(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0070023
(43) 공개일자 2024년05월21일

(51) 국제특허분류(Int. Cl.)
*G06F 21/14* (2013.01)
(52) CPC특허분류
*G06F 21/14* (2013.01)
*G06F 2221/2125* (2013.01)
(21) 출원번호 10-2022-0151321
(22) 출원일자 2022년11월14일
심사청구일자 2022년11월14일

(71) 출원인
**충남대학교산학협력단**
대전광역시 유성구 대학로 99 (궁동, 충남대학교)
(72) 발명자
**조은선**
대전광역시 유성구 어은로 57 한빛아파트, 110동 1504호
**목성균**
대전광역시 서구 가장로 131-2, 세종빌 403호
(74) 대리인
**이은철, 김재문**

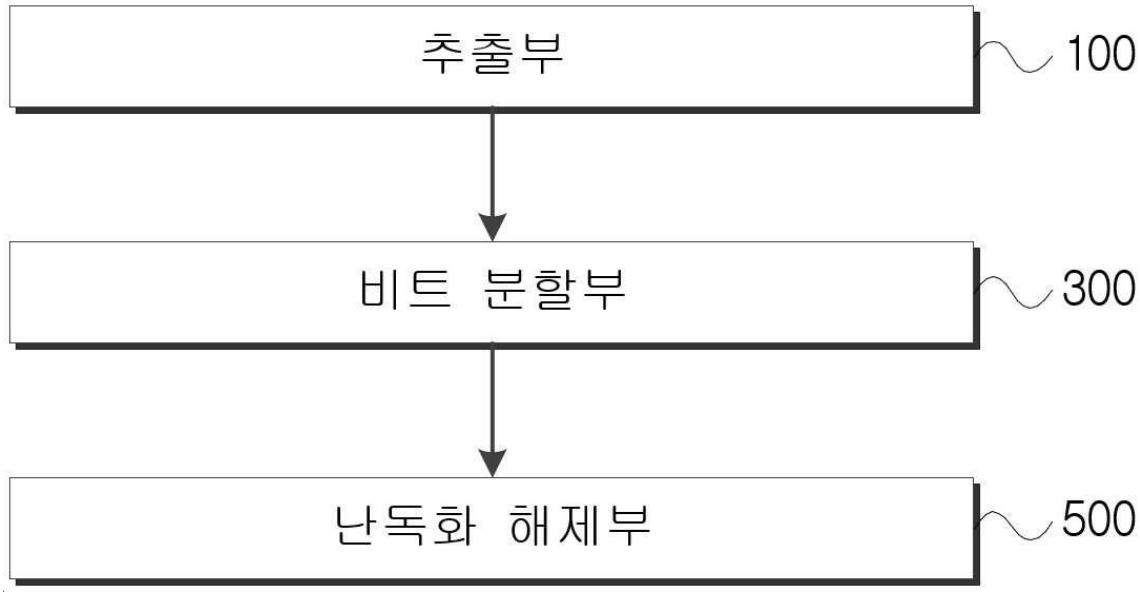
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **비선형 MBA 난독화 데이터를 단순화하는 전자 장치 및 방법**

## (57) 요 약

본 발명은 비선형 MBA 난독화 데이터를 단순화하는 전자 장치 및 방법에 관한 것으로, 컴퓨터에서 판독 가능한 적어도 하나의 명령어가 저장되는 메모리 및 메모리에 저장된 명령어를 수행하도록 구현되는 프로세서를 포함하고, 프로세서의 추출부가 입력된 난독화 데이터로부터 적어도 하나의 비트값을 추출하고, 비트 분할부가 추출된 비트값이 변경되는 지점을 기준으로 하여 난독화 데이터를 분할하고, 동일한 적어도 하나의 비트값을 그룹화하고, 난독화 해제부가 그룹화된 비트값을 단순화 알고리즘으로 연산하여 그룹별로 단순화를 수행하는 구성으로 악성코드로 인한 난독화의 해제 효율성을 향상시킬 수 있다.

*대 표 도* - 도1



이 발명을 지원한 국가연구개발사업

| | |
|---|---|
| 과제고유번호 | 1711160654 |
| 과제번호 | 2022-0-01200-001 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 정보통신기획평가원 |
| 연구사업명 | 정보통신방송혁신인재양성 |
| 연구과제명 | 융합보안대학원(충남대학교) |
| 기 여 율 | 1/1 |
| 과제수행기관명 | 충남대학교 산학협력단 |
| 연구기간 | 2022.01.01 ~ 2022.12.31 |

# 명 세 서

## 청구범위

### 청구항 1

입력된 난독화 데이터로부터 적어도 하나의 비트값을 추출하는 추출부;

추출된 상기 비트값이 변경되는 지점을 기준으로 하여 상기 난독화 데이터를 분할하고, 동일한 적어도 하나의 비트값을 그룹화하는 비트 분할부; 및

상기 그룹화된 비트값을 단순화 알고리즘으로 연산하여 그룹별로 단순화를 수행하는 난독화 해제부를 포함하는 비선형 MBA 난독화 데이터를 단순화하는 전자 장치.

