(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ
(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ
(11) **1-0026304**
(51)[7] **G06F 21/00**
(13) **B**

(21) 1-2013-04148
(22) 26/12/2012
(86) PCT/CN2012/087530 26/12/2012
(87) WO 2013/097718 A1 04/07/2013
(30) 201110445277.1 27/12/2011 CN
(45) 25/11/2020 392
(43) 27/10/2014 319A
(73) TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED (CN)
Room 403, East Block 2, SEG Park, Zhenxing Road, Futian District, Shenzhen, Guangdong 518044, P.R. China
(72) YUAN, Xiaohui (CN); LONG, Hai (CN); LI, Shuai (CN).
(74) Công ty TNHH Sở hữu trí tuệ WINCO (WINCO CO., LTD.)

(54) PHƯƠNG PHÁP VÀ THIẾT BỊ PHÁT HIỆN MÃ ĐỘC HẠI TRÊN CÁC TRANG WEB

(57) Sáng chế đề cập đến phương pháp phát hiện mã độc hại trên các trang web bao gồm các bước: thu nhận danh sách hàm bằng cách thi hành mã xác định và mã đối tượng định trước; phân tích cú pháp mã xác định và thu nhận các giá trị biến dựa vào kết quả phân tích cú pháp và danh sách hàm; và xác định xem mã độc hại có xuất hiện trên các trang web hay không dựa vào các giá trị biến. Sáng chế còn đề cập đến thiết bị phát hiện mã độc hại trên các trang web.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực công nghệ web, và cụ thể hơn sáng chế đề cập đến phương pháp và thiết bị phát hiện mã độc hại trên các trang web.

Tình trạng kỹ thuật của sáng chế

Với sự phát triển liên tục của công nghệ thông tin, mọi người quen với việc thu thập thông tin năng động về mọi vấn đề bằng cách duyệt các trang web. Công nghệ web, là một trong số các công nghệ chia sẻ thông tin quan trọng, có thể cung cấp rất nhiều thông tin cho người dùng.

Tuy nhiên, do thiếu các tính năng tương tác, khả năng tái sử dụng kém, và các vấn đề trong việc bảo trì của các trang web tĩnh ban đầu, cho nên các công nghệ web động dần dần được phát triển, và một trong số đó có ngôn ngữ lập trình VBScript (Visual Basic Script).

Ngôn ngữ lập trình VBScript có thể được sử dụng để điều hướng trình duyệt trên máy khách, triển khai động ngôn ngữ đánh dấu siêu văn bản (HyperText Markup Language, HTML), và thậm chí còn kết hợp chương trình bên ngoài vào các trang web. Tuy nhiên, do thiếu an toàn, nên kẻ tấn công với ý định xấu có thể phát tán mã độc hại trên các trang web, tải xuống phần mềm độc hại Trojan, tấn công máy chủ người dùng và truy nhập thông tin người dùng thông qua các lỗ hổng của kỹ thuật lập trình VBScript.

Hiện nay, một trong số các phương thức để phát hiện mã VBScript độc hại là chuyển đổi từ mã VBScript sang mã JavaScript và sau đó phân tích cú pháp mã JavaScript bằng cách sử dụng trình thông dịch của ngôn ngữ lập trình JavaScript (JavaScript scripting engine). Tuy nhiên, phương thức này có lỗ hổng, tức là mã VBScript không thể được chuyển đổi tương đương sang mã JavaScript và mã JavaScript đã chuyển đổi có thể có các hàm ngữ nghĩa sai khác so với các hàm ngữ nghĩa của mã VBScript ban đầu. Do đó, có thể đưa ra các kết quả kiểm tra không chính xác.

Bản chất kỹ thuật của sáng chế

Một phương án thực hiện sáng chế đề xuất phương pháp phát hiện mã độc hại trên

các trang web, phương pháp này bao gồm các bước:

thu nhận danh sách hàm bằng cách thi hành mã xác định và mã đối tượng định trước;

phân tích cú pháp mã xác định và thu nhận các giá trị biến dựa vào kết quả phân tích cú pháp và danh sách hàm; và

xác định xem mã độc hại có xuất hiện trên các trang web hay không dựa vào các giá trị biến.

Một phương án khác để thực hiện sáng chế đề xuất thiết bị phát hiện mã độc hại trên các trang web, thiết bị này bao gồm:

môđun thu nhận danh sách hàm được tạo cấu hình để thu nhận danh sách hàm bằng cách thi hành mã xác định và mã đối tượng định trước; và

môđun phân tích cú pháp và tách được tạo cấu hình để phân tích cú pháp mã xác định và thu nhận các giá trị biến dựa vào kết quả phân tích cú pháp và danh sách hàm, trong đó việc mã độc hại có xuất hiện trên các trang web hay không được xác định dựa vào các giá trị biến.

Các phương án thực hiện sáng chế mô tả phương pháp và thiết bị phát hiện mã độc hại trên các trang web. Nhờ vào việc tuần tự thu nhận danh sách hàm bằng cách thi hành mã VBScript và mã đối tượng định trước và thu nhận các giá trị biến bằng cách phân tích cú pháp mã VBScript và kết quả phân tích cú pháp, mã script độc hại trên các trang web có thể được phát hiện trước và sau đó, hệ thống liên quan có thể chặn mã VBScript độc hại này và nhắc người dùng nếu phát hiện thấy mã VBScript độc hại; do đó, quyền của người dùng được bảo đảm và người dùng có thể duyệt các trang web với độ an toàn được nâng cao.

Mô tả vắn tắt các hình vẽ

Fig.1 là lưu đồ thể hiện phương pháp phát hiện mã độc hại trên các trang web theo phương án ưu tiên để thực hiện sáng chế;

Fig.2 là lưu đồ thể hiện quy trình thu nhận danh sách hàm bằng cách thi hành mã xác định và mã đối tượng định trước trong phương pháp phát hiện mã độc hại trên các

trang web theo phương án ưu tiên để thực hiện sáng chế;

Fig.3 là lưu đồ thể hiện quy trình phân tích cú pháp mã xác định và nhờ đó thu được các giá trị biến dựa vào các kết quả phân tích cú pháp và danh sách hàm trong phương pháp phát hiện mã độc hại trên các trang web theo phương án ưu tiên để thực hiện sáng chế;

Fig.4 là lưu đồ thể hiện quy trình mở rộng mã dựa vào danh sách hàm và thông tin thủ tục hàm trong phương pháp phát hiện mã độc hại trên các trang web theo phương án ưu tiên để thực hiện sáng chế;

Fig.5 là sơ đồ cấu trúc của thiết bị phát hiện mã độc hại trên các trang web theo phương án ưu tiên để thực hiện sáng chế;

Fig.6 là sơ đồ cấu trúc của môđun thu nhận danh sách hàm trong thiết bị phát hiện mã độc hại trên các trang web theo phương án ưu tiên để thực hiện sáng chế;

Fig.7 là sơ đồ cấu trúc của môđun phân tích cú pháp và tách trong thiết bị phát hiện mã độc hại trên các trang web theo phương án ưu tiên để thực hiện sáng chế; và

Fig.8 là sơ đồ cấu trúc của bộ phận mở rộng trong thiết bị phát hiện mã độc hại trên các trang web theo phương án ưu tiên để thực hiện sáng chế.

