(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2018년09월07일
(11) 등록번호 10-1896503
(24) 등록일자 2018년09월03일

(51) 국제특허분류(Int. Cl.)
G06F 21/00 (2006.01) *G06F 21/10* (2013.01)
(21) 출원번호 10-2012-0025227
(22) 출원일자 2012년03월12일
심사청구일자 2017년03월13일
(65) 공개번호 10-2013-0116408
(43) 공개일자 2013년10월24일
(56) 선행기술조사문헌
KR1020080038198 A*
US20070240141 A1*
US8127360 B1
US20080100518 A1
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
서울대학교산학협력단
서울특별시 관악구 관악로 1 (신림동)
(72) 발명자
박지순
경기 수원시 영통구 월드컵로179번길 14-10, 2동 202호 (원천동, 덕수빌라)
김진영
서울 관악구 관악로 1, 302동 312-2호 (신림동, 서울대학교)
(뒷면에 계속)
(74) 대리인
리앤목특허법인

전체 청구항 수 : 총 17 항 심사관 : 윤혜숙

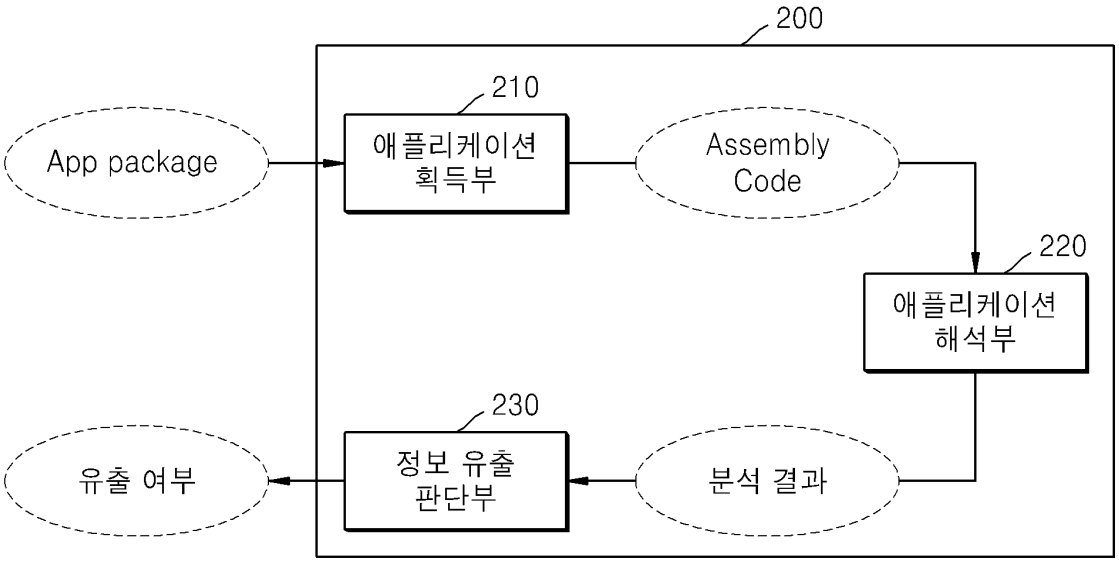
(54) 발명의 명칭 **디바이스 정보자원이 유출되는지 여부를 탐지하는 방법 및 장치**

(57) 요 약

본 발명은 디바이스의 정보자원 유출 여부를 탐지하는 방법 및 장치에 관한 것이다.

본 발명을 통해 애플리케이션으로부터 소스코드를 획득하고, 소스코드를 해석하여 디바이스의 정보자원 중에서 적어도 하나의 정보자원에 대해 수행되는 태스크를 추적하여 적어도 하나의 정보자원이 디바이스의 외부로 전송 가능한지 여부를 판단하여 애플리케이션이 디바이스로부터 정보자원을 외부로 유출시키는지 여부를 탐지하는 방법, 장치 및 기록매체가 개시된다.

대 표 도 - 도2



(72) 발명자

윤용호

서울 관악구 관악로 1, 302동 312-2호 (신림동, 서울대학교)

신준범

경기 수원시 영통구 봉영로 1526, 717동 104호 (영통동, 살구골7단지아파트)

이광근

서울 관악구 관악로 1, 302-428 (신림동, 서울대학교)

명 세 서

청구범위

청구항 1

디바이스의 정보자원 유출 여부를 탐지하는 방법에 있어서,

애플리케이션으로부터 소스코드(source code)를 획득하는 단계;

상기 디바이스의 정보자원 중에서 적어도 하나의 정보자원을 탐지 대상 정보자원으로 결정하는 단계;

상기 소스코드에 기초하여 상기 탐지 대상 정보자원에 대하여 수행되는 태스크(task)를 추적하는 단계; 및

상기 추적에 기초하여 상기 태스크에 의해 상기 결정된 탐지 대상 정보 자원이 상기 디바이스의 외부로 전송 가능한지 여부를 판단하는 단계를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 결정하는 단계는, 상기 소스코드에 기초하여 상기 애플리케이션의 액세스가 가능하다고 판단된 정보자원 중 적어도 하나의 정보자원 상기 탐지 대상 정보자원으로 결정하는 것을 특징으로 하는 방법.

청구항 3

제1항에 있어서,

상기 애플리케이션으로부터, 상기 애플리케이션이 액세스 할 수 있는 정보자원에 대한 접근 권한 정보를 포함하는 부가정보를 획득하는 단계를 더 포함하고,

상기 적어도 하나의 정보자원은, 상기 부가정보에 기초하여 상기 애플리케이션의 액세스가 가능하다고 판단된 정보자원 중 적어도 하나의 정보자원 상기 탐지 대상 정보자원으로 결정하는 것을 특징으로 하는 방법.

청구항 4

제1항에 있어서, 상기 탐지대상 정보자원은 미리 설정된 기준에 따라 결정되는 것을 특징으로 하는 방법.

청구항 5

제4항에 있어서, 상기 탐지대상 정보자원은 사용자의 입력에 기초하여 하나 이상의 정보자원을 추가 또는 제외될 수 있는 것을 특징으로 하는 방법.

청구항 6

제1항에 있어서,

상기 소스코드를 중간언어를 이용하여 중간언어코드로 변환하는 단계를 더 포함하고,

상기 추적하는 단계는 상기 중간언어코드에 기초하여 상기 탐지대상 정보자원에 대하여 수행되는 태스크를 추적하는 것을 특징으로 하는 방법.

청구항 7

제6항에 있어서,

상기 변환하는 단계는, 상기 소스코드에 포함된 하나 이상의 명령어를 소정의 기준에 따라 그룹핑하고, 상기 그룹핑된 명령어를 하나의 중간언어와 매칭하여 변환하는 것을 특징으로 하는 방법.