### 청구항 2

제 1 항에 있어서,

상기 단순화된 각 그룹의 비트값을 통합하여 상기 난독화 데이터가 단순화된 단순화 데이터를 생성하는 통합부를 더 포함하는 비선형 MBA 난독화 데이터를 단순화하는 전자 장치.

### 청구항 3

제 2 항에 있어서,

상기 통합부는 통합된 비트값과 상기 그룹별 비트값을 비교하여 일치 여부를 판단하는 것을 특징으로 하는 비선형 MBA 난독화 데이터를 단순화하는 전자 장치.

### 청구항 4

제 1 항에 있어서,

상기 추출부는 상기 비트값을 2진법 및 16진법 중 어느 하나로 추출하는 것을 특징으로 하는 비선형 MBA 난독화 데이터를 단순화하는 전자 장치.

### 청구항 5

제 1 항에 있어서,

상기 비트 분할부는 상기 비트값이 연속적으로 구성된 선형 MBA 표현(Linear Mixed Boolean arithmetic expression)을 도출하고,

도출된 상기 선형 MBA 표현을 분할하여 비선형 MBA 표현으로 변환하는 것을 특징으로 하는 비선형 MBA 난독화 데이터를 단순화하는 전자 장치.

### 청구항 6

제 5 항에 있어서,

상기 비트 분할부는 적어도 하나의 상기 선형 MBA 표현 및 적어도 하나의 상기 비선형 MBA 표현의 합인 반선형 MBA 표현(Semi-linear MBA expression)을 생성하는 것을 특징으로 하는 비선형 MBA 난독화 데이터를 단순화하는 전자 장치.

**청구항 7**

입력된 난독화 데이터로부터 적어도 하나의 비트값을 추출하는 추출 단계;

추출된 상기 비트값이 변경되는 지점을 기준으로 하여 상기 난독화 데이터를 분할하고, 동일한 적어도 하나의 비트값을 그룹화하는 비트 분할 단계; 및

상기 그룹화된 비트값을 단순화 알고리즘으로 연산하여 그룹별로 단순화를 수행하는 난독화 해제 단계를 포함하는 비선형 MBA 난독화 데이터를 단순화하는 방법.

**청구항 8**

제 7 항에 있어서,

상기 단순화된 각 그룹의 비트값을 통합하여 상기 난독화 데이터가 단순화된 단순화 데이터를 생성하는 통합 단계를 더 포함하는 비선형 MBA 난독화 데이터를 단순화하는 방법.

**청구항 9**

제 8 항에 있어서,

상기 통합 단계는 통합된 비트값과 상기 그룹별 비트값을 비교하여 일치 여부를 판단하는 것을 특징으로 하는 비선형 MBA 난독화 데이터를 단순화하는 방법.

**청구항 10**

제 7 항에 있어서,

상기 추출 단계는 상기 비트값을 2진법 및 16진법 중 어느 하나로 추출하는 것을 특징으로 하는 비선형 MBA 난독화 데이터를 단순화하는 방법.

## *발명의 설명*

### *기 술 분 야*

[0001] 본 발명은 비선형 MBA 난독화 데이터를 단순화하는 전자 장치 및 방법에 관한 것으로, MBA 표현(expression)을 단순화할 수 있도록 한 기술에 관한 것이다.

### *배 경 기 술*

[0003] MBA 난독화는 컴퓨터가 읽기 어렵게 구현된 난독화 기술로서 소프트웨어 보호에 유용한 방법이지만 난독화 도구와 기술이 보편화됨에 따라 많은 악성코드가 난독화 기법을 적용하고 있다. 악성코드가 MBA 난독화 옵션으로 난독화되면 그 동작을 이해하기 위한 분석가의 엄청난 노력이 필요하다. 반면, 분석에 활용 가능한 역난독화 도구의 부재와 자동화된 분석 기술의 한계로 난독화 코드에 대한 분석에 많은 시간과 노력이 요구된다. 따라서, 이 문제를 다루기 위한 다양한 단순화 노력이 있어 왔다.