Mô tả chi tiết sáng chế

Các mục đích, giải pháp kỹ thuật và ưu điểm của sáng chế sẽ được hiểu rõ hơn sau khi xem phần mô tả sáng chế dưới đây dựa vào các hình vẽ và các phương án thực hiện sáng chế. Cần phải hiểu rằng các phương án thực hiện sáng chế được sử dụng để mô tả sáng chế chứ không phải là để giới hạn phạm vi của sáng chế.

Giải pháp chính được đề xuất trong các phương án thực hiện sáng chế là thu nhận danh sách hàm bằng cách thi hành mã script và mã đối tượng định trước, phân tích cú pháp mã script này, tách ra các giá trị biến dựa vào các kết quả phân tích cú pháp và danh sách hàm, và kiểm tra các giá trị biến. Do đó, các trang web chứa mã script độc hại có thể được phát hiện trước và nhờ đó nâng cao độ an toàn cho người dùng duyệt các trang web.

Mã được đề cập trong sáng chế là mã script, và cụ thể là mã VBScript hoặc các loại mã script khác. Do đó, mỗi phương án dưới đây được mô tả bằng cách sử dụng mã

VBScript.

Theo phương pháp thông thường để phát hiện mã VBScript độc hại có thể xuất hiện trên trang web, trước tiên là chuyển đổi từ mã VBScript sang mã JavaScript và sau đó phân tích cú pháp mã JavaScript. Để giải quyết vấn đề là tốc độ chuyển đổi tương đối chậm trong phương pháp thông thường này, phương án thực hiện sáng chế sử dụng trình thông dịch MSScript chạy trên nền sử dụng hệ điều hành Windows để phát hiện mã VBScript độc hại. Cụ thể là, mã VBScript được thi hành thông qua trình thông dịch MSScript sao cho các thông tin như thông tin biến và thông tin hàm có thể được tách ra từ mã VBScript. Sau đó, thông tin đã tách ra được nhập vào bộ phận tách đặc điểm để tách ra các biến toàn cục từ mã VBScript. Ngoài ra, quy trình mở rộng có thể được thực hiện theo phương án thực hiện sáng chế để tách ra các biến cục bộ, các biến cục bộ này có thể xuất hiện trong thông tin hàm và không thể phát hiện được bằng cách xử lý các biến toàn cục.

Fig.1 thể hiện tóm tắt phương pháp phát hiện mã độc hại trên các trang web theo phương án ưu tiên để thực hiện sáng chế. Phương pháp này bao gồm các bước sau đây.

Ở bước S101, danh sách hàm được thu nhận bằng cách thi hành mã xác định và mã đối tượng định trước.

Trong sáng chế, mã VBScript được sử dụng để làm ví dụ. Trước tiên, đối tượng trình duyệt (Browser) và đối tượng môđun đối tượng tài liệu (Document Object Module, DOM) thường dùng, ví dụ đối tượng Navigator, đối tượng Document và đối tượng Object, được ưu tiên định nghĩa trước để tránh vấn đề bất định có thể xảy ra và dẫn đến lỗi thi hành khi đối tượng trình duyệt (Browser) và đối tượng môđun đối tượng tài liệu (DOM) được chèn trực tiếp vào trình thông dịch MSScript.

Sau đó, mã VBScript và mã đối tượng định trước được thi hành bằng cách gọi phương pháp thi hành mã, ví dụ phương pháp ExecuteStatement, được cung cấp thông qua giao diện lập trình script IScriptControl.

Sau khi các mã được thi hành thành công, phương pháp thu nhận danh sách tên thủ tục, ví dụ phương pháp GetProcedures, được cung cấp thông qua giao diện lập trình script IScriptControl được gọi để thu nhận danh sách tên thủ tục (hàm), và phương pháp

thu nhận danh sách biến, ví dụ phương pháp GetCodeObject, được cung cấp thông qua giao diện lập trình script IscriptControl được gọi để thu được con trỏ giao diện IDispatch. Sau đó, danh sách biến toàn cục trong mã VBScript được thu nhận bằng cách sử dụng cơ chế ánh xạ COM, trong đó danh sách tên thủ tục và danh sách biến toàn cục được gọi là danh sách hàm thu được. Sau đó, bước 102 và bước 103 được thực hiện để phân tích cú pháp mã VBScript và thu nhận các giá trị biến dựa vào kết quả phân tích cú pháp và danh sách hàm, và kiểm tra các giá trị biến.

Ở bước S102, mã xác định được phân tích cú pháp để thu được các giá trị biến dựa vào kết quả phân tích cú pháp và danh sách hàm.

Ở bước 103, các giá trị biến được kiểm tra. Sau đó, việc mã độc hại có xuất hiện trên các trang web hay không có thể được xác định dựa vào các giá trị biến đã kiểm tra.

Ở các bước S102, S103, các thông tin thủ tục hàm chi tiết như danh sách thông số của hàm và tập hợp câu lệnh (function body), được thu nhận bằng cách phân tích cú pháp mã VBScript ban đầu sau khi danh sách hàm được thu nhận, và sau đó một mã VBScript mới được thu nhận bằng cách thực hiện quy trình tinh chỉnh thủ tục hàm trên mã VBScript ban đầu để loại bỏ hoàn toàn tất cả các thủ tục hàm ra khỏi mã VBScript ban đầu. Mục đích của việc thực hiện quy trình loại bỏ thủ tục hàm trên mã VBScript ban đầu là để thi hành mã VBScript mở rộng trong trình thông dịch MSScript và nhờ đó tách ra các giá trị biến có mặt ở trong đó.

Trong khi đó, danh sách biến cục bộ được thu nhận bằng cách tuần tự gọi phương pháp ExecuteStatement và phương pháp GetCodeObject được cung cấp thông qua giao diện lập trình script IScriptControl đối với mỗi hàm dựa vào các thông tin thủ tục hàm chi tiết trong mã VBScript thu được. Vì mã thi hành độc hại thường xuất hiện trong các biến cục bộ, nên tốt hơn là phải tách ra và kiểm tra các biến cục bộ để xác định chính xác xem có còn mã thi hành độc hại nào nữa không. Các biến cục bộ có thể được tách ra và kiểm tra bằng bộ phận tách đặc điểm.

Thông qua quy trình nêu trên, tất cả các thông tin cơ bản cần thiết để mở rộng mã VBScript được thu nhận. Để nâng cao hơn nữa hiệu quả của quy trình mở rộng mã VBScript, phương án thực hiện sáng chế đề xuất sử dụng bảng quan hệ phụ thuộc giữa các hàm, qua đó một hàm có thể được mở rộng theo cấu trúc phân cấp và nhờ đó nâng

cao hiệu quả của quy trình mở rộng.

Cụ thể là, bảng quan hệ phụ thuộc hai chiều biểu thị quan hệ phụ thuộc giữa các hàm được tạo ra bằng cách phân tích quan hệ gọi đối với mỗi hàm. Trong sáng chế, quan hệ phụ thuộc được biểu diễn dưới dạng sự phụ thuộc ngược, ví dụ, như sau:

Đối với các hàm A, B, C, D, E, F và G, có quan hệ gọi hàm: các hàm B, D và G có thể được gọi bằng hàm A; các hàm C, E và G có thể được gọi bằng hàm B; và các hàm F và G có thể được gọi bằng hàm E.