청구항 8

제6항에 있어서,

상기 변환하는 단계는 상기 소스코드에 포함된 하나의 명령어를 소정의 기준에 따라 하나 이상의 중간언어의 조합으로 변환하는 것을 특징으로 하는 방법.

청구항 9

디바이스의 정보자원 유출 여부를 탐지하는 정보유출 탐지장치에 있어서,

애플리케이션으로부터 소스코드(source code)를 획득하는 애플리케이션 획득부;

상기 디바이스의 정보자원 중에서 적어도 하나의 정보자원을 탐지 대상 정보자원으로 결정하고, 상기 소스코드에 기초하여 상기 탐지대상 정보자원에 대하여 수행되는 태스크(task)를 추적하는 애플리케이션 해석부 및;

상기 추적에 기초하여 상기 태스크에 의해 상기 결정된 탐지 대상 정보 자원이 상기 디바이스의 외부로 전송 가능한지 여부를 판단하는 정보유출판단부를 포함하는 장치.

청구항 10

제9항에 있어서,

상기 탐지대상 정보자원은, 상기 소스코드에 기초하여 상기 애플리케이션의 액세스가 가능하다고 판단된 정보자원 중 적어도 하나의 정보자원을 포함하는 것을 특징으로 하는 장치.

청구항 11

제9항에 있어서,

상기 애플리케이션 획득부는, 상기 애플리케이션이 액세스 할 수 있는 정보자원에 대한 접근 권한 정보를 포함하는 부가정보를 더 획득하고,

상기 탐지대상 정보자원은, 상기 부가정보에 기초하여 상기 애플리케이션의 액세스가 가능하다고 판단된 정보자원 중 적어도 하나의 정보자원을 포함하는 것을 특징으로 하는 장치.

청구항 12

제9항에 있어서, 상기 탐지대상 정보자원은 미리 설정된 기준에 따라 결정되는 것을 특징으로 하는 장치.

청구항 13

제12항에 있어서, 상기 탐지대상 정보자원은 사용자의 입력에 기초하여 하나 이상의 정보자원이 추가 또는 제외될 수 있는 것을 특징으로 하는 장치.

청구항 14

제9항에 있어서,

상기 소스코드를 중간언어를 이용하여 중간언어코드로 변환하는 중간언어변환부를 더 포함하고,

상기 애플리케이션 해석부는 상기 중간언어코드에 기초하여 상기 탐지대상 정보자원에 대하여 수행되는 태스크를 추적하는 것을 특징으로 하는 장치.

청구항 15

제14항에 있어서,

상기 중간언어변환부는 상기 소스코드에 포함된 하나 이상의 명령어를 소정의 기준에 따라 그룹핑하고, 상기 그룹핑된 명령어를 하나의 중간언어와 매칭하여 변환하는 것을 특징으로 하는 장치.

청구항 16

제14항에 있어서,

상기 중간언어변환부는 상기 소스코드에 포함된 하나의 명령어를 소정의 기준에 따라 하나 이상의 중간언어의 조합으로 변환하는 것을 특징으로 하는 장치.

청구항 17

제1항 내지 제8항 중 어느 한 항의 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체.

발명의 설명

기 술 분 야

[0001] 본 발명은 스마트 디바이스에서 실행되는 애플리케이션을 분석하는 방법 및 장치에 관한 것으로, 특히 애플리케이션이 획득한 정보자원이 애플리케이션에 의해 디바이스 외부로 유출되는지 여부를 정적 분석을 통해 탐지할 수 있는 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 스마트 폰(Smart Phone), 태블릿(Tablet), 스마트 TV, 전자 리더(ex. kindle fire) 등의 스마트 디바이스(smart device)의 사용이 일반화되고 있다. 이에 따라 스마트 디바이스에서 이용되는 애플리케이션(application)의 종류와 수가 급격히 증가하고 있다. 이러한 애플리케이션은 사용자에게 보다 편리하고 유용한 서비스를 제공하기 위해서 디바이스에 저장된 자원을 사용한다.

[0003] 스마트 디바이스에서 실행되는 애플리케이션은 애플리케이션에 포함된 부가정보에서 요청한 권한과 애플리케이션의 프로그램 소스에 의해서 주소록, 사진 등의 민감한 정보자원을 네트워크 등의 수단을 이용해 외부로 전송하는 것이 가능하다. 그러나 사용자에 의해 디바이스에 저장된 정보자원이 외부로 전송되는 경우 외에, 사용자 의사에 반하여 혹은 사용자 몰래 디바이스에 저장된 정보자원이 외부로 유출되는 경우에는 심각한 보안 문제를 초래할 수 있다.

[0004] 도 1은 스마트 디바이스 플랫폼에서 다운로드 받은 애플리케이션에 대해 동적 분석을 실시하여 민감한 정보(sensitive data)의 유출을 탐지하는 방법이다. 도 1을 참조하면, 동적 분석 방법은 애플리케이션 코드 사이에 메시지를 주고 받는 부분, 가상머신, 시스템 라이브러리, 저장매체 등에 소프트웨어 추적기를 장착한다. 각 추적기는 정보자원이 사용될 때 민감한 정보인지 여부를 함께 넘겨주며, 네트워크 등을 통해 데이터가 외부로 전송될 때, 민감한 정보가 포함되어 있을 경우 경고를 발생시키도록 하고 있다.

발명의 내용

해결하려는 과제

[0005] 동적 분석에 의한 정보자원의 유출을 탐지하는 방법은 애플리케이션의 실행과 함께 분석이 이루어지므로, 필연적으로 디바이스 혹은 애플리케이션의 성능 저하를 유발한다. 또한 플랫폼을 감염시켜서 기능을 변경하는 바이러스와는 달리 애플리케이션 자체에 구현된 기능을 검사하여야 하므로 안티바이러스의 기술 적용이 어렵다.

[0006] 또한 이러한 동적 분석 방법은 자원에 대한 사용 시나리오가 애플리케이션마다 상이하므로 배치 테스트(batch test)에 한계가 있어서 많은 수의 다양한 애플리케이션에 대한 분석이 어렵다. 동적 분석 방법에 기반한 분석의 범위는 테스트 케이스(case)에 따라 달라지는데 모든 경우의 수에 대한 분석은 현실적으로 어려우며, 분석의 신뢰도가 테스트 케이스에 좌우된다는 문제가 있다.