[0004] 특히, Mixed Boolean Arithmetic Expression(MBA 표현)으로 만들어 난독화의 분석은 수학의 방정식을 풀어주는 해석기로 종래의 Satisfiability modulo theories(SMT) solver에서도 단순화하기에 어려움이 있었다.

[0005] MBA 난독화의 부분집합인 선형 MBA 표현 단순화 기법은 로지컬(Logical) 해법이 존재하나, 선형 MBA 표현은 전체 MBA 표현에 비해 범위가 매우 좁고, 난독화 악성코드에서 $x \times y$, $x\&(x+y)$, $x\&0x1$과 같은 MBA 표현은 선

형 표현이 아니기 때문에 난독화 악성코드에는 적합하지 아니한 문제가 있다. 특히, 선형 MBA 표현에서는 0 또는 -1로 나타나지만, 악성코드가 난독화된 코드에서 자주 나타나는 MBA 표현에는 여러가지 상수가 나타날 수 있다. 제어 흐름 난독화를 위해 조건 분기를 위한 jmp 문에서 조건 검사에 활용되는 flag 레지스터를 적절히 마스킹해야하나, 0xFF0000000 등 특정 비트들이 1이고 나머지는 0인 상수들의 등장은 필연적이다. 종래의 난독화된 MBA 표현을 복호화하는 기술은 난이도가 매우 높아서 선형 MBA 표현 난독화에 한정된 경우가 많다.

[0006] 또한, 최근 합성 기반 MBA 표현 단순화가 개발되었다. 이러한 접근 방식은 프로그램 합성을 통해 원래 프로그램의 간단한 표현을 검색한다. 특정 형태의 표현 외에도 복잡한 형태의 MBA 표현에 적용할 수 있으나 솔루션이 불완전하고 난독화를 단순화가 상대적으로 느리다.

[0007] 따라서, 이러한 문제를 해결하기 위해 상수로 0 또는 -1만 가질 수 있는 선형 MBA 표현뿐 아니라, 이외의 상수를 가질 수 있는 반선형(Semi-linear) MEA 표현으로 악성코드 및 난독화된 MBA 표현을 단순화할 수 있는 방법을 제안하고자 한다.

## 선행기술문헌

### 특허문헌

[0009] (특허문헌 0001) 대한민국 등록특허 제10-2438784호(2022.09.02.)

## 발명의 내용

### 해결하려는 과제

[0010] 본 발명은, 악성코드에 빈번히 등장하는 난독화를 해제할 수 있는 비선형 MBA 난독화 데이터를 단순화하는 전자 장치 및 방법을 제공할 수 있다.

### 과제의 해결 수단

[0012] 본 발명의 일 측면에 따른 비선형 MBA 난독화 데이터를 단순화하는 전자 장치는 컴퓨터에서 판독 가능한 적어도 하나의 명령어가 저장되는 메모리; 및 상기 메모리에 저장된 상기 명령어를 수행하도록 구현되는 프로세서;를 포함하고, 상기 프로세서는, 입력된 난독화 데이터로부터 적어도 하나의 비트값을 추출하는 추출부; 추출된 상기 비트값이 변경되는 지점을 기준으로 하여 상기 난독화 데이터를 분할하고, 동일한 적어도 하나의 비트값을 그룹화하는 비트 분할부; 및 상기 그룹화된 비트값을 단순화 알고리즘으로 연산하여 그룹별로 단순화를 수행하는 난독화 해제부를 포함한다.

[0013] 바람직하게는, 상기 프로세서는, 상기 단순화된 각 그룹의 비트값을 통합하여 상기 난독화 데이터가 단순화된 단순화 데이터를 생성하는 통합부를 더 포함할 수 있다.

[0014] 바람직하게는, 상기 통합부는 통합된 비트값과 상기 그룹별 비트값을 비교하여 일치 여부를 판단할 수 있다.

[0015] 바람직하게는, 상기 추출부는 상기 비트값을 2진법 및 16진법 중 어느 하나로 추출할 수 있다.

[0016] 바람직하게는, 상기 비트 분할부는 상기 비트값이 연속적으로 구성된 선형 MBA 표현(Linear Mixed Boolean arithmetic expression)을 도출하고, 도출된 상기 선형 MBA 표현을 분할하여 비선형 MBA 표현으로 변환할 수 있다.

[0017] 바람직하게는, 상기 비트 분할부는 적어도 하나의 상기 선형 MBA 표현 및 적어도 하나의 상기 비선형 MBA 표현의 합인 반선형 MBA 표현(Semi-linear MBA expression)을 생성할 수 있다.

