Vol.2018-CSEC-83 No.18
2018/12/14

Android向けPUAと正規アプリ間のAPI使用傾向の比較調査

伊藤 克恭[†1,a)]　長谷川 皓一[†2,b)]　山口 由紀子[†3,c)]　嶋田 創[†3,d)]

概要：近年スマートフォンの利用が急速に普及し，様々なアプリが利用できるようになっている．特にAndroidでは公式のアプリストアだけでなく，サードパーティのストアも手軽に利用可能である．その中には一見便利な機能を備えているが，ユーザの意図した目的とは関係ない個人情報にアクセスし，窃取することを主目的としたアプリであるPUAが存在し，問題となっている．これに対して我々はアプリケーションが呼び出すAPIの頻度に着目し，統計的手法によりPUAを検知する手法を提案している．本稿では，PUA検知精度の向上を目的とし，サードパーティストアから収集したPUAと正規アプリ間のAPI使用傾向の差の詳細分析を行った．その結果，正規アプリは画面の体裁や操作性などのUIに関わるAPIが，PUAは通信状況の調査に関わるAPIが高頻度で使用されていることが明らかになった．

キーワード：Android, PUA, 個人情報窃取, APIコール, 統計解析

Survey about API Usage Tendencies between PUAs and Proper Applications of Android Devices

Ito Katsutaka[†1,a)]　Hasegawa Hirokazu[†2,b)]　Yamaguchi Yukiko[†3,c)]　Shimada Hajime[†3,d)]

Abstract: In these years, the use of smartphone spreads rapidly and various applications are available. Especially in Android, it can use not only official application store but also third-party application stores with ease. There are applications called PUA and that is becoming serious problem. They seem to be convenience but access and steal privacy information unrelated with users' intense. Thus, we proposed a method to detect those type of applications by statistical method. It focus on frequency of API calls which show the activity of applications. In this paper, for the purpose of improving detection method, we analyzed the difference of API usage between proper applications and PUAs in third-party application stores. We found that proper applications use APIs to show UI well and PUAs use APIs for surveying condition of network and terminal.

Keywords: Android, PUA, Leakage of Privacy Information, API Calls, Statistic Analysis

1. 序論

近年スマートフォンの利用が広がり，総務省の調査によると約57%の日本人がスマートフォンを利用している [1]．特に20代から30代の若い世代に関しては9割以上の人がスマートフォンを利用している．

その一方でスマートフォンをより便利に利用するためのアプリケーションも豊富に存在している．IDC (Internationl Data Corporation) の調査 [2] によると約85%のス

[†1] 名古屋大学情報学研究科
Graduate School of Informatics, Nagoya University
[†2] 名古屋大学情報戦略室
Information Strategy Office, Nagoya University
[†3] 名古屋大学情報基盤センター
Information Technology Center, Nagoya University
[a)] itokatu@net.itc.nagoya-u.ac.jp
[b)] hasegawa@icts.nagoya-u.ac.jp
[c)] yamaguchi@itc.nagoya-u.ac.jp
[d)] shimada@itc.nagoya-u.ac.jp

マートフォンユーザが Android を利用している．Android は公式のアプリストアだけでなく，サードパーティのストアからも様々なアプリケーションを容易に利用可能であるという特徴を持つ．サードパーティストアではアプリケーションのプライバシー利用の適切性に関する事前審査がなく，中には便利な見た目とは関係なく機能外の個人情報にアクセスし窃取することを主目的としたアプリケーションである PUA(Potentially Unwanted Application) の一種が存在している．

情報処理推進機構 (IPA) の調査によると，約 71%の人が個人情報漏洩を心配しているのに対して，アプリケーションのパーミッションを適切に設定する等の対策をしている人は約 19%にとどまる [3]．そのため，機能に関係のない不正なパーミッションが見逃され，意図せず個人情報が窃取される危険性が高い人が多い状況と言える．