[0007] 본 발명은 애플리케이션의 정적 분석을 통해서 애플리케이션을 직접 실행하지 않고도 애플리케이션이 디바이스에 저장된 자원을 외부로 유출시킬 수 있는지 여부를 알려주는 방법 및 장치를 제공하여 상술한 문제점들을 해결하고자 한다.

과제의 해결 수단

[0008] 본 발명의 일 실시예에 따른 디바이스의 정보자원 유출 여부를 탐지하는 방법은 애플리케이션으로부터 소스코드(source code)를 획득하는 단계; 디바이스의 정보자원 중에서 적어도 하나의 정보자원을 탐지 대상 정보자원으로 결정하는 단계; 소스코드에 기초하여 탐지대상 정보자원에 대하여 수행되는 태스크(task)를 추적하는 단계; 및 추적에 기초하여 정보자원이 디바이스의 외부로 전송 가능한지 여부를 판단하는 단계를 포함할 수 있다.

[0009] 또한 소스코드에 기초하여 애플리케이션의 액세스가 가능하다고 판단된 정보자원 중 적어도 하나의 정보자원 탐

지 대상 정보자원으로 결정할 수 있다.

[0010] 또한 애플리케이션으로부터, 애플리케이션이 액세스 할 수 있는 정보자원에 대한 접근 권한 정보를 포함하는 부가정보를 획득하고, 부가정보에 기초하여 애플리케이션의 액세스가 가능하다고 판단된 정보자원 중 적어도 하나의 정보자원 탐지 대상 정보자원으로 결정할 수 있다.

[0011] 또한, 탐지대상 정보자원은 미리 설정된 기준에 따라 결정될 수 있고, 사용자의 입력에 기초하여 탐지대상 정보자원에 하나 이상의 정보자원이 추가 또는 제외될 수 있다.

[0012] 본 발명의 일 실시예에 따른 디바이스의 정보자원 유출 여부를 탐지하는 장치는 애플리케이션으로부터 소스코드(source code)를 획득하는 애플리케이션 획득부; 디바이스의 정보자원 중에서 적어도 하나의 정보자원을 탐지 대상 정보자원으로 결정하고, 소스코드에 기초하여 탐지대상 정보자원에 대하여 수행되는 태스크(task)를 추적하는 애플리케이션 해석부 및; 추적에 기초하여 상기 결정된 정보자원이 디바이스의 외부로 전송 가능한지 여부를 판단하는 정보유출판단부를 포함할 수 있다.

[0013] 본 발명의 다른 실시예에 따른 디바이스의 정보자원 유출여부를 탐지하는 방법은, 애플리케이션으로부터 소스코드(source code)를 획득하는 단계; 디바이스의 정보자원 중에서 적어도 하나의 정보자원을 탐지 대상 정보자원으로 결정하는 단계; 소스코드를 중간언어를 이용하여 중간언어코드로 변환하는 단계; 중간언어코드에 기초하여 탐지대상 정보자원에 대하여 수행되는 태스크(task)를 추적하는 단계; 및 추적에 기초하여 정보자원이 디바이스의 외부로 전송 가능한지 여부를 판단하는 단계를 포함할 수 있다.

[0014] 또한 소스코드를, 소스코드에 포함된 하나 이상의 명령어를 소정의 기준에 따라 그룹핑하고 상기 그룹핑된 명령어를 하나의 중간언어와 매칭하여 변환할 수 있다.

[0015] 또한 소스코드를, 소스코드에 포함된 하나의 명령어를 소정의 기준에 따라 하나 이상의 중간언어의 조합으로 변환할 수 있다.

[0016] 또한 본 발명의 실시예들에 기재된 방법은 컴퓨터로 읽을 수 있는 기록매체로 구현될 수 있다.

발명의 효과

[0017] 본 발명의 일 실시예는 정적분석 방법을 활용하므로, 기존의 동적분석과 달리 애플리케이션을 직접 실행해보지 않고 정보자원의 유출 여부를 정확하게 파악할 수 있는 효과가 있다. 또한 유출 여부를 탐지하고자 하는 정보자원 종류의 추가/제외가 용이하다. 또한 애플리케이션을 제공하는 서버, 혹은 스마트 디바이스에서 활용하는 경우에 있어서, 개인정보, 민감정보 등을 포함하는 정보자원의 유출 여부 및 그 종류를 명확하게 파악하고, 해당 애플리케이션의 설치 및 실행을 원천적으로 차단할 수 있는 효과가 있다.

[0018] 본 발명의 다른 실시예는 중간언어로 변환된 소스코드를 이용하여 애플리케이션에 의한 민감한 정보의 유출 여부를 판단하므로 분석의 속도를 높이고 시스템 자원을 절감하는 효과가 있다.

도면의 간단한 설명

[0019] 도 1은 스마트 디바이스 플랫폼에서 다운로드 받은 애플리케이션에 대해 동적 분석을 실시하여 민감한 정보(sensitive data)의 유출을 탐지하는 방법이다.

도 2는 본 발명의 일 실시예에 따른 정보자원의 유출 여부를 탐지하는 장치를 나타낸 도면이다.

도 3은 애플리케이션에 의해 정보자원에 수행되는 태스크(task)를 추적하는 일 예시이다.

도 4는 본 발명의 일 실시예에 따른 정보자원이 디바이스 외부로 유출되는지 여부를 판단하는 방법을 나타낸 흐름도이다.

도 5는 본 발명의 다른 실시예에 따른 정보유출 판단장치를 나타낸다.

도 6은 본 발명의 다른 실시예에 따른 기계어-중간언어 변환 예시를 나타낸다.

도 7은 중간언어를 이용하여 정보자원에 수행되는 태스크를 추적하는 개념도이다.

도 8은 본 발명의 다른 실시예에 따른 정보자원이 디바이스 외부로 유출되는지 여부를 판단하는 방법을 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0020] 도 2는 본 발명의 일 실시예에 따른 정보자원의 유출 여부를 탐지하는 장치를 나타낸 도면이다. 도 2를 참조하면, 정보유출 탐지장치(200)는 애플리케이션 획득부(210), 애플리케이션 해석부(220), 정보유출판단부(230)를 포함한다. 본 명세서에서 정보자원은 계정 정보, 전화번호, 이름, 생년월일을 포함하는 개인정보, 주소록, 사진, 메모, 인터넷 접속 기록 등의 사용자 경험(user experience) 등을 포함하는 민감정보(sensitive data), 기타 디바이스에 저장되고 애플리케이션이 접근할 수 있는 정보 및 데이터를 포함하는 개념으로 사용된다.