[0018] 본 발명의 다른 측면에 따른 비선형 MBA 난독화 데이터를 단순화하는 방법은 프로세서가 입력된 난독화 데이터로부터 적어도 하나의 비트값을 추출하는 추출 단계; 상기 프로세서가 추출된 상기 비트값이 변경되는 지점을 기준으로 하여 상기 난독화 데이터를 분할하고, 동일한 적어도 하나의 비트값을 그룹화하는 비트 분할 단계; 및 상기 프로세서가 상기 그룹화된 비트값을 단순화 알고리즘으로 연산하여 그룹별로 단순화를 수행하는 난독화 해

제 단계를 포함한다.

[0019] 바람직하게는, 상기 프로세서가 상기 단순화된 각 그룹의 비트값을 통합하여 상기 난독화 데이터가 단순화된 단순화 데이터를 생성하는 통합 단계를 더 포함한다.

[0020] 바람직하게는, 상기 통합 단계는 통합된 비트값과 상기 그룹별 비트값을 비교하여 일치 여부를 판단할 수 있다.

[0021] 바람직하게는, 상기 추출 단계는 상기 비트값을 2진법 및 16진법 중 어느 하나로 추출할 수 있다.

## *발명의 효과*

[0023] 본 발명에 따르면, 악성코드로 인한 난독화의 해제 효율성을 향상시킬 수 있다.

## *도면의 간단한 설명*

[0025] 도 1는 일 실시예에 따른 프로세서를 구체화한 구성도이다.

도 2는 일 실시예에 따른 SSLEM 기법을 도시한 도이다.

도 3는 일 실시예에 따른 비선형 MBA 난독화 데이터를 단순화하는 방법을 도시한 순서도이다.

## *발명을 실시하기 위한 구체적인 내용*

[0026] 이하에서는 본 발명에 따른 비선형 MBA 난독화 데이터를 단순화하는 전자 장치 및 방법을 첨부된 도면들을 참조하여 상세하게 설명한다. 이러한 과정에서 도면에 도시된 구성요소의 크기 등은 설명의 명료성과 편의상 과장되게 도시되어 있을 수 있다. 또한, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 운용자의 의도 또는 관례에 따라 달라질 수 있다. 그러므로, 이러한 용어들에 대한 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

[0027] 본 발명의 목적 및 효과는 하기의 설명에 의해서 자연스럽게 이해되거나 보다 분명해질 수 있으며, 하기의 기재만으로 본 발명의 목적 및 효과가 제한되는 것은 아니다. 또한, 본 발명을 설명함에 있어서 본 발명과 관련된 공지 기술에 대한 구체적인 설명이, 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다.

[0028] 도 1는 일 실시예에 따른 프로세서를 구체화한 구성도이다.

[0029] 도 1에서 나타낸 바와 같이, 비선형 MBA 난독화 데이터를 단순화하는 전자 장치는 컴퓨터에서 판독 가능한 적어도 하나의 명령어가 저장되는 메모리 및 상기 메모리에 저장된 명령어를 수행하도록 구현되는 프로세서를 포함한다. 여기서, 프로세서는 추출부(100), 비트 분할부(300), 및 난독화 해제부(500)를 포함하고, 통합부(700)를 더 포함할 수 있다.

[0030] 추출부(100)는 입력된 난독화 데이터(10)로부터 적어도 하나의 비트값(20)을 추출한다. 여기서, 추출부(100)는 상기 비트값(20)을 2진법 및 16진법 중 어느 하나로 추출할 수 있다.

[0031] 비트 분할부(300)는 추출된 상기 비트값(20)이 변경되는 지점을 기준으로 하여 상기 난독화 데이터(10)를 분할하고, 동일한 적어도 하나의 비트값(20)을 그룹화한다. 여기서, 비트 분할부(300)는 상기 비트값(20)이 연속적으로 구성된 선형 MBA 표현(Linear Mixed Boolean arithmetic expression)을 도출하고, 도출된 상기 선형 MBA 표현을 분할하여 비선형 MBA 표현으로 변환할 수 있다. 또한, 비트 분할부(300)는 적어도 하나의 상기 선형 MBA 표현 및 적어도 하나의 상기 비선형 MBA 표현의 합인 반선형 MBA 표현(Semi-linear MBA expression)을 생성할 수 있다. 즉, 분할된 난독화 데이터는 분할된 제1 난독화 데이터(30) 및 분할된 제2 난독화 데이터(40)로 구성될 수 있다. 분할된 제1 난독화 데이터(30) 및 분할된 제2 난독화 데이터(40)는 서로 다른 비트값이지만, 분할된 제1 난독화 데이터(30) 및 분할된 제2 난독화 데이터(40) 각각에 존재하는 비트값들은 동일 비트값으로 구성될 수 있다.

