

PCC 기반 안드로이드 앱 신뢰성 검증 시스템 개발 과제 완료 보고

김진영, 윤용호, 이광근

서울대학교

컴퓨터공학부 프로그래밍연구실

SW무결점연구센터(연구재단 선도연구센터)

12/15/2011 @ 삼성전자



ROSAECcenter
Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 KOSEF ERC



안드로이드 앱 신뢰성 자동 검증 시스템

- 동기: 안드로이드 앱 안심 마켓/생태계 구축
- 기술: 정적분석(static analysis) 기술



앱 신뢰성 문제 (1/2)

안드로이드 앱의 개인정보 누출

chosun.com

갤럭시S의 '가을 앱', 주소록과 통화녹음 유출 가능

NEWSis. ()

주식앱으로 개인정보 빼낸 증권방송 간부 기소

서울신문

[당신의 스마트폰은 안녕하십니까] '오픈소스' 안드로이드 폰 보안 장치 안하면 더 위험

dongA.com

[뉴스 파일] 안드로이드폰 악성 프로그램 주의보

연합뉴스

"혹시 내 위치정보도..." 스마트폰 불안감 확산

경향신문

"누군가 당신의 개인정보를 노린다" ... 스마트폰보안 비상

NETWORKWORLD

Many Android apps leak user privacy data

Researchers find permitted apps transmit phone numbers, location, and SIM card IDs

cnet News

Google Android apps found to be sharing data

September 26, 2013 4:52 PM PDT

What's that Android app doing with my data?

The Register

2 out of 3 Android apps use private data 'suspiciously'
Google protections 'insufficient'

WIRED

Study Shows Some Android Apps Leak User Data Without Clear Notifications

BBC

Slashdot

Your Rights Online: Many More Android Apps Leaking User Data

msnbc.com

Smartphone Apps Spread Personal Info, Study Finds



ROSAECcenter

Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 KOSEF ERC

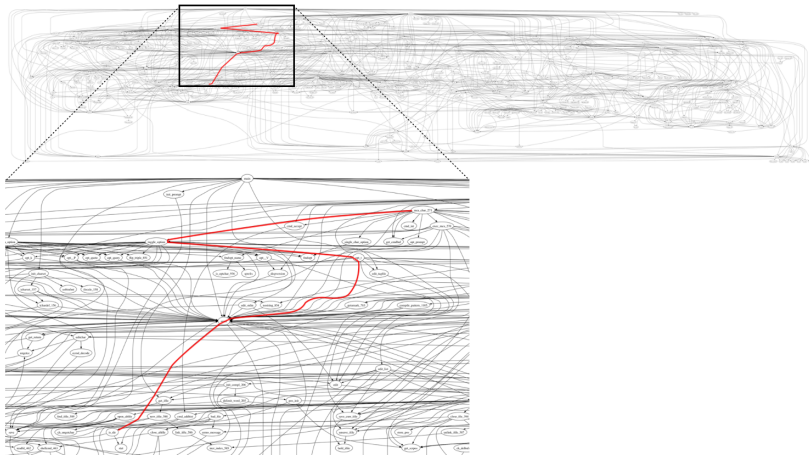


앱 신뢰성 문제 (2/2)

- 현재는 **권한** 명시 방식
 - 각 앱이 필요한 권한을 요청하도록
 - 믿고 쓰거나, 아니면 설치하지 않거나
- 문제는
 - **인플레**: 과도한 권한 명시 \Rightarrow 효력상실
 - **정교하지 않음**:
 - 실제로 어떤 정보에 접근했는가?
 - 읽기만? 아니면, 누출까지?
- 민감한 정보 누출이 발생할 수 있는 지를 미리 분석하면 해결



실행복잡도의 실제 × 난독화된 소스 × 앱 기계어 Dalvik



Project SCANDAL



- **기반 이론:** 요약해석(abstract interpretation)
 - 강력: any property + any language
- **엔지니어링 실제**
 - 노하우: 분석비용 vs 분석정확도의 발란스
 - 앱 기계어(Dalvik) 대상 (not Java, not C)



추적하는 민감한 정보의 흐름 경로

- 민감한 정보들은
 - 전화번호부, 문자메시지
 - 위치정보
 - 사진, 영상
 - 기기 고유번호(IMEI)
- 새어나가는 구멍은
 - 인터넷
 - URL에 담아서
 - 서버로 직접
 - 문자메시지로



민감한 정보 누출의 예 (1/5)

앱 : Google Wallpaper 4.2.2

```
Wallpapers.onCreate(Bundle)
```

```
...  
callv TelephonyManager.getDeviceId()  
move-result r3  
puts r3 eWallpaperConst.IMEI
```

