dữ liệu, tài sản giữa các chuỗi được khóa lại ở một chuỗi này và giải khóa những biểu diễn tương đương ở chuỗi kia thông qua cơ chế neo hai chiều dựa trên hợp đồng đa chữ ký.

Khi có các giao dịch xuyên chuỗi diễn ra, các nút của sidechain sẽ ghi lại các thông tin, dữ liệu của giao dịch đó như địa chỉ của các hợp đồng thông minh tham gia, dấu thời gian diễn ra, các lời gọi hàm, các tham số được truyền vào, giá trị trả về, cũng như các ngoại lệ nếu có, sẽ được ghi lại trong nhật ký giao dịch. Nhật ký này sau đó được tổng hợp và xử lý để tạo nên một tập dữ liệu,

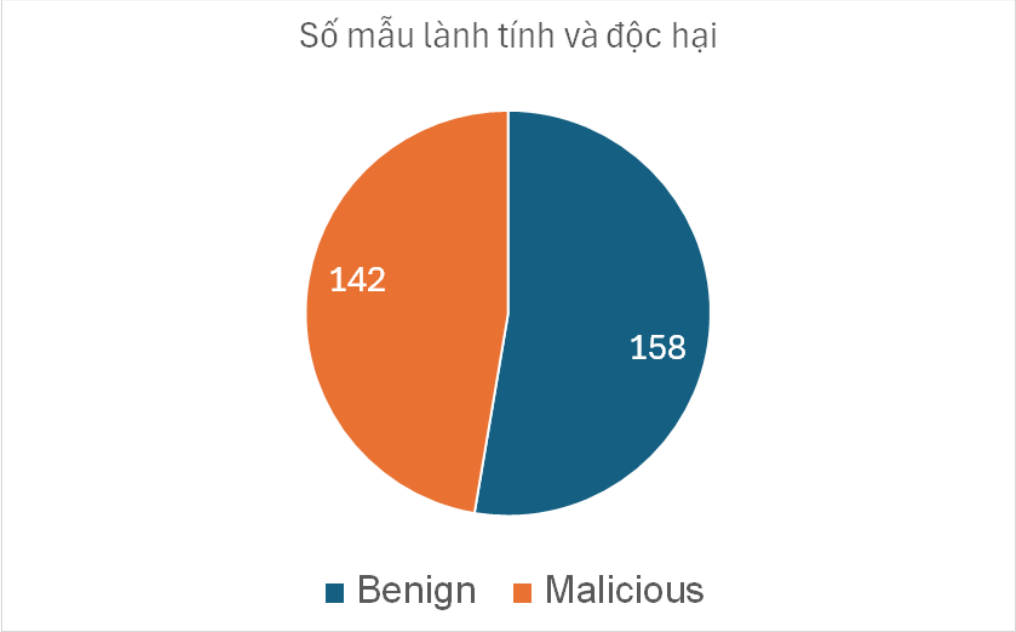
cung cấp cái nhìn sâu sắc về hành vi và cách thức hoạt động của hợp đồng thông minh, cũng như các mẫu thực thi khác nhau của chúng.
Các bước thực hiện việc chuyển đổi dữ liệu qua sidechain:
1. Chuẩn bị dữ liệu: Xác định cấu trúc và bảo đảm tính nguyên vẹn của dữliệu trước khi thực hiện chuyển đổi.
2. Khởi tạo quá trình chuyển đổi: Kết nối với sidechain để khởi tạo quá trình chuyển đổi.
3. Nhận dữ liệu ở sidechain: Sidechain nhận dữ liệu đến từ hệ thống blockchain A sau khi quá trình chuyển đổi được khởi tạo.
4. Sidechain thông báo nhận dữ liệu: blockchain A nhận thông báo từ sidechain về sự kiện chuyển đổi dữ liệu.
5. Kiểm tra Dữ liệu Trên Sidechain: Sidechain tiến hành xác nhận độ chính xác và tính nguyên vẹn của dữ liệu nhận được từ hệ thống blockchain chính.
6. Chuẩn bị dữ liệu (lần 2): Chuẩn bị lại dữ liệu cho lần chuyển đổi tiếp theo
7. Khởi tạo quá trình chuyển đổi (lần 2): Blockchain B tiếp tục kết nối với sidechain để khởi tạo lần chuyển đổi tiếp theo
8. Nhận dữ liệu (lần 2): Sidechain gửi dữ liệu mới đến Blockchain B.
9. Thông báo: Blockchain B nhận thông báo từ sidechain về sự kiện chuyển đổi dữ liệu lần thứ hai.
10. Xác minh dữ liệu đã gửi từ sidechain (lần 2): Sidechain tiến hành kiểm tra và đảm bảo tính chính xác cũng như nguyên vẹn của dữ liệu vừa được chuyển giao.
11. Bảo quản Dữ liệu: Dữ liệu được bảo lưu trên Blockchain B để sử dụng trong các hoạt động tiếp theo.
12. Xác nhận Chuyển đổi Dữ liệu: Blockchain B kiểm chứng rằng quá trình chuyển đổi đã hoàn tất thành công và dữ liệu đã được biến đổi một cách chính xác.
13. Quản lý Lỗi: Trong trường hợp xuất hiện lỗi, các mạng blockchain phải tiến hành xử lý, bao gồm cả việc gửi thông báo lỗi và ghi nhận chúng vào log để có thể theo dõi và giải quyết vấn đề.