[0032] 난독화 데이터는 MBA 표현일 수 있고, 분할된 난독화 데이터들은 선형 MBA 표현일 수 있다.

[0033] 난독화 해제부(500)는 상기 그룹화된 비트값(20)을 단순화 알고리즘으로 연산하여 그룹별로 단순화를 수행한다.

난독화 해제부(500)는 논리 연산 기반의 단순화 알고리즘을 사용하여 각 그룹의 비트값(20)을 단순화할 수 있다.

[0034] 프로세서는 통합부(700)를 더 포함할 수 있다. 통합부(700)는 상기 단순화된 각 그룹의 비트값(20)을 통합하여 상기 난독화 데이터(10)가 단순화된 단순화 데이터를 생성할 수 있고, 통합된 비트값(20)과 상기 그룹별 비트값(20)을 비교하여 일치 여부를 판단할 수 있다.

[0035] 일 실시예에 따른 비트 분할부(300)는 난독화 데이터(10)에서 반선형 MBA 표현을 생성할 수 있다. 비트 분할부(300)는 비트값(20)이 변경되는 지점을 기준으로 난독화 데이터(10)를 분할할 수 있고, 반선형 MBA 표현은 분할된 난독화 데이터(10)로 인해 선형 MBA 데이터와 비선형 MBA 데이터를 포함할 수 있다. 비트 분할부(300)에서 적어도 하나의 선형 MBA 표현을 추출하되, 각 선형 MBA 표현은 동일 비트값(20)끼리 그룹화될 수 있다. 추출된 적어도 하나의 선형 MBA 표현 및 비선형 MBA 표현을 단순화 알고리즘에 입력으로 단순화를 수행할 수 있다. 단순화된 각 선형 MBA 표현의 합은 원본 데이터와 동일한 값임을 의미할 수 있다.

[0036] 여기서, 비트값(20)은 n 진법으로 추출될 수 있으며, 더욱 상세하게는 2진법 및 16진법으로 추출된 상수 또는 문자일 수 있으며, 추출된 비트값(20)을 순차적으로 스캔을 수행할 수 있고, 스캔된 비트값(20)이 변경되는 지점을 비트변환지점으로 설정할 수 있다.

[0037] 선형 MBA 표현은 산술 연산을 의미하고, 계수 $a_i$와 순수 부울(boolean) 표현 $e_i(x_1, \dots, x_T)$의 곱셈을 의미하며, T는 변수의 수(N)이다. N은 다음 수학식에서 항의 수로 표현될 수 있다.

[0038] [수학식 1]

[0039] $$\sum_{i=0}^{N} a_i e_i(x_1, \dots, x_T)$$

[0041] 선형 표현의 예는 $10 \times (x \,\&\, y)$, $5 \times (x \mid y) + 6 \times (x \,\&\, y)$, $3 \times (x \,\&\, y) - 20 \times (x \mid y) - 1$ 그리고 $y + 376 \times (x \,\&\, -1)$일 수 있다.

[0042] 앞서 언급했듯이, 선형 MBA 표현은 MBA 표현의 가장 전통적인 단순화 방식이다. 그러나, 패턴 매칭 및 룰 재작성에 따라, 확장성이 부족하여 복잡한 비선형 표현에는 적합하지 아니하다.

[0043] 1-비트(bit) 변수를 사용하는 선형 MBA 표현의 단순화된 형태가 n비트 변수를 갖는 선형 MBA 표현의 단순화된 형식을 나타내며, 그 반대의 경우도 마찬가지로 입증되었다. 이는 1-비트 변수를 사용하여 선형 MBA 표현을 단순화하면 단순화 결과가 n비트 변수를 사용하여 대응하는 동일한 모양의 표현의 단순화로 간주될 수도 있음을 의미한다. 이것은 단순화 과정을 훨씬 더 효율적으로 만든다.

[0044] 입력 변수의 조합을 출력에 매핑하는 함수 표현을 고려한다. 복잡한 표현이든 단순한 표현이든 관계없이 고유한 매핑 수는 입력 및 출력 변수의 크기와 수에 따라 제한된다. 예를 들어, 두 개의 1비트 변수 x와 y를 입력으로 사용하는 경우 x와 y의 가능한 조합은 00, 01, 10 및 11이 된다. 주어진 선형 MBA 표현의 단순화는 검색 문제가 될 수 있다. 각 입력 비트 조합에 대해 가능한 모든 출력 비트 조합 중에서 단순화자는 주어진 표현에 해당하는 단순한 표현을 찾아야 한다. 이때, 속성은 검색 공간을 $2^{2^{2^{16}}}$ (즉, $16^{16} = 1.84e + 19$)에서 $2^{2^{2^1}}$(즉, 8)일 수 있다.