本研究は，個人情報の漏洩を恐れながらもパーミッション設定を気にしていない人を，PUA が及ぼす情報窃取の脅威から守るために，PUA の特徴を分析し検知する手法を提案することを目的としている．本稿では検知すべき PUA の傾向を捉えるために，アプリケーションの行動が現れると考えられる Android API に着目して，特徴分析のために正規アプリ（3.2 節で定義）との差を分析して特徴を調査する．正規アプリと PUA の間で API の呼出頻度の平均を元にした統計的分析を行い，差が大きい API や API クラス，および使用頻度が高い API の比較を行った．

以下に本稿の構成を示す．2 章では，Android の不正アプリケーションの検知に関わる関連研究を挙げる．3 節では，API を用いて分析を行う本研究の概要について説明する．4 節では，API コールに着目した正規アプリと PUA の比較調査の方法について説明する．5 節では API コールの分析結果と考察を示す．6 節では結論と今後の課題について述べる．

2. 関連研究

Android は iOS と異なり，審査のないサードパーティアプリケーションを手軽に利用できることから，Android はセキュリティのリスクを負いやすいと考えられる．このリスクに関連する研究として，インターネット上で公開されているアプリなどの実態調査に関わる研究が行われている．

畑田らは，PUA に関するセキュリティアラートのトリアージを行うために，DNS クエリに基づいて PUA の脅威の分類を行うことを提案している [4]．この研究において，同種の PUA が行う DNS クエリに類似度が高いこと，Android が行うものでない DNS クエリについて類似度を学習させ分類を行った結果，PUA を 9 割以上の割合で検知できたこと，亜種の分類も 8 割以上の正確さで可能であることが示されている．韓らは，端末情報の取得を行う API とシステムコールを特徴量とした K-means ベースの機械学習を用いて Android マルウェアによる個人情報窃取を検知する手法を提案している [5]．システムコール呼出回数のみで個人情報窃取の検知を行う先行研究と比較して，検知精度が向上したと述べている．

細谷らは，公式ストアのアプリケーションを収集し，APK 内のコードを解析し API に対応付けることにより，どのような情報をアプリケーションが使用しているか分析を行っている [6]．約 7 割のアプリケーションが位置情報を取得していること，約 4 割のアプリケーションが端末情報を取得していることから，多くのアプリケーションにおいてユーザトラッキングされる危険性があることが述べられている．三村らは，Twitter 投稿のリンク先の APK を収集し解析を行い，特徴や危険性を分析している [7]．その中には，セキュリティリスクを伴うルート化を行うアプリケーションや，既存アプリの一部を改変して再配布されるリパッケージアプリの存在が明らかになり，Twitter 投稿のリンク先の APK をインストールする危険性が示唆されている．

3. Android API に着目したアプリケーションの分類手法

3.1 API に着目した動機

PUA は，スケアウェアによるインストールの強要や，一見便利な見た目，宣伝文の利用を行いつつサードパーティストアで配布されることにより，広まると考えられる．1 節で示したように，多くの人が個人情報流出を恐れながらもパーミッション等の設定を気にかけていない．そのため，このような人が PUA をインストールすると同時に情報取得に関連するパーミッションを与えると，意図せず個人情報が流出することが危惧される．正規アプリも PUA も，初回起動時やインストール時にユーザへパーミッションを要求し，許可を得た個人情報に対してアクセスを行う点では，Android の保護システムやメモリアクセスなどの挙動から見て個人情報を取り扱う様子に差はない．そこで我々は，機能外の個人情報にアクセスするアプリケーションを検知するために，アプリケーションの行動の全容が把握できると考えられる Android API に着目して，特徴を分析し，アプリケーションの分類を試みた．

3.2 アプリケーションの分類

本研究では，アプリケーションが求めるパーミッションと，機能の実現に必要となるパーミッションの関係を元に，アプリケーションを以下に示す 2 種類に分類を行う．

- 正規アプリ
  アプリケーションの目的や機能に関わる個人情報のみに対してパーミッションを要求し，使用するもの
- PUA
  機能に関係のない個人情報に対してパーミッションを要求し，実際にアクセスするもの