[0021] 애플리케이션 혹은 애플리케이션 패키지는 애플리케이션의 실행에 필요한 프로그램 소스파일과 애플리케이션과 관련된 서명정보, 인증정보, 퍼미션(권한) 요청 정보 포함되어 있다. 예를 들어 안드로이드 플랫폼에서 애플리케이션은 resources.arcs의 프로그램 소스파일과 AndroidManifest.xml 파일 및 기타 폴더를 포함하는 부가정보를 포함한다. 소스파일에는 복수개의 함수 혹은 명령어를 포함하는 소스코드가 기록되어 있으며, 매니페스트 파일에는 애플리케이션이 접근할 수 있는 자원 및 기능에 관한 권한 요청 정보가 포함되어 있다.

[0022] 정보유출 탐지장치(200)는 스마트 디바이스의 저장공간으로부터 스마트 디바이스에 설치된 애플리케이션을 가져올 수 있다. 또한, 유/무선 네트워크로 연결된 외부 서버로부터 애플리케이션을 수신할 수도 있다. 이 경우에 정보유출 탐지장치(200)는 애플리케이션 수신부(미도시)를 추가적으로 포함할 수 있다. 애플리케이션 획득부(210)는 획득한 애플리케이션의 소스코드를 애플리케이션이 실행되는 플랫폼에 적합한 어셈블리 코드(assembly code)로 변환할 수 있다. 혹은 애플리케이션 패키지로부터 적절한 어셈블리 코드로 변환된 소스코드를 추출할 수 있다. 안드로이드 플랫폼이라면 애플리케이션 획득부(210)는 소스코드를 애플리케이션이 실행되는 달빅 가상머신(Dalvik Virtual Machine)의 바이트코드의 형태로 추출할 수 있다.

[0023] 또한 애플리케이션 획득부(210)는 애플리케이션으로부터 부가정보를 획득할 수 있다. 부가정보에는 애플리케이션이 접근하고자 하는 자원 및 기능에 관한 퍼미션 정보가 포함될 수 있다.

[0024] 애플리케이션 해석부(220)는 애플리케이션의 프로그램 소스코드를 해석한다. 애플리케이션 해석부(220)는 소스코드를 이용하여 애플리케이션이 액세스 할 수 있는 정보자원을 판단할 수 있다. 예를 들어 소스코드에 주소록을 호출하는 함수 혹은 명령어가 포함되어 있다면, 애플리케이션 해석부(220)는 애플리케이션이 주소록 자원에 액세스할 수 있다고 판단할 수 있다. 판단된 정보자원들 중에서 하나 이상의 정보자원에 대해 유출 여부의 탐지 대상 정보자원으로 결정할 수 있다.

[0025] 또한 애플리케이션 해석부(220)는 디바이스에 저장된 정보자원이 소스코드에 의해 프로그램의 다른 부분으로 어떻게 흘러가는지 판단한다. 애플리케이션 해석부(220)는 디바이스에 저장된 정보자원 중에서, 소스코드의 제어 흐름을 분석한 결과 애플리케이션이 액세스 할 수 있다고 판단된 정보자원에 대해서 그 이동 경로를 추적할 수 있다. 도 3을 참조하여 자세하게 설명한다.

[0026] 도 3은 애플리케이션에 의해 정보자원에 수행되는 태스크(task)를 추적하는 일 예시이다. 예를 들어 사진을 찍고 웹앨범에 업로드하는 기능을 수행하는 카메라 애플리케이션을 가정하면, 단계 310에서 카메라 애플리케이션이 실행된다. 단계 320에서는 사용자의 입력에 기초하여 대상체를 촬영하게 되고, 이때 애플리케이션은 사진 데이터를 생성한다. 단계 330에서 생성된 사진 데이터는 디바이스에 저장된다. 단계 340에서 사진데이터는 웹앨범에 업로드되기 적절한 형태로 사이즈, 해상도, 용량 등이 조절될 수 있고, 단계 350에서 조절된 사진 데이터는 다시 디바이스에 저장된 후 단계 360에서 웹앨범으로 업로드된다.

[0027] 이와 같이 애플리케이션에서 수행 가능한 모든 태스크(task)는 애플리케이션의 소스코드에 기반하여 수행되므로, 소스코드의 분석을 통해 사진데이터와 같은 정보자원에 수행되는 태스크를 추적하는 것이 가능하다. 이러한 태스크 추적은 상술한 예에 국한되지 아니하고 애플리케이션의 정보자원 데이터에 대하여 수행할 수 있는 모든 종류의 태스크를 추적하는 것이 가능하다. 이러한 추적을 통해 개인정보, 민감정보 등의 데이터가 생성되는 부분에서 만들어진 값이 프로그램의 다른 부분으로 어떻게 흘러가는지 파악할 수 있다.

[0028] 다시 도 2를 참조하면, 정보유출판단부(230)는 태스크의 추적에 기초하여 정보자원의 외부 유출 여부를 판단한다. 도 3의 예시에서 촬영된 사진은 최종적으로 웹앨범으로 업로드될 수 있으므로, 이 경우 정보유출판단부(230)는 이 애플리케이션은 사진 데이터를 디바이스의 외부로 유출시킬 가능성이 있다고 판단할 수 있다. 만일 웹앨범이 로그인을 필요로 하면 애플리케이션은 사진과 함께 아이디/비밀번호를 전송할 수 있으며, 이러한 경우 소스코드는 웹 앨범으로 사진데이터와 함께 계정정보가 전송되는 것으로 분석될 수 있다. 정보유출판단부(230)는 애플리케이션 해석부(220)의 추적 결과에 이와 같이 정보자원을 디바이스의 외부로 유출시킬 수 지점이 있는

지를 탐색한다.

[0029] 이러한 정보자원의 유출 여부 판단은 플랫폼에서 디바이스에 존재할 수 있는 하나 이상의 정보자원에 대해서 수행될 수 있다. 정보자원의 유출 여부에 대한 판단은 애플리케이션의 소스코드의 해석에 기초하여 이루어지므로, 일반적으로는 해당 애플리케이션이 액세스 할 수 있는 정보자원에 대해서 이루어질 수 있다. 그러나 애플리케이션에 포함된 부가정보에서 요청한 퍼미션 정보에 의해 애플리케이션이 액세스 할 수 없는 자원에 대해서는 이러한 정보자원의 유출여부 판단이 불필요 할 수 있다. 예를 들어, 카메라 애플리케이션의 프로그램 소스코드에는 웹앨범 등에 사진을 업로드하는 코드가 포함되어 있으나, 애플리케이션에 포함된 부가정보에는 인터넷 혹은 네트워크 접근 권한의 요청이 포함되어 있지 않을 수 있다. 비록 소스코드의 해석에만 기초하면 정보자원이 외부로 유출될 수 있지만, 권한이 없어서 실제 유출은 일어나지 않기 때문에 이 경우까지 해당 정보자원의 유출가능성이 있다고 판단하는 것은 바람직하지 않다. 따라서 애플리케이션 획득부(210)에서 획득한 부가정보에 포함된 정보자원에 대한 접근권한에 기초하여 정보유출 가능성을 판단할 수 있다. 이러한 작업은 애플리케이션 해석부(220) 단계에서 부가정보를 참조하여 접근권한을 가지고 있는 정보자원 데이터에 대해서 태스크를 추척함으로써 수행될 수도 있으며, 애플리케이션 해석부(220)에서는 소스코드만 기초하여 태스크를 추적하고, 최종적으로 정보유출판단부(230)에서 부가정보와 태스크 추적 결과를 참조하여 수행될 수도 있다.