Do đó, bảng quan hệ phụ thuộc hai chiều có thể được thiết lập như sau:

A→NIL;

B→A;

C→B;

D→A;

E→B;

F→E;

G→A, B, E.

Đối với mỗi hàm, quy trình mở rộng chủ yếu được xây dựng dựa vào bảng quan hệ phụ thuộc giữa các hàm; do đó, bộ phận chọn mở rộng hàm được đưa vào sử dụng và được thiết kế để trả về hàm cần mở rộng kế tiếp. Cụ thể là, bộ phận chọn mở rộng hàm được tạo cấu hình để quét ngang qua danh sách hàm hiện thời để thu nhận hàm thứ nhất không được mở rộng và có quan hệ phụ thuộc giữa các hàm là NIL, bộ phận chọn mở rộng hàm này trả về hàm cần mở rộng kế tiếp, và tuần tự mở rộng mỗi hàm cần mở rộng trong danh sách hàm.

Đối với các hàm A, B, C, D, E, F và G nêu trên, quy trình mở rộng được mô tả để làm ví dụ như sau:

1. Mở rộng hàm A nếu hàm A không phụ thuộc vào bất kỳ hàm nào khác;

2. Mở rộng, sau khi hàm A được mở rộng và do đó các quan hệ phụ thuộc giữa các hàm B và D là NIL, mỗi hàm trong số hàm B hoặc hàm D kế tiếp sau hàm A (trong ví dụ này chọn trước tiên quét hàm B);

3. Mở rộng, sau khi hàm B được mở rộng và do đó các quan hệ phụ thuộc giữa các

hàm C, D và E là NIL, hàm C kế tiếp sau hàm B;

4. Tuần tự mở rộng các hàm D và E kế tiếp sau hàm C;

5. Tuần tự mở rộng, sau khi hàm E được mở rộng và do đó các quan hệ phụ thuộc giữa các hàm F và G là NIL, các hàm F và G.

Đối với mỗi hàm, quy trình mở rộng chủ yếu được thực hiện bằng cách tìm hàm cần gọi, xây dựng một tập hợp câu lệnh mới, và thực hiện việc thay thế. Việc xây dựng một tập hợp câu lệnh mới được thực hiện bằng cách đặt lại tên cho các thông số của hàm và các biến cục bộ của hàm dưới dạng là tên-hàm_tên-biến (tên-thông-số)_ID-gọi. Ngoài ra, các thông số ở phần trước của tập hợp câu lệnh có dạng biến cục bộ, và giá trị ước tính tương ứng với các thông số được đưa vào trong lúc gọi được kết hợp vào trong các biến. Giá trị ID gọi biểu thị số lần gọi của hàm phát hiện được hiện thời, giá trị này được thực hiện để ngăn chặn tình trạng xung đột biến xảy ra do nhiều lần gọi và mở rộng hàm.

Sau khi quy trình mở rộng tất cả các hàm đã hoàn thành, một mã VBScript mới được thu nhận.

Mã VBScript mới thu được sau khi hoàn thành quy trình mở rộng hàm được nhập vào và thi hành trong trình thông dịch MSScript. Một danh sách gồm tất cả các giá trị biến được thu nhận theo cơ chế ánh xạ giao diện COM, và sau đó các giá trị biến thu được được nhập vào bộ phận tách đặc điểm để tách ra và kiểm tra nhằm hoàn thành việc phát hiện mã VBScript độc hại.

Như được thể hiện trên Fig.2, trong quy trình thực hiện nêu trên sử dụng mã VBScript để làm ví dụ, bước S101 còn bao gồm:

Bước S1011: thi hành mã VBScript và mã đối tượng định trước bằng cách gọi phương pháp ExecuteStatement được cung cấp thông qua giao diện lập trình script;

Bước S1012: thu nhận danh sách tên thủ tục trong mã VBScript bằng cách gọi phương pháp GetProcedures được cung cấp thông qua giao diện lập trình script;

Bước S1013: thu nhận con trỏ giao diện IDispatch bằng cách gọi phương pháp GetCodeObject được cung cấp thông qua giao diện lập trình script và thu nhận danh sách biến toàn cục trong mã VBScript bằng cách sử dụng cơ chế ánh xạ COM.

Như được thể hiện trên Fig.3, bước S102 bao gồm:

Bước S1021: thu nhận thông tin thủ tục hàm bằng cách phân tích cú pháp mã xác định;

Bước S1022: mở rộng mã xác định dựa vào danh sách hàm và thông tin thủ tục hàm;

Bước S1023: tách ra các giá trị biến bằng cách thi hành mã xác định mở rộng.

Như được thể hiện trên Fig.4, bước S1022 bao gồm:

Bước S10221: thu nhận quan hệ gọi đối với mỗi hàm dựa vào thông tin thủ tục hàm;

Bước S10222: tạo ra bảng quan hệ phụ thuộc hai chiều dựa vào quan hệ gọi đối với mỗi hàm;

Bước S10223: mở rộng mã VBScript dựa vào danh sách hàm và bảng quan hệ phụ thuộc hai chiều.

Bằng cách quét ngang qua danh sách hàm, bộ phận chọn mở rộng hàm thu nhận hàm thứ nhất không được mở rộng và có quan hệ phụ thuộc giữa các hàm là NIL, bộ phận chọn mở rộng hàm này trả về hàm cần mở rộng kế tiếp, và tuần tự mở rộng mỗi hàm cần mở rộng trong danh sách hàm.

Phương án thực hiện sáng chế nêu trên có thể phát hiện thành công trang web chứa mã VBScript độc hại trên nền sử dụng hệ điều hành Windows, và do đó chặn mã VBScript độc hại này và nhắc người dùng nếu phát hiện thấy mã VBScript độc hại. Vì vậy, quyền của người dùng được bảo đảm và người dùng có thể duyệt các trang web với độ an toàn được nâng cao. Ngoài ra, phương án thực hiện sáng chế nêu trên ngăn chặn các lỗi, có thể xuất hiện trong khi chuyển đổi từ mã VBScript sang mã JavaScript, để phát hiện mã VBScript độc hại một cách có hiệu quả.

Như được thể hiện trên Fig.5, phương án ưu tiên để thực hiện sáng chế mô tả thiết bị phát hiện mã độc hại trên các trang web, thiết bị này bao gồm môđun thu nhận danh sách hàm 401, môđun phân tích cú pháp và tách 402 và môđun kiểm tra 403, trong đó:

môđun thu nhận danh sách hàm 401 được tạo cấu hình để thu nhận danh sách hàm bằng cách thi hành mã xác định, ví dụ mã VBSscript, và mã đối tượng định trước;

môđun phân tích cú pháp và tách 402 được tạo cấu hình để phân tích cú pháp mã VBScript và thu nhận các giá trị biến dựa vào kết quả phân tích cú pháp và danh sách hàm; và

môđun kiểm tra 403 được tạo cấu hình để kiểm tra các giá trị biến.

Trong sáng chế, mã VBScript được sử dụng để làm ví dụ trong phương án thực hiện sáng chế. Theo quan điểm thực tế là việc đưa đối tượng trình duyệt (Browser) và đối tượng môđun đối tượng tài liệu (DOM), thường được sử dụng trong mã VBScript trên các trang web, trực tiếp vào trong trình thông dịch MSScript sẽ gây ra lỗi đối tượng bất định, và do đó dẫn đến lỗi thi hành. Vì vậy, bằng cách định nghĩa trước các đối tượng Browser và DOM thường dùng này như đối tượng Navigator, đối tượng Document và đối tượng Object theo phương án thực hiện sáng chế, vấn đề do lỗi đối tượng bất định gây ra có thể được giải quyết.