기기 고유번호

```
SearchTagsActivity.initTagWebView()
```

```
...  
get r0 r3 SearchTagsActivity.mWebView  
get r1 r3 SearchTagsActivity.mSharedPreferences  
calld XMLTools.getSearchURL(r1)  
move-result r1  
callv WebView.loadUrl(r1)
```

initTagWebView()

getSearchURL()

getLocale_version_IMEI_W_H0



민감한 정보 누출의 예 (2/5)

앱 : Google Wallpaper 4.2.2

```
Wallpapers.onCreate(Bundle)
```

```
...
```

```
XMLTools.getSearchURL(SharedPreferences)
```

```
...
```

```
calld XMLTools.getLocale_version_IMEI_W_H(r4)
```

```
move-result r2
```

```
callv StringBuilder.append(r1,r2)
```

```
move-result r1
```

```
Se
```

```
calld EWallpaperHttpHelper.getSignatureParamString()
```

```
move-result r2
```

```
callv StringBuilder.append(r1,r2)
```

```
move-result r1
```

```
callv StringBuilder.toString(r1)
```

```
move-result r0
```

```
return r0
```

```
initTagWebView()
```

```
getSearchURL()
```

```
getLocale_version_IMEI_W_H()
```



ROSAEC center

Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 KOSF ERC

민감한 정보 누출의 예 (3/5)

앱 : Google Wallpaper 4.2.2

```
Wallpapers.onCreate(Bundle)
...
XMLTools.getSearchURL(SharedPreferences)
move-result r3
```

XMLTools.getLocale_version_IMEI_W_H(SharedPreferences)

```
...
gets r5 eWallpaperConst.IMEI
callv StringBuilder.append(r4,r5)
move-result r4
callv StringBuilder.toString(r4)
move-result r4
return r4
```

기기 고유번호

initTagWebView0

getSearchURL0

getLocale_version_IMEI_W_H0



민감한 정보 누출의 예 (4/5)

앱 : Google Wallpaper 4.2.2

Wallpapers.onCreate(Bundle)

XMLTools.getSearchURL(SharedPreferences)

...

callv XMLTools.getLocale_version_IMEI_W_H(r4)

move-result r2

callv StringBuilder.append(r1,r2)

move-result r1

callv EWallpaperHttpHelper.getSignatureParamString()

move-result r2

callv StringBuilder.append(r1,r2)

move-result r1

callv StringBuilder.toString(r1)

move-result r0

return r0

"http://www.imnet.us/api/wallpapers/photos/
search_keywords?" + IMEI + SignatureParamString

initTagWebView()

getSearchURL()

getLocale_version_IMEI_W_H()

기기 고유번호



ROSAEC center

Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 KOSF ERC



민감한 정보 누출의 예 (5/5)

앱 : Google Wallpaper 4.2.2

```
Wallpapers.onCreate(Bundle)
```

```
...  
callv TelephonyManager.getDeviceId()  
move-result r3  
puts r3 eWallpaperConst.IMEI
```

기기 고유번호

```
SearchTagsActivity.initTagWebView()
```

```
...  
get r0 r3 SearchTagsActivity.mWebView  
get r1 r3 SearchTagsActivity.mSharedPreferences  
callv XMLTools.getSearchURL(r1)  
move-result r1  
callv WebView.loadUrl(r1)
```

"http://www.imnet.us/api/wallpapers/
photos/search_keywords?" + IMEI
+ SignatureParamString

```
initTagWebView()
```

```
getSearchURL()
```

```
getLocale_version_IMEI_W_H0
```



KSAEC center

Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 KOSF ERC

비공식 마켓 앱 (1/2)

Android Market



Monkey Jump 2
DS Effects

INSTALL

INSTALLS:
100,000 - 500,000



last 30 days

VS

Free Monkey Jump 2 Download

[Monkey Jump 2 2.1](#)

The second episode of Monkey Jump, help the monkey to collect

儿童 > Monkey Jump 2

**Monkey Jump 2 2.1
download**

Monkey Jump 2 2.1 - [more info](#)

Monkey Jump 2

★★★★★ (0份评价)