[0045] 그러나 선형 MBA 표현 전용 솔루션은 일반 MBA 표현을 처리할 수 없으며 $2^{2^{2^1}}$ 표현만 처리할 수 있다. 따라서, 주어진 복잡한 표현에서 선형 하위 표현을 찾기 위해 휴리스틱(heuristics)을 적용할 수 있다. 반면, $2^{2^{2^{16}}} - 2^{2^{2^1}}$ MBA 표현을 처리할 수 없으며 그 중 일부는 난독화된 악성코드로 나타날 수 있다.

[0046] 도 2는 일 실시예에 따른 SSLEM 기법을 도시한 도이다.

[0047] 도 2에서 나타낸 바와 같이, 일 실시예에 따른 SSLEM은 새로운 MBA 표현 단순화 솔루션으로, SSLEM은 두 종류의 기존 도구(논리 기반 단순화 및 프로그램 합성 기반 단순화)를 모두 사용할 수 있다. 전술한 바와 같이, 논리

기반 단순화 알고리즘은 빠르고 효율적이지만 제한된 범주의 표현(선형 MBA 표현)만 단순화한다. 이러한 문제점을 극복하기 위해, SSLEM은 일반 MBA 표현을 선형 MBA 표현으로 전처리하고 그 결과를 논리 기반 단순화기에 입력할 수 있다. 그러나 고유한 의미 체계의 수를 비교한 것처럼 선형 MBA 표현은 전체 MBA 표현의 작은 부분 집합에 불과하기 때문에 선형 표현으로의 변환은 대부분의 경우 실패한다.

[0048] 따라서, SSLEM은 반선형(semi-linear) MBA 표현이라고 하는 MBA 표현의 중간 형태를 사용하고, 프로그램 합성 기반 단순화기에 의해 일반 MBA 표현을 반선형 MBA 표현으로 변환한 다음 결과로 생성되는 반선형 MBA 표현을 사용할 수 있다. 표현은 여러 선형 표현으로 구성된 결합 표현으로 변환된다. 각각의 결과에 대한 선형 MBA 표현은 논리 기반 단순화 방법에 의해 단순화될 수 있으므로 최종적으로 논리 기반 단순화로 비선형 표현의 단순화를 얻을 수 있다.

[0049] PLASynth-SL은 III 단계로 작동하는 논리 기반 단순화에 대한 예비 후보를 찾을 수 있다. 비선형 MBA 표현에 해당하는 반선형 MBA 표현을 찾아 의미를 동일하게 유지하는 검색 기반 합성기이다.

[0050] LEM은 반선형 MBA 표현을 선형 MBA 표현의 조합으로 변환할 수 있다. LEM으로 인한 반선형 MBA 표현의 선형 하위 표현은 차례로 논리 기반 단순화기에 공급된다. 이 단계는 MBA Blast와 같은 기존의 강력한 선형 MBA 표현 단순화기를 활용할 수 있다.

[0051] 비트값(20)이 다른 상수(예: 0xFE, 0x1, 0x003F)인 "이종 비트값(20) 상수"로 시작해보면, PLASynth-SL의 선형 표현은 부울에서 "동종 비트값(20) 상수"만 0xFF, -1 및 0x0을 포함하는 하위 표현으로, 각각은 동일한 비트값(20)(모두 1 또는 모두 0)로 구성될 수 있다. Semi-linear MBA 표현은 다음과 같이 정의된 선형 표현의 약간 확장된 형태일 수 있다. 여기서, 반선형 MBA 표현은 부울 하위 표현에 "이종 비트값(20) 상수"가 없는 경우 선형 MBA 표현이 될 수 있는 MBA 표현일 수 있다. 반선형 표현에는 $"30 \times (0xFE \mid y)"$, $"21 \times (x\&0x01)"$이 포함된다.

[0052] PLASynth-SL은 더 나은 성능을 위한 최적화 및 타임아웃 관리 기능을 구비하고 있다. 예비 실험에서 Tigress가 MBA 난독화 옵션으로 난독화한 500개의 MBA 표현 샘플 중에서 PLASynth-SL은 68개의 반선형 MBA 표현을 얻은 반면 46개의 선형 MBA 표현을 획득할 수 있다.

[0053] 16비트 반선형 MBA 표현 f가 $"x \wedge 0x003F"$ 와 같다고 가정한다. 여기서, "0x003F"는 이종 비트값(20) 상수이며, 이것이 없으면 f는 선형 MBA 표현이 될 수 있다. 이 상수는 0000 0000 00과 11 1111로 나뉠 수 있다.