[0030] 또한 소정의 기준에 따라 미리 설정된 정보자원에 대해서 유출 여부를 판단하는 것도 가능하다. 예를 들어, 디바이스의 제원과 같이 유출여부에 크게 민감하지 않은 항목은 판단 대상에서 제외할 수 있다. 통신사 혹은 제조업체 등의 정책에 따라서 별도의 정보자원을 분류하여 그 유출여부를 판단하는 것도 가능하다. 사용자입력수신부(미도시)를 통해 유출 여부를 판단하고자 하는 정보자원의 종류를 추가하거나 삭제하는 입력을 수신하고, 그 결과에 기초하여 애플리케이션이 유출시킬 수 있는 정보자원을 판단하는 것도 가능하다.

[0031] 또한 정보유출판단부(230)는 판단결과를 디스플레이하거나 애플리케이션 제공 서버, 애플리케이션 개발자 디바이스 등에 전송하는 것도 가능하다.

[0032] 도 4는 본 발명의 일 실시예에 따른 정보자원이 디바이스 외부로 유출되는지 여부를 판단하는 방법을 나타낸 흐름도이다. 전술한 설명과 중복되는 내용은 생략한다. 도 4를 참조하면, 단계 410에서 애플리케이션으로부터 애플리케이션의 프로그램 소스코드를 획득한다. 이때 애플리케이션이 액세스할 수 있는 정보자원에 대한 접근권한에 대한 정보를 포함하는 부가정보를 추가적으로 획득할 수 있다. 단계 420에서는 획득한 소스코드의 해석에 기초하여 탐지의 대상이 되는 정보자원을 결정하고, 결정된 정보자원에 대하여 애플리케이션이 실행됨에 따라 수행될 수 있는 태스크를 추적한다. 단계 430에서는 추적 결과에 정보자원을 디바이스 외부로 유출시키는 태스크가 포함되어 있는지 판단한다. 판단 대상이 되는 정보자원은 전술한 설명에 의하여 설정, 추가, 삭제되는 것이 가능하다.

[0033] 상술한 본 발명의 실시예는 정적분석 방법을 활용하므로, 기존의 동적분석과 달리 애플리케이션을 직접 실행해보지 않고 정보자원의 유출 여부를 정확하게 파악할 수 있는 효과가 있다. 또한 유출 여부를 탐지하고자 하는 정보자원 종류의 추가/제외가 용이하다. 또한 애플리케이션을 제공하는 서버, 혹은 스마트 디바이스에서 활용하는 경우에 있어서, 개인정보, 민감정보 등을 포함하는 정보자원의 유출 여부 및 그 종류를 명확하게 파악하고, 해당 애플리케이션의 설치 및 실행을 원천적으로 차단할 수 있는 효과가 있다.

[0034] 도 5는 본 발명의 다른 실시예에 따른 정보유출 판단장치(500)를 나타낸다. 본 실시예에 다른 판단장치(500)는 애플리케이션 획득부(510), 중간언어변환부(520), 애플리케이션 해석부(530), 정보유출판단부(540)를 포함한다. 애플리케이션 획득부(510), 애플리케이션 해석부(530), 정보유출판단부(540)는 각각 도 2에 도시된 애플리케이션 획득부(210), 애플리케이션 해석부(220), 정보유출판단부(230)과 동일하거나 유사한 기능을 수행한다. 이하에서는 중복되는 설명은 생략한다.

[0035] 중간언어변환부(520)는 애플리케이션의 프로그램 소스를 분석이 용이한 중간언어 소스로 번역하는 부분이다. 애플리케이션은 일반적으로 기계어 혹은 기계어에 가까운 언어로 작성되어 있다. 이러한 기계어는 목적이 되는 기계에 따라 다르지만 일반적으로 매우 많은 종류의 명령어를 포함하며, 이러한 명령어들 중에는 핵심적인 동작은 같으나, 최적화를 위해 다양한 형태를 취하는 유사한 명령어들이 많다. 애플리케이션의 분석을 위해서 이러한 명령어들을 핵심 기능별로 묶어서 소스코드의 구조를 단순화할 수 있는 언어를 정의하여 활용하는 것이 효율적이다.

[0036] 안드로이드 플랫폼에서는 애플리케이션이 실행되는 달빅 가상머신의 기계어를 모두 220여개에 달한다. 이러한 기계어는 유사한 명령어를 포함하고 있다. 예를 들어, move는 13개, return은 4개, if는 12개의 유사한 명령어

를 포함한다. 이에 관한 자세한 설명은 도 6에서 후술한다. 본 발명의 실시예에서는 이러한 기계어를 모두 포괄할 수 있는 12개의 명령어(instruction)를 포함하는 중간언어를 정의한다. 중간언어에 포함된 명령어는 개발자에 따라서 자유롭게 설정될 수 있으며, 본 실시예에서 설명한 예시에 한정되지 아니한다. 또한, 다른 종류의 가상머신의 경우에도 같은 원리를 이용하여 중간언어를 정의하여 소스코드의 분석에 활용하는 것이 가능한다.

[0037] 도 6은 본 발명의 다른 실시예에 따른 기계어-중간언어 변환 예시를 나타낸다. OPcode는 operation code를 의미하며, 기계어의 일부이며 수행할 명령어를 지정하는 역할을 한다. 각 OPcode는 대응되는 opcode name을 가지고 있다. 예를 들어, OPcode "0"은 OPcode name "nop"에, OPcode "0E"는 OPcode name return-void에 대응된다.