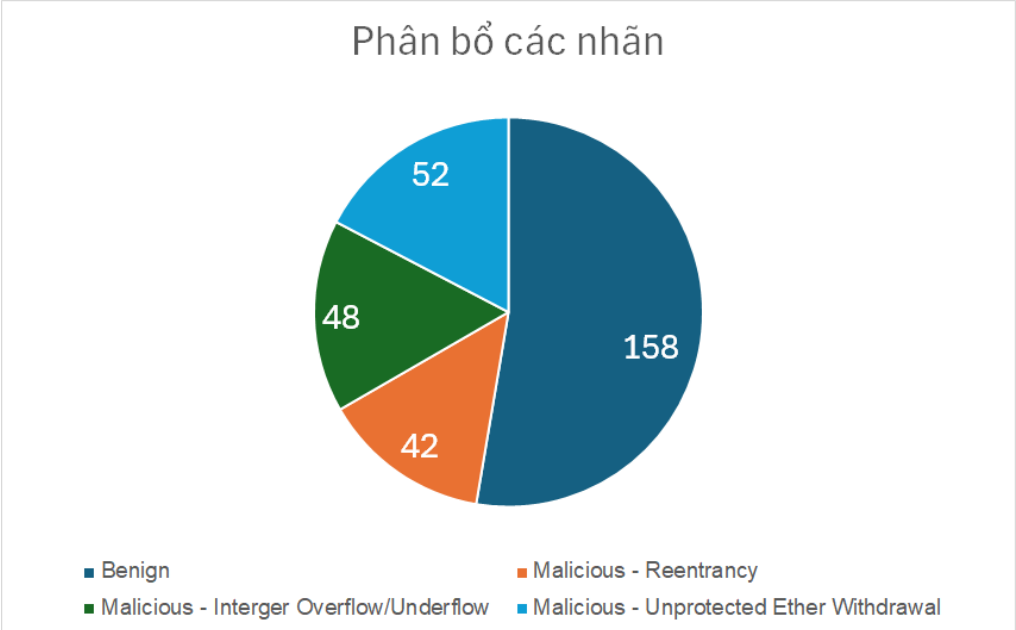
Hình 3: Quá trình chuyển đổi dữ liệu giữa 2 chuỗi

Về việc xây dựng tập dữ liệu và xử lý dữ liệu:

Tập dữ liệu CrossChainSentinel được tạo ra với mục đích phân tích các lỗ hổng tiềm ẩn trong các hợp đồng thông minh trên cầu nối sidechain. Tập dữ liệu này chứa 300 tệp hợp đồng thông minh, trong đó 158 mẫu được xác định là lành tính, 142 mẫu còn lại được xếp vào loại độc hại. Cụ thể, số mẫu độc hại bao gồm 42 hợp đồng bị lỗ hổng tái tham chiếu (Reentrancy), 48 hợp đồng chứa lỗi tràn/cạn số nguyên (Interger Overflow/Underflow) và 52 hợp đồng có vấn đề rút Ether ra mà không được bảo vệ (Unprotected Ether Withdrawal):



Hình 4: Phân bố mẫu độc hại và mẫu lành tính

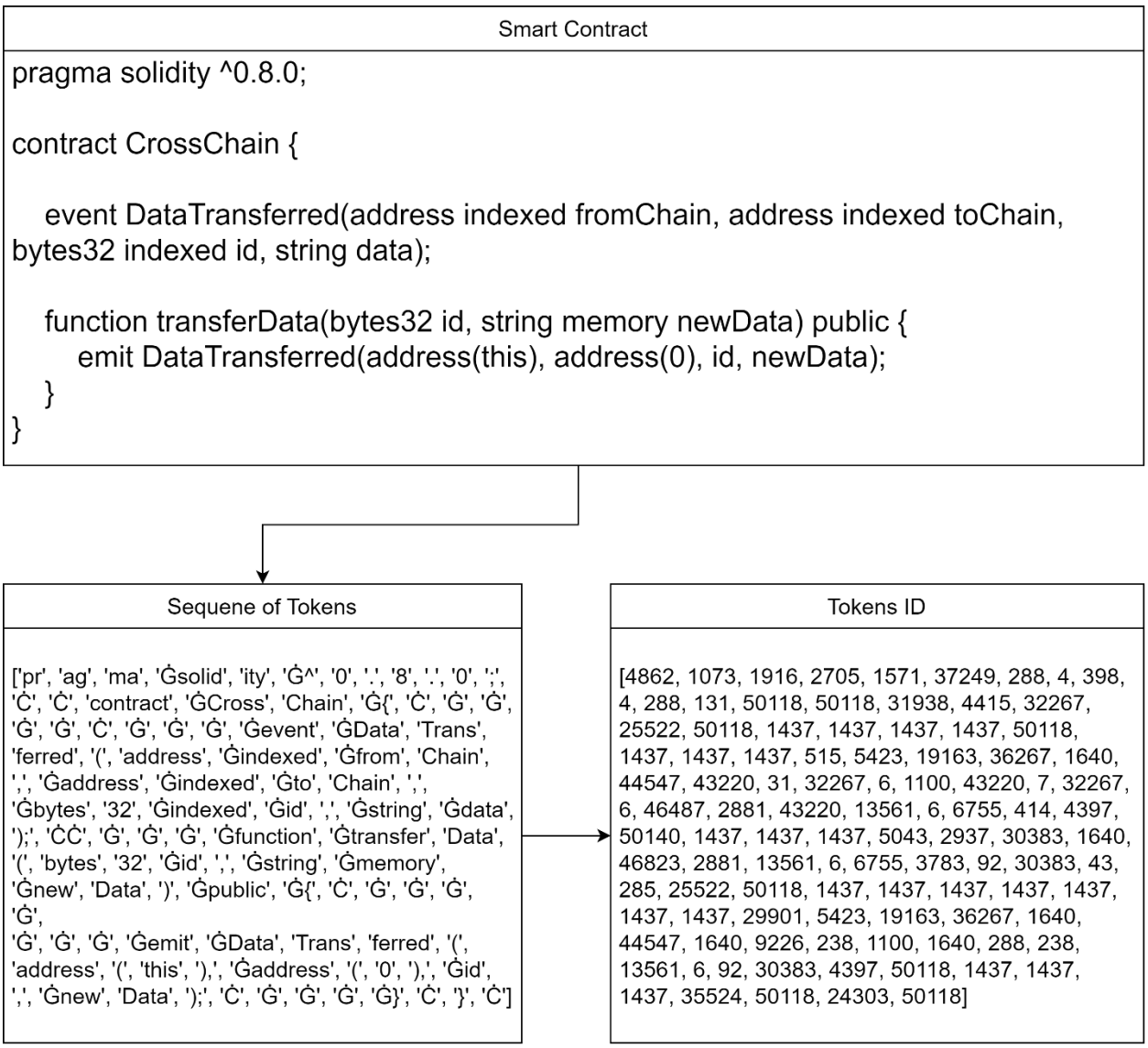


Hình 5: Phân bố chi tiết các mẫu

Tập dữ liệu cdhứa hai loại nhãn để phân loại hợp đồng thông minh. Tập nhãn đầu tiên là một tập nhãn nhị phân, xác định xem một hợp đồng có đặc điểm an toàn hay nguy hiểm. Tập nhãn thứ hai chi tiết hóa các nhãn đa lớp, phân loại các hợp đồng độc hại thành các loại lỗ hổng cụ thể như an toàn, có khả năng bị Reentrancy, Interger Overflow/Underflow và Unprotected Ether Withdrawal. Dữ liệu được xác định bằng các đặc điểm như tên file dự án, ID commit, nhãn mục tiêu, cấu trúc

và nội dung của hợp đồng, cũng như số thứ tự file. Trong tập nhãn nhị phân, giá trị 0 biểu thị an toàn và 1 là nguy hiểm. Trong tập nhãn đa lớp, giá trị 0 là an toàn, 1 là lỗ hổng Reentrancy, 2 là lỗ hổng Integer Overflow/Underflow và 3 là lỗ hổng Unprotected Ether Withdrawal.