[0054] MBA 표현 f는 $f_1$과 $f_2$의 조합으로 간주된다. 여기서, $x = [X_1^{10}, X_0^6]$은 다중 입상 표현에서 $f_1 = X_1^{10} \wedge 0x00000$, $f_2 = X_0^6 \wedge 0x3F$이다. 이때, $f_1$과 $f_2$는 각각 선형 MBA 표현이며, 이는 $f_1$의 단순화된 표현, 즉 $F_1$ 이 구문적으로 $F_1'$와 동일하다는 것을 의미한다. 또한, $F_1'$은 1비트 표현으로 해석될 때 $f_1$의 단순화된 표현이다(마찬가지로 $F_2$는 $F_2'$와 구문상 동일하다.). 따라서, f에서 단순화한 최종 결과는 $(f_1 \ll 6) + f_2$로 공식화된다.

[0055] 반선형 MBA 표현 변환의 새로운 개념을 기반으로 하는 "SSLEM"의 MBA 단순화와 대상을 선형 MBA 표현으로 제한하는 논리 기반 단순화를 제안할 수 있다. 선형 MBA 표현보다 더 많은 반선형 MBA 표현이 실제 난독화된 악성코드에 존재한다고 가정한다. 비트값(20) 상수에서 동일한 비트의 더 짧은 연속 시퀀스는 결과를 복잡하게 만든다. 예를 들어, 0101 0101 0101 01 0101은 16개의 분리된 선형 표현과 단순화 결과에 15개의 이동이 필요하며 이는 원래 표현보다 간단하지 않을 수 있다.

[0056] 그러나 0x01 및 0xFE와 같은 상수 비트의 작은 수의 변경으로 프로그램 합성과 논리적으로 입증된 솔루션의 이점을 활용하여 SSLEM이 더 나은 성능을 보일 것으로 예상한다. 일부 표현은 프로그램 합성에 의해 단순화될 가능성이 있고, 일부는 논리 기반 재작성 방식으로 단순화되기 때문에 SSLEM은 기존 솔루션이 처리할 수 없는 코너 케이스를 처리할 수 있을 것으로 기대한다.

[0057] 도 3는 일 실시예에 따른 비선형 MBA 난독화 데이터를 단순화하는 방법을 도시한 순서도이다.

[0058] 도 3에서 나타낸 바와 같이, 비선형 MBA 난독화 데이터(10)를 단순화하는 방법은 추출 단계(S100), 비트 분할 단계(S300), 난독화 해제 단계(S500)를 포함하고, 통합 단계(S700)를 더 포함할 수 있다.

[0059] 추출 단계(S100)는 프로세서가 입력된 난독화 데이터(10)로부터 적어도 하나의 비트값(20)을 추출한다. 여기서, 상기 추출 단계(S100)는 상기 비트값(20)을 2진법 및 16진법 중 어느 하나로 추출하는 것을 특징으로 하는 비선형 MBA 난독화 데이터(10)를 단순화할 수 있다.

[0060] 비트 분할 단계(S300)는 상기 프로세서가 추출된 상기 비트값(20)이 변경되는 지점을 기준으로 하여 상기 난독화 데이터(10)를 분할하고, 동일한 적어도 하나의 비트값(20)을 그룹화한다. 상기 비트 분할 단계(S300)는 상기 비트값(20)이 연속적으로 구성된 선형 MBA 표현(Linear Mixed Boolean arithmetic expression)을 도출한다.

[0061] 난독화 해제 단계(S500)는 상기 프로세서가 상기 그룹화된 비트값(20)을 단순화 알고리즘으로 연산하여 그룹별로 단순화를 수행한다.

[0062] 통합 단계(S700)는 상기 프로세서가 단순화된 각 그룹의 비트값(20)을 통합하여 상기 난독화 데이터(10)가 단순화된 단순화 데이터를 생성할 수 있고, 통합된 비트값(20)과 상기 그룹별 비트값(20)을 비교하여 일치 여부를 판단할 수 있다.

[0063] 이상에서 대표적인 실시예를 통하여 본 발명을 상세하게 설명하였으나, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리 범위는 설명한 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허청구범위뿐만 아니라 특허청구범위와 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태에 의하여 정해져야 한다.