Sau đó, môđun thu nhận danh sách hàm 401 thu nhận danh sách hàm bằng cách gọi giao diện lập trình script để thi hành mã VBScript và mã đối tượng định trước. Cụ thể là, mã VBScript và mã đối tượng định trước được thi hành bằng phương pháp thi hành mã, ví dụ phương pháp ExecuteStatement, được cung cấp thông qua giao diện lập trình script IScriptControl.

Sau khi mã được thi hành thành công, môđun thu nhận danh sách hàm 401 thu nhận danh sách tên thủ tục (hàm) trong mã VBScript bằng cách gọi phương pháp thu nhận danh sách tên thủ tục, ví dụ phương pháp GetProcedures, được cung cấp thông qua giao diện lập trình script IscriptControl, thu nhận con trỏ giao diện IDispatch bằng cách gọi phương pháp thu nhận danh sách biến, ví dụ phương pháp GetCodeObject, được cung cấp thông qua giao diện lập trình script IscriptControl, và sau đó thu nhận danh sách biến toàn cục trong mã VBScript bằng cách sử dụng cơ chế ánh xạ COM; trong đó, danh sách tên thủ tục nêu trên và danh sách biến toàn cục được gọi là danh sách hàm thu được.

Sau khi danh sách hàm được thu nhận, môđun phân tích cú pháp và tách 402 thu nhận các thông tin thủ tục hàm chi tiết như danh sách thông số của hàm và tập hợp câu lệnh bằng cách phân tích cú pháp mã VBScript ban đầu, và thu nhận mã VBScript mới bằng cách thực hiện quy trình tinh chỉnh thủ tục hàm trên mã VBScript ban đầu để loại bỏ hoàn toàn tất cả các thủ tục hàm ra khỏi mã VBScript ban đầu; trong đó, mục đích của

việc thực hiện quy trình tinh chỉnh thủ tục hàm trên mã VBScript ban đầu là để thi hành mã VBScript mở rộng trong trình thông dịch MSScript và nhờ đó tách ra các giá trị biến có mặt ở trong đó.

Trong khi đó, môđun phân tích cú pháp và tách 402 thu nhận danh sách biến cục bộ bằng cách tuần tự gọi phương pháp ExecuteStatement và phương pháp GetCodeObject được cung cấp thông qua giao diện lập trình script IScriptControl đối với mỗi hàm dựa vào các thông tin thủ tục hàm chi tiết trong mã VBScript thu được. Vì mã thi hành độc hại thường xuất hiện trong các biến cục bộ, cho nên sự xuất hiện của mã thi hành độc hại có thể được xác định bằng cách trước tiên thu nhận các biến cục bộ và sau đó thi hành các biến cục bộ trong bộ phận tách đặc điểm để kiểm tra.

Thông qua quy trình nêu trên, tất cả các thông tin cơ bản cần thiết để mở rộng mã VBScript được thu nhận; và sau đó mã VBScript được mở rộng dựa vào danh sách hàm và thông tin thủ tục hàm.

Ngoài ra, để nâng cao hiệu quả của quy trình mở rộng mã VBScript, phương án thực hiện sáng chế đề xuất sử dụng bảng quan hệ phụ thuộc giữa các hàm, qua đó một hàm có thể được mở rộng theo cấu trúc phân cấp và nhờ đó nâng cao hiệu quả của quy trình mở rộng.

Cụ thể là, bảng quan hệ phụ thuộc hai chiều biểu thị các quan hệ phụ thuộc giữa các hàm được tạo ra bằng cách phân tích quan hệ gọi đối với mỗi hàm. Trong sáng chế, quan hệ phụ thuộc được biểu diễn dưới dạng sự phụ thuộc ngược, ví dụ, như sau:

Đối với các hàm A, B, C, D, E, F và G, có quan hệ gọi hàm: các hàm B, D và G có thể được gọi bằng hàm A; các hàm C, E và G có thể được gọi bằng hàm B; và các hàm F và G có thể được gọi bằng hàm E.

Do đó, bảng quan hệ phụ thuộc hai chiều có thể được thiết lập như sau:

A→NIL;

B→A;

C→B;

D→A;

E→B;

F→E;

G→A, B, E.

Đối với mỗi hàm, quy trình mở rộng chủ yếu được xây dựng dựa vào bảng quan hệ phụ thuộc giữa các hàm; do đó, bộ phận chọn mở rộng hàm được đưa vào sử dụng và được thiết kế để trả về hàm cần mở rộng kế tiếp. Cụ thể là, bộ phận chọn mở rộng hàm được tạo cấu hình để quét ngang qua danh sách hàm hiện thời để thu nhận hàm thứ nhất không được mở rộng và có quan hệ phụ thuộc giữa các hàm là NIL, bộ phận chọn mở rộng hàm này trả về hàm cần mở rộng kế tiếp, và tuần tự mở rộng mỗi hàm cần mở rộng trong danh sách hàm.

Đối với ví dụ nêu trên (các hàm A, B, C, D, E, F và G), quy trình mở rộng được mô tả như sau:

1. Mở rộng hàm A nếu hàm A không phụ thuộc vào bất kỳ hàm nào khác;

2. Mở rộng, sau khi hàm A được mở rộng và do đó các quan hệ phụ thuộc giữa các hàm B và D là NIL, mỗi hàm trong số hàm B hoặc hàm D kế tiếp sau hàm A (trong ví dụ này chọn trước tiên quét hàm B), sau đó quan hệ phụ thuộc giữa các hàm C, D và E là NIL;

3. Mở rộng, sau khi hàm B được mở rộng và do đó các quan hệ phụ thuộc giữa các hàm C, D và E là NIL, hàm C kế tiếp sau hàm B;

4. Tuần tự mở rộng các hàm D và E kế tiếp sau hàm C;

5. Tuần tự mở rộng, sau khi hàm E được mở rộng và do đó các quan hệ phụ thuộc giữa các hàm F và G là NIL, các hàm F và G.

Đối với mỗi hàm, quy trình mở rộng chủ yếu được thực hiện bằng cách tìm hàm cần gọi, xây dựng một tập hợp câu lệnh mới, và thực hiện việc thay thế. Việc xây dựng một tập hợp câu lệnh mới được thực hiện bằng cách đặt lại tên cho các thông số của hàm và các biến cục bộ của hàm dưới dạng là tên-hàm_tên-biến (tên-thông-số)_ID-gọi. Ngoài ra, các thông số ở phần trước của tập hợp câu lệnh có dạng biến cục bộ, và giá trị ước tính tương ứng với các thông số được đưa vào trong lúc gọi được kết hợp vào trong các biến. Giá trị ID gọi biểu thị số lần gọi của hàm phát hiện được hiện thời, giá trị này được thực hiện để ngăn chặn tình trạng xung đột biến xảy ra do nhiều lần gọi và mở rộng hàm.

Sau khi quy trình mở rộng tất cả các hàm đã hoàn thành, một mã VBScript mới được thu nhận.