Để xử lý dữ liệu, em bắt đầu bằng việc chuyển đổi các file mã nguồn hợp đồng thông minh trở thành các token thông qua quá trình tokenization. Mỗi file code smart contract được phân chia thành các token, đơn vị nhỏ nhất của mã nguồn, như từ vựng, ký tự đặc biệt, hoặc biểu thức. Việc này giúp em biểu diễn cấu trúc và ý nghĩa của mã nguồn một cách có cấu trúc. Tiếp theo, mỗi token được chuyển đổi thành một tokenID, là một số nguyên duy nhất đại diện cho loại cụ thể của token đó. Quá trình này giúp giảm kích thước của dữ liệu và tạo ra một biểu diễn số hóa hiệu quả cho mã nguồn hợp đồng. Quá trình này tối ưu hóa việc đưa dữ liệu vào mô hình học máy, giúp mô hình có khả năng "nắm bắt" cấu trúc và ngữ cảnh của mã nguồn hợp đồng thông minh. Từ đó, em đã phát triển một bộ các tokenID, trong đó mỗi tokenID tương ứng với một phần cụ thể của mã nguồn. Điều này cho phép em tạo ra các vector đặc trưng cho mỗi hợp đồng thông minh, và cuối cùng, sử dụng chúng để đưa vào mô hình học máy nhằm phân loại các lỗ hổng một cách chính xác.



Hình 6: Quá trình xử lý dữ liệu

Về việc ứng dụng các mô hình học máy trong việc phát hiện lỗ hổng:

Việc chọn các phương pháp học máy bởi chúng cho phép tự động phát hiện lỗ hổng bảo mật trong các hợp đồng thông minh bằng cách phân biệt các mẫu trong mã hợp đồng và cấu trúc. So với kiểm toán thủ công, các mô hình học máy có thể phân tích hợp đồng một cách hiệu quả và nhất quán hơn. Em thử nghiệm vài mô hình học máy cổ điển, bao gồm Cây quyết định, Rừng ngẫu nhiên, Máy vector hỗ trợ, XGBoost và Hồi quy Logistic. Nhờ khả năng giải thích, các mô hình này cho phép hiểu được lý do tại sao một số hợp đồng bị gắn cờ là dễ bị tấn công.