超级猴子跳 **Monkey Jump 2 2.1**

- 비공식 마켓의 변조된 앱
 - apk 파일을 손쉽게 다운로드 가능
 - 공식 마켓의 앱과 같은 기능
 - 민감한 정보를 몰래 수집

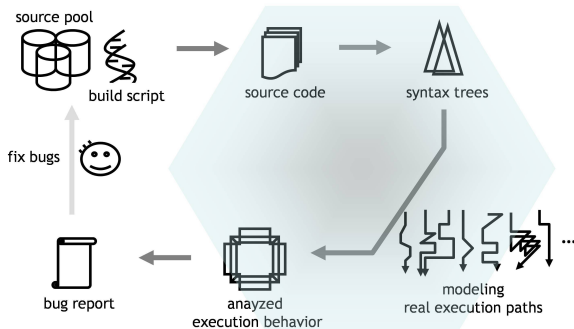


SAECcenter
Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 KOSEF ERC

비공식 마켓 앱 (2/2)

- 변조된 앱들
 - 게임(Monkey Jump 2, Gold Miner, Mini Army, Baseball Superstars 2010, Shot Gun Free)
 - 생활 앱(Xing Metro)
- SCANDAL로 분석 결과
 - 공식 마켓의 앱에서는 악성 서버로의 누출 없음
 - 비공식 마켓의 앱에서는 악성 서버로의 누출 검출



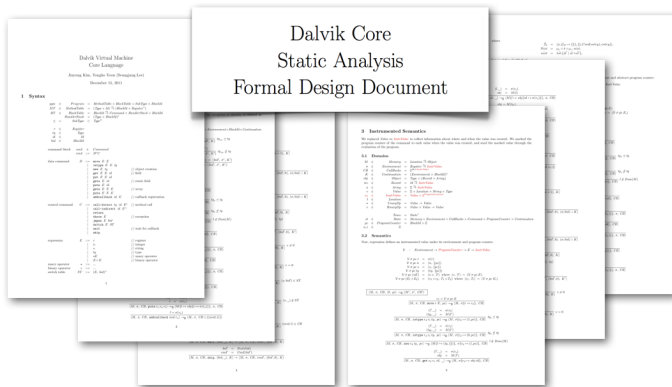


Dalvik 언어 전처리 인프라

- 앱 기계어 Dalvik
 - 220여개의 명령어
 - 같은 일을 하는 서로 다른 명령어가 많음
 - 모두에 대한 의미 정의, 분석기 설계는 낭비
- 분석용 Dalvik Core 정의 + 번역기구축
 - 12개의 명령어로 충분
 - 데이터 접근: `move`, `istype`, `new`, `get`, `put`
 - 흐름 제어: `call-direct`, `call-virtual`, `return`, `throw`, `jmpnz`, `switch`, `skip`
 - 단, 원래 의미를 보존하도록
 - `move*`, `const*`, `unop*`, `binop*`, ... → `move`
 - `return v` → `move rr rv ; return`



엄밀한 디자인



- 요약해석 이론 기반
- 앱 기계어(Dalvik) 대상

김진영, 윤용호, 이광근

ROSAECcenter
Research On Software Analysis for Error-free Computing
소프트웨어 무결성 연구센터 KOSF ERC



ScanDal : 안드로이드 앱 신뢰성 검증 시스템

Dalvik 프로그램 정적 분석을 어렵게 하는 특성들과 SCANDAL의 대응



- 객체 지향 프로그램(OOP)의 핵심
 - 타입의 상속을 활용(Subtype Systems)
 - 타입에 따라 다른 메소드 호출
- 우리의 분석기는
 - 타입 정보도 분석에 포함
 - 실제로 일어나지 않는 호출을 안전하게 제거



- Java의 콜백 메소드(callbacks)
 - 시스템이 호출하는 메소드
 - 프로그램 코드에선 직접 호출되지 않는
 - 프로그램 외부와의 상호작용을 담당
 - 유저의 입력(화면 터치, 버튼, 위치 변화)
 - OS의 상태변화(WiFi On/Off, 배터리 부족)
- 우리의 분석기는
 - Listener를 다루는 명령어를 핵심언어에 추가
 - Listener를 환경에 등록하는
 - 이벤트를 기다리다 적절한 Listener를 호출하는
 - 이벤트를 기다리는 명령어를 프로그램 맨 끝에 무한 루프 후



- Java의 예외상황 관리 방법
- 예외(Exception) 타입의 객체를 던짐
- 메소드 호출 관계와 예외 객체의 타입에 따라 예외를 처리하는 부분이 달라짐
- 우리의 분석기는 두 가지 모두 안전하게 분석



- 문자열로부터 클래스, 객체를 만들고 함수를 호출
- 단순한 형태의 다단계 프로그래밍
- 우리의 분석기는
 - 대부분의 프로그램은 Reflection에 문자열 상수를 이용하므로 정확히 분석
 - 상수가 아닌 경우도 안전하게 분석