Mã VBScript mới thu được sau khi hoàn thành quy trình mở rộng hàm được nhập vào và thi hành trong trình thông dịch MSScript. Một danh sách gồm tất cả các giá trị biến được thu nhận theo cơ chế ánh xạ giao diện COM, và sau đó các giá trị biến thu được được nhập vào bộ phận tách đặc điểm bằng môđun kiểm tra 403 để tách ra và kiểm tra nhằm hoàn thành việc phát hiện mã VBScript độc hại.

Như được thể hiện trên Fig.6, trong quy trình thực hiện cụ thể sử dụng mã VBScript để làm ví dụ, môđun thu nhận danh sách hàm 401 bao gồm: bộ phận thi hành 4011, bộ phận thu nhận danh sách tên thủ tục 4012 và bộ phận thu nhận danh sách biến toàn cục 4013, trong đó:

bộ phận thi hành 4011 được tạo cấu hình để thi hành mã VBScript và mã đối tượng định trước bằng cách gọi phương pháp ExecuteStatement được cung cấp thông qua giao diện lập trình script;

bộ phận thu nhận danh sách tên thủ tục 4012 được tạo cấu hình để thu nhận danh sách tên thủ tục trong mã VBScript bằng cách gọi phương pháp GetProcedures được cung cấp thông qua giao diện lập trình script;

bộ phận thu nhận danh sách biến toàn cục 4013 được tạo cấu hình để thu nhận con trỏ giao diện IDispatch bằng cách gọi phương pháp GetCodeObject được cung cấp thông qua giao diện lập trình script và thu nhận danh sách biến toàn cục trong mã VBScript bằng cách sử dụng cơ chế ánh xạ COM.

Như được thể hiện trên Fig.7, môđun phân tích cú pháp và tách 402 bao gồm:

bộ phận phân tích cú pháp và nhận biết 4021 được tạo cấu hình để phân tích cú pháp mã xác định và nhận biết thông tin thủ tục hàm trong mã xác định;

bộ phận mở rộng 4022 được tạo cấu hình để mở rộng mã xác định dựa vào danh sách hàm và thông tin thủ tục hàm;

bộ phận tách ra giá trị biến 4023 được tạo cấu hình để tách ra các giá trị biến bằng cách thi hành mã xác định mở rộng.

Như được thể hiện trên Fig.8, bộ phận mở rộng 4022 bao gồm: bộ phận con thu nhận quan hệ gọi 40221, bộ phận con tạo ra 40222 và bộ phận con mở rộng 40223, trong đó:

bộ phận con thu nhận quan hệ gọi 40221 được tạo cấu hình để thu nhận quan hệ gọi đối với mỗi hàm dựa vào thông tin thủ tục hàm;

bộ phận con tạo ra 40222 được tạo cấu hình để tạo ra bảng quan hệ phụ thuộc hai chiều dựa vào quan hệ gọi đối với mỗi hàm;

bộ phận con mở rộng 40223 được tạo cấu hình để mở rộng mã VBScript dựa vào danh sách hàm và bảng quan hệ phụ thuộc hai chiều.

Cụ thể là, bộ phận con mở rộng 40223 quét ngang qua danh sách hàm để thu nhận hàm thứ nhất không được mở rộng và có quan hệ phụ thuộc giữa các hàm là NIL, bộ phận con mở rộng này hàm trả về hàm cần mở rộng kế tiếp, và tuần tự mở rộng mỗi hàm cần mở rộng trong danh sách hàm.

Tóm lại, sáng chế mô tả phương pháp và thiết bị phát hiện mã độc hại trên các trang web. Nhờ vào việc tuần tự thu nhận danh sách hàm bằng cách thi hành mã VBScript và mã đối tượng định trước thông qua giao diện script, thu nhận thông tin thủ tục hàm trong mã VBScript bằng cách phân tích cú pháp mã VBScript, mở rộng mã VBScript dựa vào danh sách hàm và thông tin thủ tục hàm và tách ra các giá trị biến bằng cách thi hành mã VBScript mở rộng trong trình thông dịch MSScript và kiểm tra, mã script độc hại trên các trang web có thể được phát hiện trước và sau đó, hệ thống liên quan có thể chặn mã VBScript độc hại này và nhắc người dùng nếu phát hiện thấy mã VBScript độc hại; do đó, quyền của người dùng được bảo đảm và người dùng có thể duyệt các trang web với độ an toàn được nâng cao.

Phần mô tả sáng chế trên đây dựa vào các phương án ưu tiên để thực hiện sáng chế chỉ được sử dụng để làm ví dụ chứ không được sử dụng để giới hạn phạm vi của sáng chế. Trái lại, phạm vi của sáng chế được hiểu là bao gồm nhiều phương án cải biến khác nhau và các cấu trúc tương tự nằm trong phạm vi được xác định dựa vào các điểm yêu cầu bảo hộ kèm theo, theo cách hiểu rộng nhất là bao gồm tất cả các phương án cải biến và các cấu trúc tương tự như vậy.

YÊU CẦU BẢO HỘ

1. Phương pháp phát hiện mã độc hại trên các trang web bao gồm các bước:

thu nhận danh sách hàm bằng cách thi hành mã xác định và mã đối tượng định trước, trong đó mã xác định là mã script và bước thu nhận danh sách hàm bằng cách thi hành mã xác định và mã đối tượng định trước bao gồm các bước: thi hành mã script và mã đối tượng định trước bằng cách gọi phương pháp thi hành mã được cung cấp thông qua giao diện lập trình script, thu nhận danh sách tên thủ tục trong mã script bằng cách gọi phương pháp thu nhận danh sách tên thủ tục được cung cấp thông qua giao diện lập trình script, và thu nhận con trỏ giao diện bằng cách gọi phương pháp thu nhận danh sách biến được cung cấp thông qua giao diện lập trình script và thu nhận danh sách biến toàn cục trong mã script bằng cách sử dụng cơ chế ánh xạ;

phân tích cú pháp mã xác định và thu nhận các giá trị biến dựa vào kết quả phân tích cú pháp và danh sách hàm, trong đó bước phân tích cú pháp mã xác định và thu nhận các giá trị biến dựa vào kết quả phân tích cú pháp và danh sách hàm bao gồm các bước: thu nhận thông tin thủ tục hàm trong mã xác định bằng cách phân tích cú pháp mã xác định, mở rộng mã xác định dựa vào danh sách hàm và thông tin thủ tục hàm, và tách ra các giá trị biến bằng cách thi hành mã xác định mở rộng; trong đó bước mở rộng mã xác định dựa vào danh sách hàm và thông tin thủ tục hàm bao gồm các bước: thu nhận quan hệ gọi đối với mỗi hàm dựa vào thông tin thủ tục hàm, tạo ra bảng quan hệ phụ thuộc hai chiều dựa vào quan hệ gọi đối với mỗi hàm, quét ngang qua, bằng bộ phận chọn mở rộng hàm, danh sách hàm để thu nhận hàm thứ nhất không được mở rộng và có quan hệ phụ thuộc giữa các hàm là NIL, bộ phận chọn mở rộng hàm này trả về hàm cần mở rộng kế tiếp, và tuần tự mở rộng mỗi hàm trong danh sách hàm; và

xác định xem mã độc hại có xuất hiện trên các trang web hay không dựa vào các giá trị biến.

2. Phương pháp theo điểm 1, trong đó phương pháp này còn bao gồm bước:

kiểm tra các giá trị biến.

3. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 2, trong đó bước thu nhận

thông tin thủ tục hàm trong mã xác định bằng cách phân tích cú pháp mã xác định bao gồm bước:

thu nhận danh sách biến cục bộ bằng cách tuần tự gọi phương pháp thi hành mã và phương pháp thu nhận danh sách biến đối với mỗi hàm.

4. Thiết bị phát hiện mã độc hại trên các trang web bao gồm:

môđun thu nhận danh sách hàm được tạo cấu hình để thu nhận danh sách hàm bằng cách thi hành mã xác định và mã đối tượng định trước, trong đó mã xác định là mã script, và môđun thu nhận danh sách hàm bao gồm:

bộ phận thi hành được tạo cấu hình để thi hành mã script và mã đối tượng định trước bằng cách gọi phương pháp thi hành mã được cung cấp thông qua giao diện lập trình script;

bộ phận thu nhận danh sách tên thủ tục được tạo cấu hình để thu nhận danh sách tên thủ tục trong mã script bằng cách gọi phương pháp thu nhận danh sách tên thủ tục được cung cấp thông qua giao diện lập trình script; và

bộ phận thu nhận danh sách biến toàn cục được tạo cấu hình để thu nhận con trỏ giao diện bằng cách gọi phương pháp thu nhận danh sách biến được cung cấp thông qua giao diện lập trình script và thu nhận danh sách biến toàn cục trong mã script bằng cách sử dụng cơ chế ánh xạ, quét ngang qua, bằng bộ phận chọn mở rộng hàm, danh sách hàm để thu nhận hàm thứ nhất không được mở rộng và có quan hệ phụ thuộc giữa các hàm là NIL, bộ phận chọn mở rộng hàm này trả về hàm cần mở rộng kế tiếp, và tuần tự mở rộng mỗi hàm trong danh sách hàm; và

môđun phân tích cú pháp và tách được tạo cấu hình để phân tích cú pháp mã xác định và thu nhận các giá trị biến dựa vào kết quả phân tích cú pháp và danh sách hàm, trong đó môđun phân tích cú pháp và tách bao gồm:

bộ phận phân tích cú pháp và thu nhận được tạo cấu hình để phân tích cú pháp mã xác định và thu nhận thông tin thủ tục hàm trong mã xác định;

bộ phận mở rộng được tạo cấu hình để mở rộng mã xác định dựa vào danh sách hàm và thông tin thủ tục hàm; và

bộ phận tách ra giá trị biến được tạo cấu hình để tách ra các giá trị biến bằng cách thi hành mã xác định mở rộng,

trong đó bộ phận mở rộng được tạo cấu hình để mở rộng mã xác định dựa vào danh sách hàm và thông tin thủ tục hàm và bao gồm:

bộ phận con thu nhận quan hệ gọi được tạo cấu hình để thu nhận quan hệ gọi đối với mỗi hàm dựa vào thông tin thủ tục hàm;

bộ phận con tạo ra được tạo cấu hình để tạo ra bảng quan hệ phụ thuộc hai chiều dựa vào quan hệ gọi đối với mỗi hàm;

bộ phận con mở rộng được tạo cấu hình để mở rộng mã xác định dựa vào danh sách hàm và bảng quan hệ phụ thuộc hai chiều,

trong đó bộ phận con mở rộng còn được tạo cấu hình để quét ngang qua, bằng bộ phận chọn mở rộng hàm, danh sách hàm để thu nhận hàm thứ nhất không được mở rộng và có quan hệ phụ thuộc giữa các hàm là NIL, bộ phận chọn mở rộng hàm này trả về hàm cần mở rộng kế tiếp, và tuần tự mở rộng mỗi hàm trong danh sách hàm,

trong đó việc mã độc hại có xuất hiện trên các trang web hay không được xác định dựa vào các giá trị biến.

5. Thiết bị theo điểm 4, trong đó thiết bị này còn bao gồm:

môđun kiểm tra được tạo cấu hình để kiểm tra các giá trị biến.

6. Thiết bị theo điểm bất kỳ trong số các điểm từ 4 đến 5, trong đó môđun phân tích cú pháp và tách còn được tạo cấu hình để thu nhận danh sách biến cục bộ bằng cách tuần tự gọi phương pháp thi hành mã và phương pháp thu nhận danh sách biến đối với mỗi hàm.


Bắt đầu
Thu nhận danh sách hàm bằng cách
thi hành mã xác định và mã đối tượng định trước
S101
Phân tích cú pháp mã xác định và thu nhận các giá trị biến
dựa vào kết quả phân tích cú pháp và danh sách hàm
S102
Kiểm tra các giá trị biến
Kết thúc
S103

FIG. 1

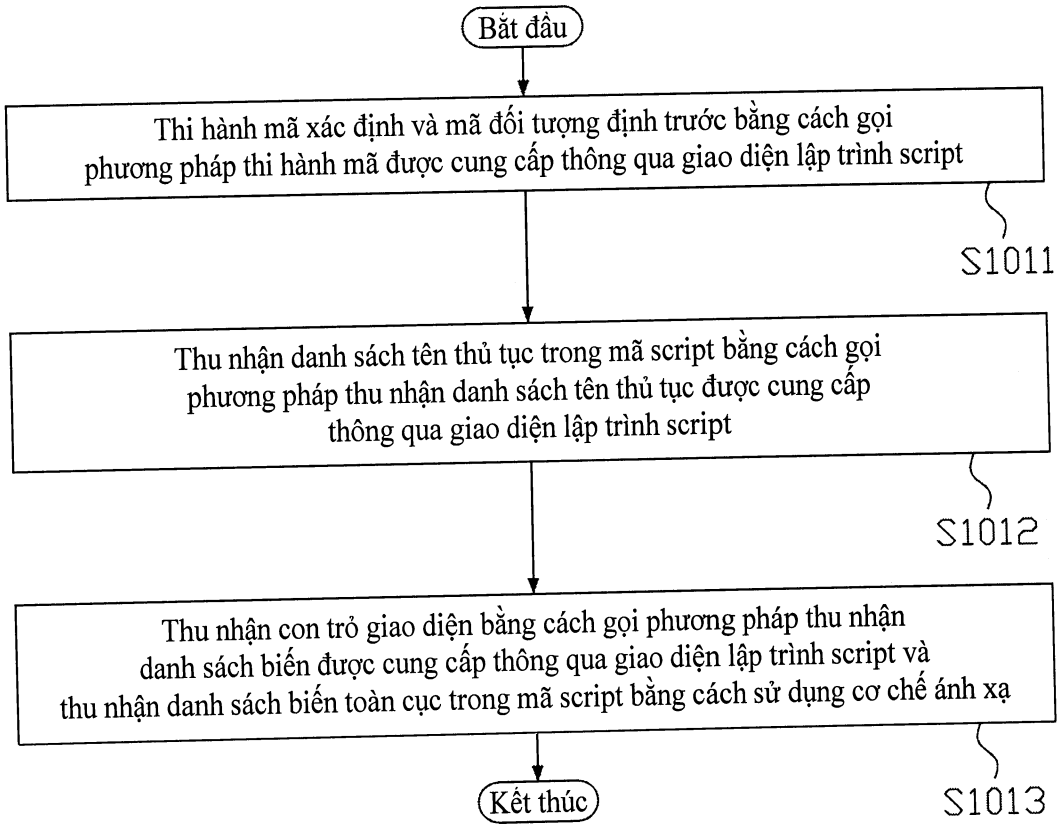
Bắt đầu
Thi hành mã xác định và mã đối tượng định trước bằng cách gọi
phương pháp thi hành mã được cung cấp thông qua giao diện lập trình script
S1011
Thu nhận danh sách tên thủ tục trong mã script bằng cách gọi
phương pháp thu nhận danh sách tên thủ tục được cung cấp
thông qua giao diện lập trình script
S1012
Thu nhận con trỏ giao diện bằng cách gọi phương pháp thu nhận
danh sách biến được cung cấp thông qua giao diện lập trình script và
thu nhận danh sách biến toàn cục trong mã script bằng cách sử dụng cơ chế ánh xạ
Kết thúc
S1013