Decision tree và Random forest xây dựng các quy tắc phân cấp dựa trên các thuộc tính mã để phân loại hợp đồng. Trong khi đó, SVM tìm ranh giới tối ưu giữa hợp đồng dễ bị tấn công và an toàn trong không gian đặc trưng. XGBoost tiếp tục tăng độ chính xác thông qua một tập hợp các học viên yếu. Em trích xuất các đặc trưng số liệu có giá trị thông tin như độ phức tạp hàm, luồng điều khiển và các mẫu cú pháp.

Bằng cách huấn luyện trên dữ liệu đại diện thích hợp của cả ví dụ có chiều hướng tích cực lẫn chiều hướng tiêu cực, các mô hình học cách phát hiện bền vững các lỗ hổng an toàn thông tin. Em áp dụng các kỹ thuật này trong ChainSniper để xây dựng một máy quét tự động cho các lỗ hổng như Reentrancy, Interger Overflow/Underflow và Unprotected Ether Withdrawal. Tính linh hoạt của học máy cung cấp một phương pháp luận để phát hiện lỗ hổng trong khi tránh nỗ lực thủ công mở rộng.

Về việc ứng dụng các mô hình học sâu trong việc phát hiện lỗ hổng:

Quá trình lựa chọn các phương pháp học sâu được thực hiện vì khả năng của chúng trong việc mô hình hóa các mối quan hệ phức tạp. của logic mã và phụ thuộc mà không cần trích xuất thủ công các đặc trưng. Các kỹ thuật như CNN, RNN, LSTM và các mô hình Transformer như RoBERTa cho phép học từ đầu đến cuối trực tiếp từ mã nguồn hợp đồng thông minh thô để phát hiện các mẫu cú pháp và ngữ nghĩa phức tạp mà đặc trưng cho các lỗ hổng.

Việc thử nghiệm các mạng nơ-ron sâu bao gồm các mô hình tuần tự dựa trên LSTM có khả năng diễn giải các luồng điều khiển, CNN trích xuất các mẫu cú pháp cục bộ và bộ mã hóa Transformer như RoBERTa cung cấp các vector từ ngữ cảnh phong phú. Nhờ xử lý phân cấp theo tầng, các mạng này tự động học quan hệ tiềm ẩn giữa các token và cấu trúc dẫn đến sự hiểu biết về logic cho thấy lỗ hổng bảo mật.

Hơn thế nữa, các phụ thuộc dài hạn được mô hình hóa bởi LSTM có thể truy tìm các luồng thông tin trong hợp đồng để xác định các vấn đề kiểm tra cuộc gọi không kiểm soát. Các vector ngữ cảnh của RoBERTa có khả năng phân biệt các sắc thái giữa các cấu trúc có vẻ tương tự nhưng có thể hoặc không dễ bị tấn công. Các lớp chú ý có thể giải thích được của học sâu cũng hỗ trợ xác định các thành phần gây vấn đề. Việc áp dụng các kỹ thuật này trong ChainSniper để xây dựng các máy quét tự động có thể cờ các nguy cơ Reentrancy, Interger Overflow/Underflow và Unprotected Ether Withdrawal.

B2. Mục tiêu, nội dung, kế hoạch nghiên cứu

B2.1 Mục tiêu

Mục tiêu đề tài:

- Triển khai được mô hình cầu nối dựa trên phương pháp SideChain và tiến hành được quá trình chuyển đổi dữ liệu qua các mạng chuỗi khối.
- Xây dựng tập dữ liệu CrossChainSentinel bao gồm các mẫu lành tính và các mẫu độc hại được gán nhãn thủ công, đồng thời tiến hành xử lý dữ liệu.
- Ứng dụng các mô hình học máy và học sâu trong việc phát hiện các mẫu độc hại một cách tự động, đánh giá và nhận xét các mô hình.