- Java와 Android의 풍부한 Library
- 개발자에게 편리, 정적 분석에는 걸림돌
 - 무슨 일을 하는 메소드인가?
 - 프로그램 코드에는 드러나지 않는
- 자주 쓰이면
 - 메소드의 실행 의미를 하드코딩
 - 단, 분석하고자 하는 성질에 따라 필요한 것만
- 자주 쓰이지 않으면
 - 가장 안전하게 메소드의 실행 의미를 가정하고 분석



기술 소개 & 기회



정적 프로그램 분석(static program analysis)

프로그램의 실행 성질을
실행전에 자동으로
안전하게 어림잡는
일반적인 방법

- 응용: sw 오류검증, sw 관리, sw 테스트, sw 최적화, 등등
- 다양한 레벨/목적/이름으로 존재:

theory	“요약해석 abstract interpretation”
pl, se, veri.	“type system”, “model checking”, “theorem proving”
cmplr	“data-flow analysis”, etc.

 **SAEC** center
Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 KOSEF ERC



정적 프로그램 분석

- “**실행전**”: 프로그램을 실행시키지 않고
- “**자동으로**”: 프로그램이 프로그램을 분석
- “**안전하게**”: 모든 가능성을 포섭
- “**어림잡는**”: 실제 이외의 것들이 포함됨
 - 어림잡지 않으면 불가능
- “**일반적**”: 소스언어와 성질에 무제한
 - C, Java, ML, Bluespec, Dalvik, x86, binary, etc.
 - “buffer overrun?”, “memory leak?”, “unhandled exn?”
“x=y at line 2?”, “memory use \leq 2K?”, “terminate?”,
“race?” etc.
- “**One-on-One**”: 분석기1개당, 1언어 1성질.



모든 정적 분석은 3스텝

1. “연립방정식”을 세운다
 - 요약된 세계에서 (abstract semantic domains)
 - 프로그램의 실행ダイナ믹스에 관한 (abstract execution flows)
2. 그 방정식을 푼다
3. 그 해를 가지고 결론을 내린다
 - 있는가 없는가? 같은가 다른가?

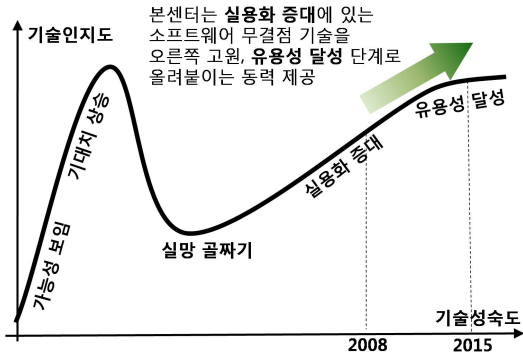


다른 분야와 다르지 않은

<u>소프트웨어</u>		<u>기계공학</u>
프로그램	↔	기계디자인
실행 = 컴퓨터 (언어정의)	↔	작동 = 자연 (자연법칙)
실행에대한 방정식	↔	작동에대한 방정식
방정식 풀기	↔	방정식 풀기
“생각대로 돌것이다”	↔	“생각대로 작동할것이다”
“무인비행기에 심자”	↔	“만들어 팔자”
PL & Logic 이론	↔	물리-화학법칙, XX방정식



기술 위치 (우물안)



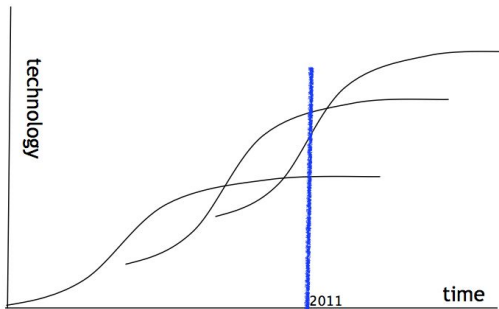
ROSAECcenter

Research On Software Analysis for Error-free Computing

소프트웨어 무결점 연구센터 KOSEF ERC



기술 단계 (우물박)



안드로이드 앱 신뢰성 자동 검증 시스템

- 동기: 안드로이드 앱 안심 마켓/생태계 구축
- 기술: 정적분석(static analysis) + α 기술





● SCANDAL/X

- X = 민감정보 누출 검증 (SCANDAL/privacy)
- X = 다른 악성 실행 가능성
- 기술: 정적분석 × 자동증명기술
- Cloud system sw 검증/안심 생태계 공헌
 - e.g. Xen Hypervisor core 검증
 - 기술: 정적분석 × 자동증명기술

감사합니다