FIG. 2

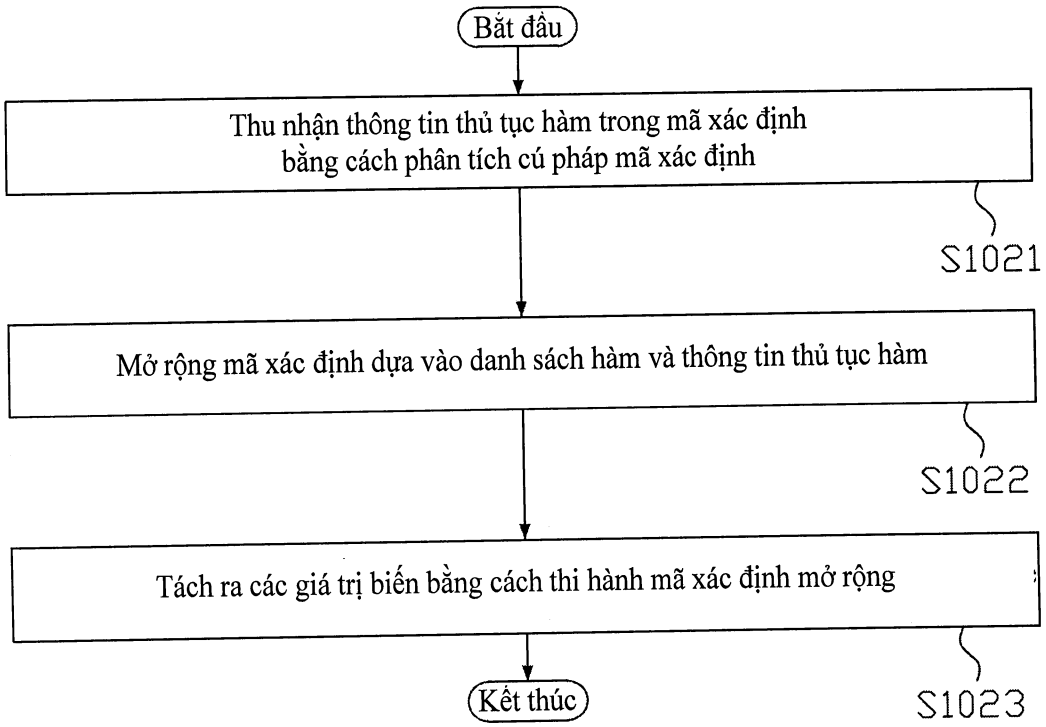
Bắt đầu
Thu nhận thông tin thủ tục hàm trong mã xác định bằng cách phân tích cú pháp mã xác định
S1021
Mở rộng mã xác định dựa vào danh sách hàm và thông tin thủ tục hàm
S1022
Tách ra các giá trị biến bằng cách thi hành mã xác định mở rộng
Kết thúc
S1023

FIG. 3

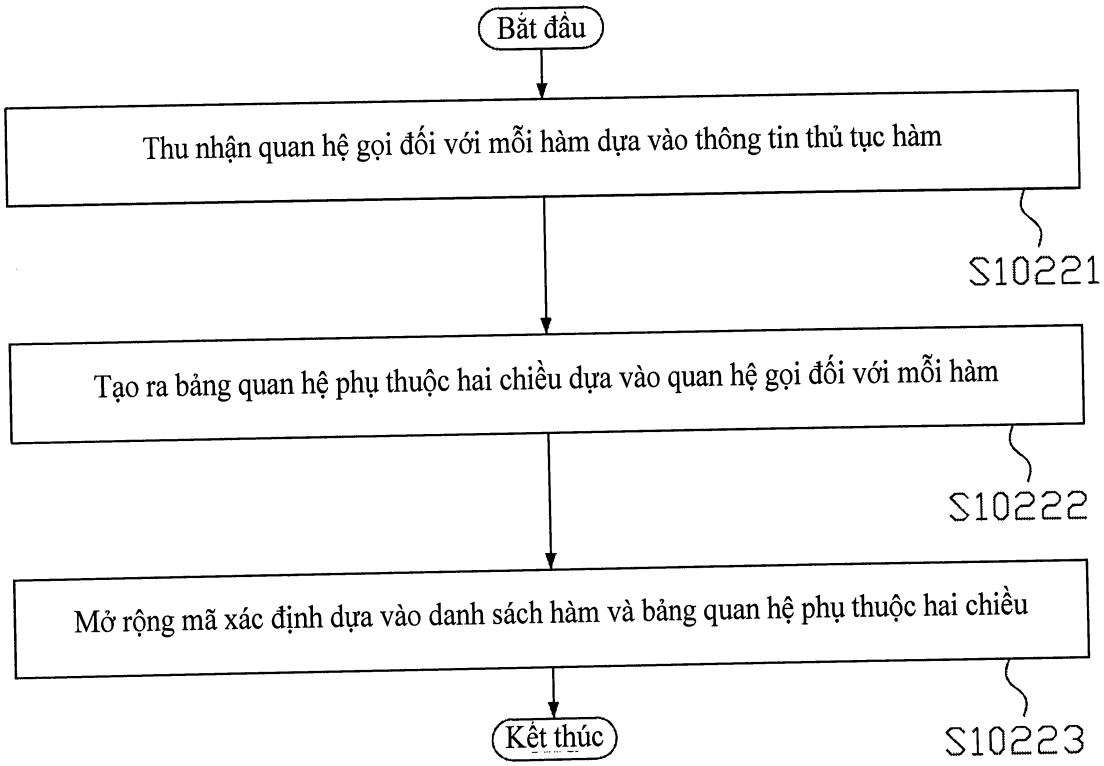
Bắt đầu
Thu nhận quan hệ gọi đối với mỗi hàm dựa vào thông tin thủ tục hàm
S10221
Tạo ra bảng quan hệ phụ thuộc hai chiều dựa vào quan hệ gọi đối với mỗi hàm
S10222
Mở rộng mã xác định dựa vào danh sách hàm và bảng quan hệ phụ thuộc hai chiều
Kết thúc
S10223