B2.2 Nội dung và phương pháp nghiên cứu

Nội dung 1: Triển khai mô hình cầu nối bằng phương pháp Sidechain và tiến hành chuyển đổi dữ liệu

Mục tiêu:

- Triển khai phương pháp sidechain để kết nối các blockchain khác nhau và cho phép chuyển dữ liệu giữa chúng.
- Mô tả quá trình chuyển đổi dữ liệu giữa các blockchain thông qua cầu nối sidechain.

Phương pháp:

- Triển khai hệ thống với 2 blockchain blockchain A – Ethereum và blockchain B – Avalanche.
- Sử dụng sidechain làm cầu nối để chuyển dữ liệu giữa 2 blockchain trên.
- Ghi lại log giao dịch và dữ liệu hợp đồng thông minh trong quá trình chuyển đổi giữa các chuỗi.

Kết quả đạt được:

- Xây dựng thành công mô hình cầu nối sidechain để kết nối và chuyển dữ liệu giữa 2 blockchain khác nhau.
- Thu thập được dữ liệu log giao dịch và hợp đồng thông minh trong quá trình chuyển đổi giữa các chuỗi blockchain.
- Dữ liệu thu thập được sẽ được sử dụng để huấn luyện mô hình máy học phát hiện lỗ hổng bảo mật trong hợp đồng thông minh liên chuỗi.

Nội dung 2: Xây dựng tập dữ liệu mới được đặt tên là "CrossChainSentinel", bao gồm 300 mẫu hợp đồng thông minh được gán nhãn thủ công

Mục tiêu:

- Xây dựng một tập dữ liệu mới về hợp đồng thông minh liên chuỗi (cross-chain smart contracts) để sử dụng trong việc huấn luyện mô hình phát hiện lỗ hổng bảo mật.
- Tập dữ liệu cần bao gồm cả mẫu hợp đồng thông minh an toàn và có lỗ hổng bảo mật.

Phương pháp:

- Tổng cộng thu thập 300 mẫu hợp đồng thông minh.
- Các mẫu được gán nhãn thủ công để phân loại là hợp đồng an toàn hay có lỗ hổng bảo mật.

- Trong số 300 mẫu, có 158 mẫu an toàn và 142 mẫu có lỗ hổng bao gồm: 42 mẫu lỗi reentrancy, 48 mẫu lỗi tràn số nguyên, 52 mẫu lỗi rút Ether không được bảo vệ.
- Các mẫu được lấy từ các nguồn thực tế như Smartbugs-wild, SolidiFi-benchmark, và địa chỉ Ethereum liên quan đến các cuộc tấn công đa chuỗi.

Kết quả đạt được:

- Tập dữ liệu CrossChainSentinel với 300 mẫu hợp đồng thông minh liên chuỗi đã được xây dựng thành công.
- Tập dữ liệu bao gồm cả mẫu an toàn và có lỗ hổng, với các loại lỗ hổng phổ biến như tái nhập, tràn số, rút Ether không bảo vệ.
- Dữ liệu được gán nhãn bởi chuyên gia, đảm bảo chất lượng và sự đa dạng của tập dữ liệu.

Nội dung 3: Ứng dụng các mô hình học máy và học sâu vào việc phát hiện lỗ hổng, tiến hành đánh giá kết quả

Mục tiêu:

- Ứng dụng các mô hình học máy (machine learning) và học sâu (deep learning) để phát hiện lỗ hổng bảo mật trong hợp đồng thông minh liên chuỗi.
- Đánh giá và so sánh hiệu suất của các mô hình khác nhau trong việc phát hiện lỗ hổng.