## *부호의 설명*

[0065] 100: 추출부

300: 비트 분할부

500: 난독화 해제부

700: 통합부

10: 난독화 데이터

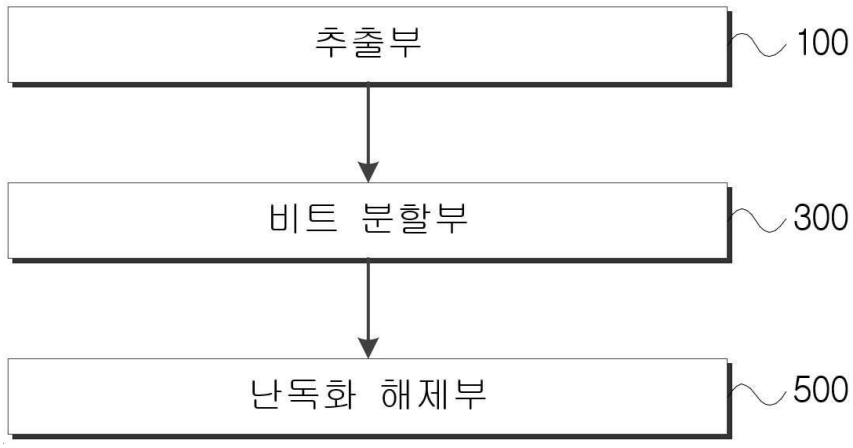
20: 비트값

30: 분할된 제1 난독화 데이터

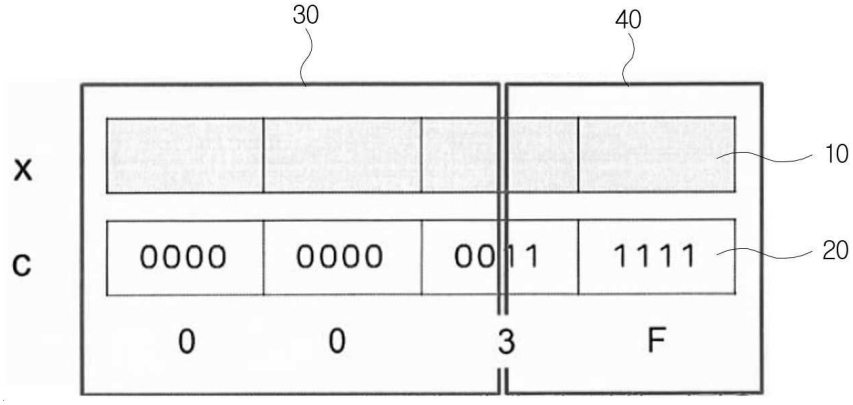
40: 분할된 제2 난독화 데이터

## *도면*

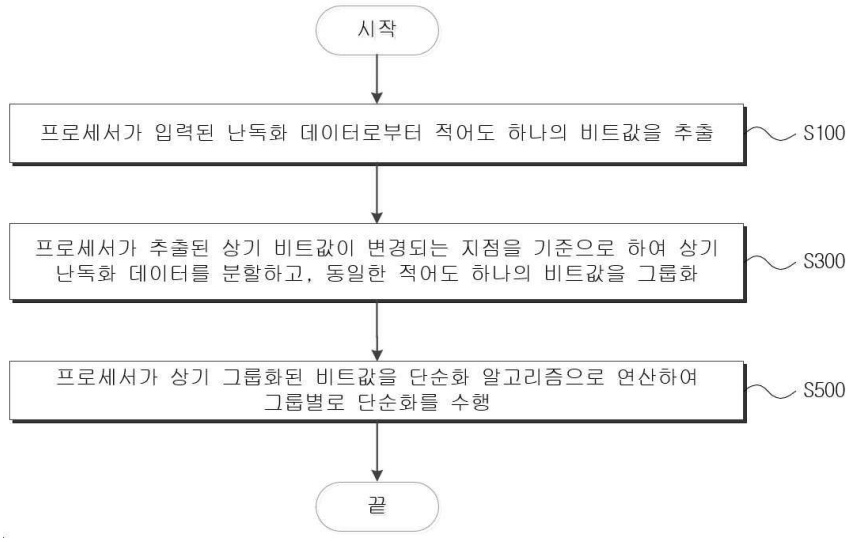
### *도면1*



도면2


30
40
X
C
10
20
0000
0000
0011
1111
0
0
3
F

도면3


시작
프로세서가 입력된 난독화 데이터로부터 적어도 하나의 비트값을 추출
S100
프로세서가 추출된 상기 비트값이 변경되는 지점을 기준으로 하여 상기 난독화 데이터를 분할하고, 동일한 적어도 하나의 비트값을 그룹화
S300
프로세서가 상기 그룹화된 비트값을 단순화 알고리즘으로 연산하여 그룹별로 단순화를 수행
S500
끝