FIG. 4

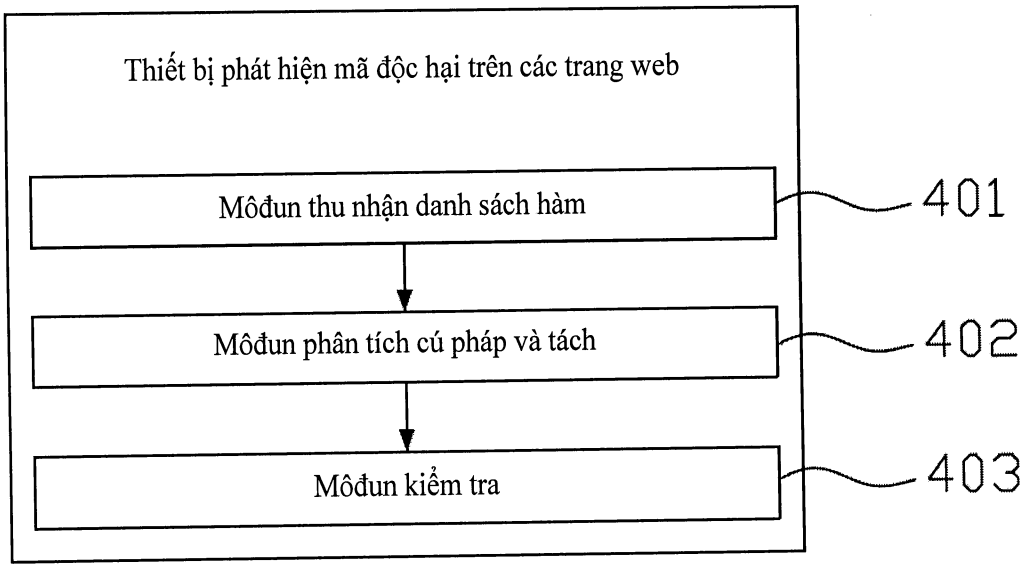
Thiết bị phát hiện mã độc hại trên các trang web
Môđun thu nhận danh sách hàm
401
Môđun phân tích cú pháp và tách
402
Môđun kiểm tra
403

FIG. 5

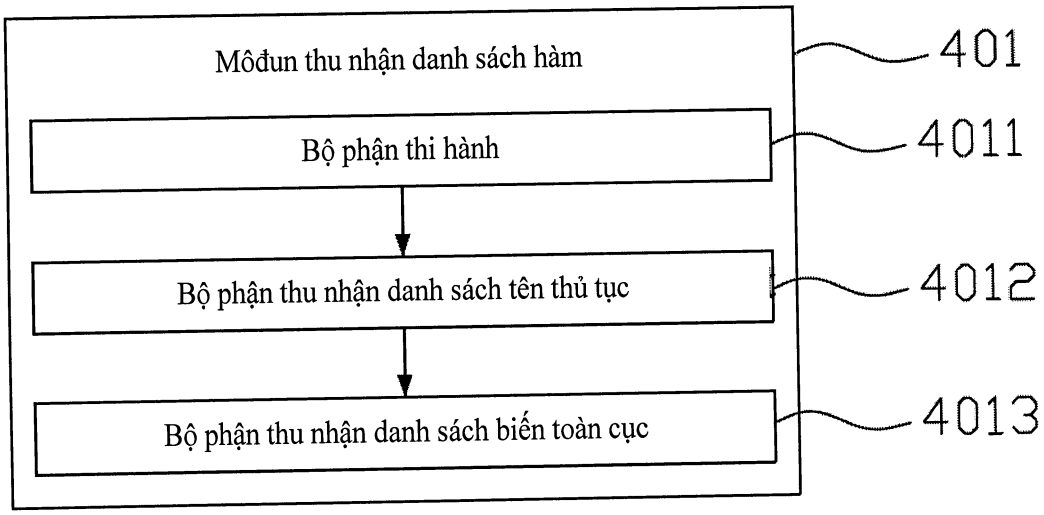
Môđun thu nhận danh sách hàm
401
Bộ phận thi hành
4011
Bộ phận thu nhận danh sách tên thủ tục
4012
Bộ phận thu nhận danh sách biến toàn cục
4013

FIG. 6

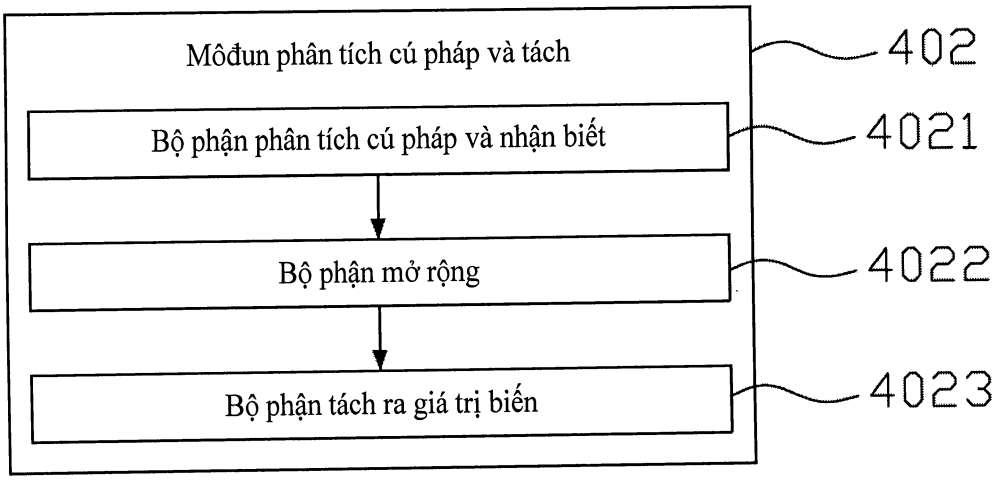
Môđun phân tích cú pháp và tách
402
Bộ phận phân tích cú pháp và nhận biết
4021
Bộ phận mở rộng
4022
Bộ phận tách ra giá trị biến
4023

FIG. 7

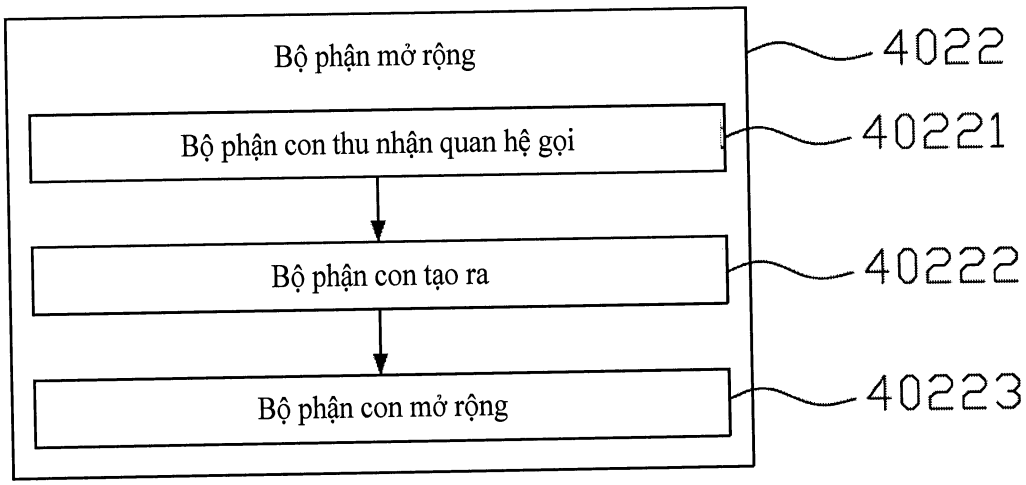
Bộ phận mở rộng
4022
Bộ phận con thu nhận quan hệ gọi
40221
Bộ phận con tạo ra
40222
Bộ phận con mở rộng
40223

FIG. 8