Phương pháp:

- Sử dụng tập dữ liệu CrossChainSentinel để huấn luyện các mô hình học máy và học sâu khác nhau.
- Các mô hình học máy được áp dụng bao gồm: Cây quyết định, Rừng ngẫu nhiên, XGBoost, Máy vector hỗ trợ (SVM), Hồi quy logistic.
- Các mô hình học sâu được áp dụng bao gồm: Mạng CNN, LSTM, Mạng Neural tích chập (FFN), RoBERTa.
- Chia dữ liệu thành tập huấn luyện và kiểm tra với tỷ lệ 80/20.
- Sử dụng các chỉ số đánh giá như Độ chính xác, Độ đầy đủ, Độ phủ, F1-score để đo lường hiệu suất của các mô hình.

Kết quả đạt được:

- Đánh giá và nhận xét các mô hình học máy và học sâu trong việc phát hiện các mẫu độc hại một cách tự động, các mô hình.
- Kiểm tra về thời gian thực thi trong việc phát hiện lỗ hổng một cách tự động.

B2.3 Kế hoạch nghiên cứu.

Tháng đầu tiên: Hoàn thành dàn ý cho ý tưởng, nghiên cứu các công trình liên quan, đưa ra mô hình tổng quan. Hoàn thành thực nghiệm triển khai cầu nối và thu thập các mẫu tấn công trên cầu nối
Tháng thứ 2 và tháng thứ 3: Hoàn thành thu thập dữ liệu nhằm dựng file dataset
Tháng thứ 4 và tháng thứ 5: Hoàn thành prediction model, đưa ra số liệu đánh giá, tối ưu hoá mô hình
Tháng thứ 6: Hoàn thành báo cáo, hiệu chỉnh những thông tin cần thiết

B3. Kết quả dự kiến

Sau khi thực hiện đề tài, kết quả dự kiến đạt được như sau:

- Hoàn tất quá trình triển khai được mô hình cầu nối dựa trên phương pháp SideChain và tiến hành được quá trình chuyển đổi dữ liệu qua các mạng chuỗi khối
- Thành công trong việc tập dữ liệu CrossChainSentinel bao gồm các mẫu lành tính và các mẫu độc hại được gán nhãn thủ công, đồng thời tiến hành xử lý dữ liệu.
- Đánh giá và nhận xét các mô hình học máy và học sâu trong việc phát hiện các mẫu độc hại một cách tự động, các mô hình.

B4. Tài liệu tham khảo

[1] P. Patel and H. Patel, “Achieving a secure cloud storage mechanism using blockchain technology,” p. 130–142, 2023. [Online]. Available: http://dx.doi.org/10.7763/IJCTE.2023.V15.1342

[2] M. Kumarathunga, R. N. Calheiros, and A. Ginige, “Sustainable microfinance outreach for farmers with blockchain cryptocurrency and smart contracts,” p. 9–14, 2022. [Online]. Available: http://dx.doi.org/10.7763/IJCTE.2022.V14.1304

[3] T. Xie, J. Zhang, Z. Cheng, F. Zhang, Y. Zhang, Y. Jia, D. Boneh, and D. Song, “zkbridge: Trustless cross-chain bridges made practical,” in Pro- ceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, 2022, pp. 3003–3017.

[4] P. Han, Z. Yan, W. Ding, S. Fei, and Z. Wan, “A survey on cross-chain technologies,” Distributed Ledger Technologies: Research and Practice, vol. 2,no. 2, pp. 1–30, 2023.

[5] Y. Hei, D. Li, C. Zhang, J. Liu, Y. Liu, and Q. Wu, “Practical agentchain: A compatible cross-chain exchange system,” Future Generation Computer Systems, vol. 130, pp. 207–218, 2022.

[6] R. Lan, G. Upadhyaya, S. Tse, and M. Zamani, “Horizon: A gas-efficient, trustless bridge for cross-chain transactions,” arXiv preprintarXiv:2101.06000, 2021.

[7] K. Qin and A. Gervais, “An overview of blockchain scalability, interoperability and sustainability,” Hochschule Luzern Imperial College London Liquidity Network, pp. 1–15, 2018.

[8] T. Hardjono, “Blockchain gateways, bridges and delegated hash-locks,” arXiv preprint arXiv:2102.03933, 2021.