[0038] 도 6을 참조하면, OPcode "01~0D" 및 OPcode "12~1C"는 모두 move와 관련된 공통적인 기능을 포함하는 명령어이므로 중간언어에서는 모두 move로 정의할 수 있다. 이와 같은 방식으로 move*, const*, unop*, binop*에 해당하는 기계어를 모두 중간언어 move로 정의할 수 있다. 이와 같은 방식으로 원래의 의미를 보전하면서 다른 기계어 명령어들도 축약할 수 있다. 또한 복잡한 기능을 하는 기계어는 중간언어의 조합으로도 표현할 수 있다. 예를 들어, "return v"는 "move rrrv ; return"을 의미하는데, 이는 중간언어 move와 return의 조합으로 표현될 수 있다.

[0039] 달빅 가상머신의 기계어를 포괄하는 중간언어는 12개의 명령어 집합(instruction set)을 가질 수 있다. 또 이 명령어 집합은 그 명령어의 속성에 따라 데이터 처리 명령어(data handling instruction)와 흐름 제어 명령어(flow control instruction) 그룹으로 분류할 수 있다. 데이터 처리 명령어에는 move, istype, new, get, put의 중간언어를 포함하고, 흐름 제어 명령어에는 call-direct, call-virtual, return, throw, jmpnz, switch, skip의 중간언어를 포함할 수 있다.

[0040] 다시 도 5를 참조하면, 중간언어변환부(520)는 애플리케이션 획득부(510)로부터 획득한 소스코드를 상술한 중간언어를 이용하여 중간언어 소스코드로 변환한다. 애플리케이션 해석부(530)은 중간언어 소스코드를 해석한다. 최초의 애플리케이션 소스코드는 그 분량이 방대하고 많은 수의 명령어와 함수로 이루어져 있어서 애플리케이션의 구조가 복잡한 경우에는 그 분석에 많은 시간이 필요할 수 있다. 그러나 중간언어로 변환된 소스코드를 이용하여 분석하면 최초의 소스코드에 포함되어 있는 유사하거나 중복되는 명령어를 일일이 분석할 필요 없이 중간언어에 의해 요약된 소스코드를 해석하는 작업을 통해서 정보자원의 흐름을 판단할 수 있다.

[0041] 도 7은 중간언어를 이용하여 정보자원에 수행되는 태스크를 추적하는 개념도이다. 도 7은 도 3의 예시를 활용한다. 도 7을 참조하면, 중간언어로 변환된 소스코드에 의해서 카메라 애플리케이션을 실행하여 대상을 촬영하는 태스크는 버퍼에 정보자원을 생성하는 태스크(710)로, 촬영된 사진을 저장하고, 편집하고, 다시 저장하는 태스크는 정보자원을 메모리, 혹은 저장매체에 기록하는 태스크(720)로, 웹앨범에 업로드 하는 태스크는 정보자원을 서버로 전송하는 태스크(730)로 해석될 수 있다.

[0042] 도 8은 본 발명의 다른 실시예에 따른 정보자원이 디바이스 외부로 유출되는지 여부를 판단하는 방법을 나타낸 흐름도이다. 전술한 설명과 중복되는 내용은 생략한다. 도 8을 참조하면, 단계 810에서 애플리케이션으로부터 애플리케이션의 프로그램 소스코드를 획득한다. 이때 애플리케이션이 액세스할 수 있는 정보자원에 대한 접근권한에 대한 정보를 포함하는 부가정보를 추가적으로 획득할 수 있다. 단계 820에서는 획득한 소스코드를 중간언어를 이용하여 중간언어코드로 변환한다. 중간언어는 최초 획득된 애플리케이션의 소스코드에 포함된 명령어를 그 공통 기능에 기초하여 그룹화하여 정의한 중간언어 명령어를 포함한다. 단계 830에서는 중간언어 코드의 해석에 기초하여 적어도 하나의 정보자원에 대하여 수행될 수 있는 태스크를 추적한다. 단계 840에서는 추적 결과에 정보자원을 디바이스 외부로 유출시키는 태스크가 포함되어 있는지 판단한다. 판단 대상이 되는 정보자원은 전술한 설명에 의하여 설정, 추가, 삭제되는 것이 가능하다.

[0043] 본 발명의 다른 실시예는 중간언어로 변환된 소스코드를 이용하여 애플리케이션에 의한 민감한 정보의 유출 여부를 판단하므로 분석의 속도를 높이고 시스템 자원을 절감하는 효과가 있다.

[0044] 본 발명에서 개시된 블록도들은 본 발명의 원리들을 구현하기 위한 회로를 개념적으로 표현한 형태라고 당업자에게 해석될 수 있을 것이다. 유사하게, 임의의 흐름 차트, 흐름도, 상태 전이도, 의사코드 등은 컴퓨터 판독가능 매체에서 실질적으로 표현되어, 컴퓨터 또는 프로세서가 명시적으로 도시되든지 아니든지 간에 이러한 컴퓨터 또는 프로세서에 의해 실행될 수 있는 다양한 프로세스를 나타낸다는 것이 당업자에게 인식될 것이다. 따라서, 상술한 본 발명의 실시예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성가능하고, 컴퓨터로 읽을 수 있는 기록매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다. 상기 컴퓨터로 읽을 수 있는 기록매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체

(예를 들면, 시디롬, 디브이디 등)와 같은 저장매체를 포함한다.

[0045] 도면들에 도시된 다양한 요소들의 기능들은 적절한 소프트웨어와 관련되어 소프트웨어를 실행할 수 있는 하드웨어뿐만 아니라 전용 하드웨어의 이용을 통해 제공될 수 있다. 프로세서에 의해 제공될 때, 이런 기능은 단일 전용 프로세서, 단일 공유 프로세서, 또는 일부가 공유될 수 있는 복수의 개별 프로세서에 의해 제공될 수 있다. 또한, 용어 "프로세서" 또는 "제어부"의 명시적 이용은 소프트웨어를 실행할 수 있는 하드웨어를 배타적으로 지칭하는 것으로 해석되지 말아야 하며, 제한 없이, 디지털 신호 프로세서(DSP) 하드웨어, 소프트웨어를 저장하기 위한 판독 전용 메모리(ROM), 랜덤 액세스 메모리(RAM), 및 비휘발성 저장장치를 묵시적으로 포함할 수 있다.

[0046] 본 명세서의 청구항들에서, 특정 기능을 수행하기 위한 수단으로서 표현된 요소는 특정 기능을 수행하는 임의의 방식을 포괄하고, 이러한 요소는 특정 기능을 수행하는 회로 요소들의 조합, 또는 특정 기능을 수행하기 위한 소프트웨어를 수행하기 위해 적합한 회로와 결합된, 펌웨어, 마이크로코드 등을 포함하는 임의의 형태의 소프트웨어를 포함할 수 있다.