[9] A. Singh, K. Click, R. M. Parizi, Q. Zhang, A. Dehghantanha, and K.-K. R. Choo, “Sidechain technologies in blockchain networks: An examination and state-of-the-art review,” Journal of Network and Computer Applications, vol. 149, p. 102471, 2020.

[10] M. H. Miraz and D. C. Donald, “Atomic cross-chain swaps: development, trajectory and potential of non-monetary digital token swap facilities,”arXiv preprint arXiv:1902.04471, 2019.

[11] J. Zhang, J. Gao, Y. Li, Z. Chen, Z. Guan, and Z. Chen, “Xscope: Hunting for cross-chain bridge attacks,” in Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering, 2022, pp. 1–4.

[12] S. Sayeed, H. Marco-Gisbert, and T. Caira, “Smart contract: Attacks and protections,” IEEE Access, vol. 8, pp. 24 416–24 427, 2020.

[13] N. Atzei, M. Bartoletti, and T. Cimoli, “A survey of attacks on ethereum smart contracts (sok),” in Principles of Security and Trust: 6th International Conference, POST 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings 6. Springer, 2017, pp. 164–186.

[14] S. Wang, Y. Yuan, X. Wang, J. Li, R. Qin, and F.-Y. Wang, “An overview of smart contract: architecture, applications, and future trends,” in 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018, pp. 108–113.

[15] T. H.-D. Huang, “Hunting the ethereum smart contract: Color-inspired inspection of potential attacks,” arXiv preprint arXiv:1807.01868, 2018.

[16] L. Zhang, W. Chen, W. Wang, Z. Jin, C. Zhao, Z. Cai, and H. Chen, “Cbgru: A detection method of smart contract vulnerability based on a hybrid model,” Sensors, vol. 22, no. 9, p. 3577, 2022.

[17] E. Lai and W. Luo, “Static analysis of integer overflow of smart contracts in ethereum,” in Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy, 2020, pp. 110–115.

[18] M. Staderini, C. Palli, and A. Bondavalli, “Classification of ethereum vulnerabilities and their propagations,” in 2020 Second International Conference on Blockchain Computing and Applications (BCCA). IEEE, 2020, pp. 44–51.

[19] H. Mao, T. Nie, H. Sun, D. Shen, and G. Yu, “A survey on cross-chain technology: Challenges, development, and prospect,” IEEE Access, 2022.

[20] M. Rodler, W. Li, G. O. Karame, and L. Davi, “Sereum: Protecting existing smart contracts against re-entrancy attacks,” arXiv preprint arXiv:1812.05934, 2018.

[21] P. Praitheeshan, L. Pan, J. Yu, J. Liu, and R. Doss, “Security analysis methods on ethereum smart contract vulnerabilities: a survey,” arXiv preprint arXiv:1908.08605, 2019.

[22] S. S. Kushwaha, S. Joshi, D. Singh, M. Kaur, and H.-N. Lee, “Systematic review of security vulnerabilities in ethereum blockchain smart contract,” IEEE Access, vol. 10, pp. 6605–6621, 2022.

[23] J.-W. Liao, T.-T. Tsai, C.-K. He, and C.-W. Tien, “Soliaudit: Smart contract vulnerability assessment based on machine learning and fuzz testing,” in 2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS). IEEE, 2019, pp. 458–465.

[24] S. Badillo, B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. Kam-Thong, J. Siebourg-Polster, B. Steiert, and J. D. Zhang, “An introduction to machine learning,” Clinical pharmacology & therapeutics, vol. 107, no. 4,pp. 871–885, 2020.

[25] M. Krichen, “Strengthening the security of smart contracts through the power of artificial intelligence,” Computers, vol. 12, no. 5, p. 107, 2023.

*Ngày __ tháng __ năm 20_*	*Ngày __ tháng __ năm 20_*
Giảng viên hướng dẫn	**Chủ nhiệm đề tài**
(Ký và ghi rõ họ tên)	(Ký và ghi rõ họ tên)