[0047] 본 명세서에서 본 발명의 원리들의 '일 실시예'와 이런 표현의 다양한 변형들의 지칭은 이 실시예와 관련되어 특정 특징, 구조, 특성 등이 본 발명의 원리의 적어도 하나의 실시예에 포함된다는 것을 의미한다. 따라서, 표현 '일 실시예에서'와, 본 명세서 전체를 통해 개시된 임의의 다른 변형례들은 반드시 모두 동일한 실시예를 지칭하는 것은 아니다.

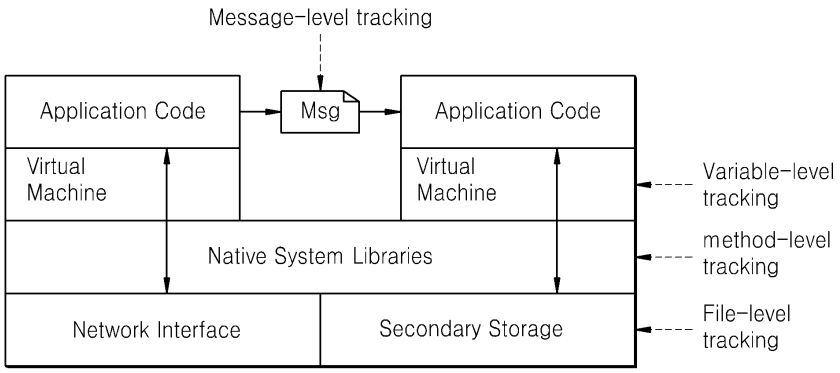
[0048] 본 명세서에서, 'A와 B 중 적어도 하나'의 경우에서 '~중 적어도 하나'의 표현은, 첫 번째 옵션 (A)의 선택만, 또는 두 번째 열거된 옵션 (B)의 선택만, 또는 양쪽 옵션들 (A와 B)의 선택을 포괄하기 위해 사용된다. 추가적인 예로 'A, B, 및 C 중 적어도 하나'의 경우는, 첫 번째 열거된 옵션 (A)의 선택만, 또는 두 번째 열거된 옵션 (B)의 선택만, 또는 세 번째 열거된 옵션 (C)의 선택만, 또는 첫 번째와 두 번째 열거된 옵션들 (A와 B)의 선택만, 또는 두 번째와 세 번째 열거된 옵션 (B와 C)의 선택만, 또는 모든 3개의 옵션들의 선택(A와 B와 C)이 포괄할 수 있다. 더 많은 항목들이 열거되는 경우에도 당업자에게 명백하게 확장 해석될 수 있다.

[0049] 이제까지 본 발명에 대하여 그 바람직한 실시예들을 중심으로 살펴보았다.

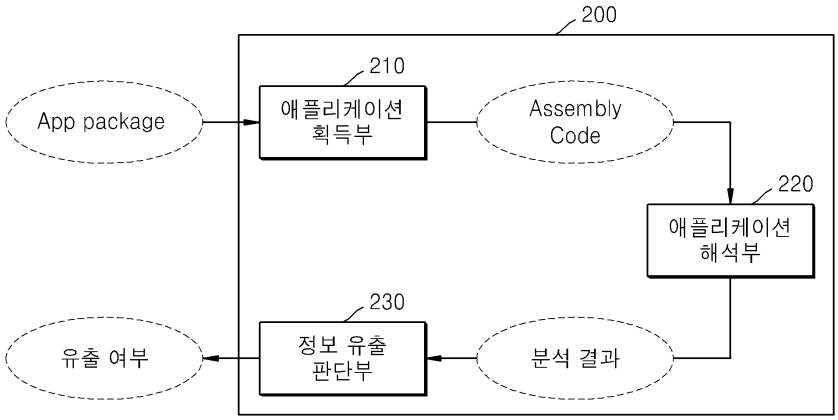
[0050] 본 명세서를 통해 개시된 모든 실시예들과 조건부 예시들은, 본 발명의 기술 분야에서 통상의 지식을 가진 당업자가 독자가 본 발명의 원리와 개념을 이해하도록 돕기 위한 의도로 기술된 것으로, 당업자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

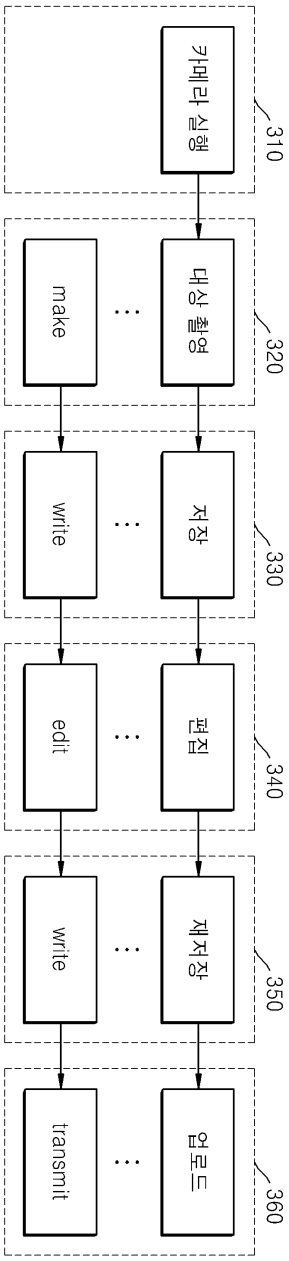
도면1



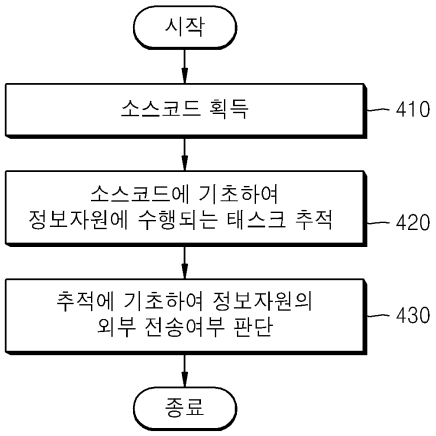
도면2


200
210
애플리케이션
획득부
App package
Assembly
Code
220
애플리케이션
해석부
230
정보 유출
판단부
유출 여부
분석 결과

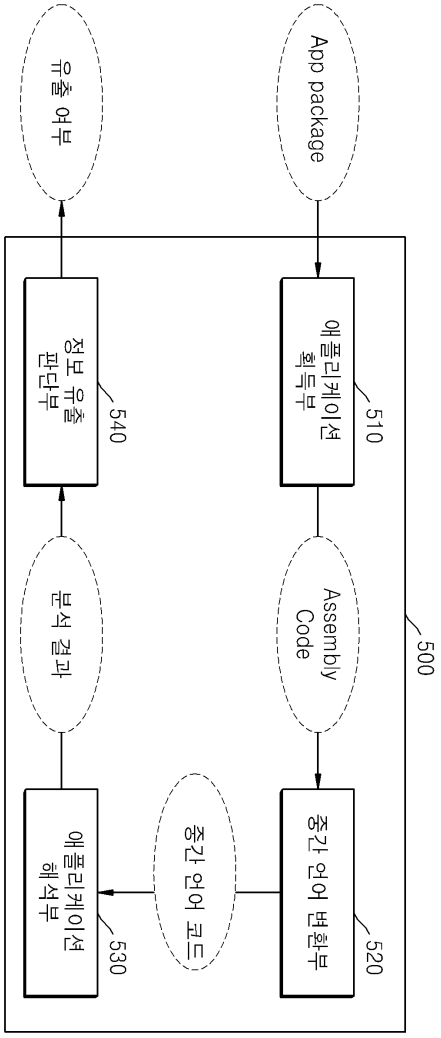
도면3


카메라 실행
310
make
...
대상 촬영
320
write
...
저장
330
edit
...
편집
340
write
...
재저장
350
transmit
...
업로드
360

도면4


시작
소스코드 획득
410
소스코드에 기초하여
정보자원에 수행되는 태스크 추적
420
추적에 기초하여 정보자원의
외부 전송여부 판단
430
종료

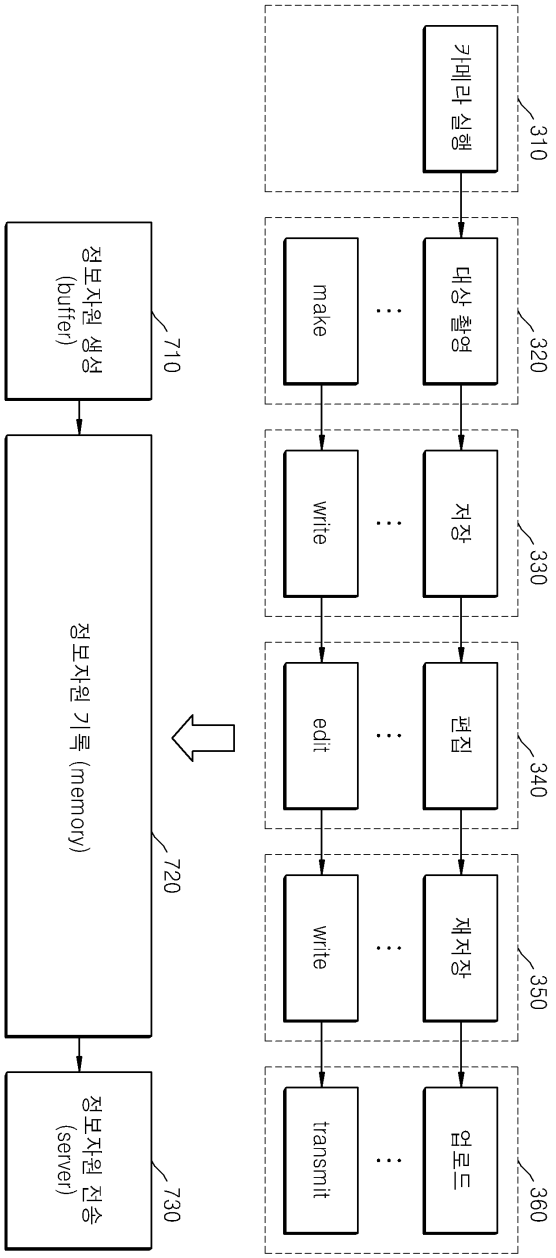
도면5


유출 여부
App package
정보 유출
판단부
540
애플리케이션
획득부
510
분석 결과
Assembly
Code
500
애플리케이션
해석부
530
중간 언어 코드
중간 언어 변환부
520

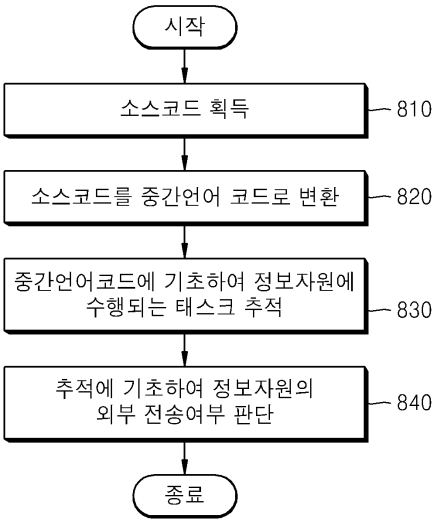
도면6

OPcode (hex)	Opcode name	중간언어 Instruction
0	nop	skip
1	move vx,vy	move
2	move/from16 vx,vy	
3	move/16	
4	move-wide	
5	move-wide/from16 vx,vy	
6	move-wide/16	
7	move-object vx,vy	
8	move-object/from16 vx,vy	
9	move-object/16	
0A	move-result vx	
0B	move-result-wide vx	
0C	move-result-object vx	
0D	move-exception vx	
0E	return-void	return
0F	return vx	combination of move and return
10	return-wide vx	
11	return-object vx	
12	const/4 vx,lit4	move
13	const/16 vx,lit16	
14	const vx, lit32	
15	const/high16 v0, lit16	
16	const-wide/16 vx, lit16	
17	const-wide/32 vx, lit32	
18	const-wide vx, lit64	
19	const-wide/high16 vx,lit16	
1A	const-string vx,string_id	
1B	const-string-jumbo	
1C	const-class vx,type_id	
28	goto target	call-direct
29	goto/16 target	
2A	goto/32 target	
32	if-eq vx,vy,target	jmpnz
33	if-ne vx,vy,target	
34	if-lt vx,vy,target	
35	if-ge vx, vy,target	
36	if-gt vx,vy,target	
37	if-le vx,vy,target	
38	if-eqz vx,target	
39	if-nez vx,target	
3A	if-ltz vx,target	
3B	if-gez vx,target	
3C	if-gtz vx,target	
3D	if-lez vx,target	

도면7


카메라 실행
310
대상 촬영
make
320
저장
write
330
편집
edit
340
재저장
write
350
업로드
transmit
360
정보자원 생성
(buffer)
710
정보자원 기록 (memory)
720
정보자원 전송
(server)
730

도면8


시작
소스코드 획득
810
소스코드를 중간언어 코드로 변환
820
중간언어코드에 기초하여 정보자원에 수행되는 태스크 추적
830
추적에 기초하여 정보자원의 외부 전송여부 판단
840
종료