3.3 これまでの研究

これまで我々は，個人情報にアクセスする API(ACCESS) とネットワーク送信時に呼び出される API(SEND)，ユーザに処理結果を示す際に呼び出される API(RESPONSE) を定義し，個人情報を適切に使用するアプリケーション群と PUA 群の間で比較し分析した [8]．また，サードパーティストア内のアプリケーションも検査するには端末上で PUA 検知を行う必要があり，軽量かつ単純な方法が必要となる．そこで，API の呼出頻度の平均値を用いた特徴分析を提案した．その結果，アプリケーション群で ACCESS と SEND の比に約 11 倍の差があることが判明し，PUA 検知の足がかりを得た．本稿では正規アプリと PUA 間のさらなる API コールの特徴の差を得るため，API コール分析の範囲を広げ，Android API 全体を分析し，PUA 群と個人情報を適切に扱うアプリケーションとの API 呼出の特徴の差を調査した．

3.4 API コールログ記録環境の構築

Google Developer の調査 [9] によると 2018 年 10 月 26 日時点で一番ユーザが多い Android のバージョンは 6.0(Marshmallow) であり，それを用いて実験を行う．一方で，デフォルトの Android は API コールログを出力する機能を持たないため，Android 6.0 のソースコードを改変し，API コールログを出力できるようにした [8]．

4. Android API の分析手法

1 節で述べたように，本研究では API コールを元に PUA を検知する手法を提案することを目的として研究を進めている [8]．より高精度な検知の実現を目指すため，本稿では，API コール全体の傾向の差について PUA を正規アプリとを比較した際に現れる特徴を細かく分析する．

本稿で行った API コール分析の流れを図 1 に示す．まず，サードパーティストアから個人情報を扱う様々なアプリケーションを収集する．次に，API コールログを出力できるように改変した Android 上で収集したアプリケーションを実行し，API コールログを得る．アプリケーションが要求したパーミッションと API コールログから，PUA と正規アプリに分類し，それぞれで API コールログに関して特徴分析を行う．

評価に用いたアプリケーションは，1Mobile Market[*1]や baidu アプリストア[*2]などのサードパーティストアを通じて収集を行った．収集したアプリケーションを，改変を行った Android 上で実行し，アプリケーションの起動から，アプリケーションの目的の達成まで API コールログ記録した．この際，アプリケーション一つ一つのパーミッションや実際に要求された個人情報にアクセスしているかを確認しながら手動でアプリケーションとログの収集を行ったため，評価に用いたアプリケーション数は少数となってしまった．結果，正規アプリを 68 個，PUA を 31 個の API コールログを収集することができた．表 1 に収集した API コールログの諸元を示す．

[*1] http://www.1mobile.com/
[*2] https://shouji.baidu.com/

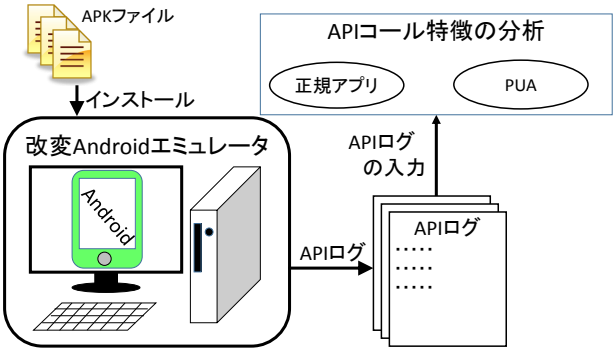

図 1: API コール分析手法の概要
Fig. 1 Overview of API Call Analysis Method.

表 1: 収集した API コールログの長さ
Table 1 Length of Collected API Call Logs.

	アプリ数	最短	最長	平均
正規アプリ	68	931	20,477	10,273
PUA	31	3,144	20,071	10,614

収集した API コールログより，アプリケーション毎に各 API や API クラスの呼出頻度を計算し，正規アプリ群と PUA 群それぞれで平均値を求めた．その結果を元に，以下の 3 種の分析を行った．なお，API クラスおよび API の種類について Google Developer 提供の仕様 [10] を調査し，93 クラスと 1717 種の API を特定した．

調査 1: API クラス毎の呼出頻度差の調査
93 クラスの API クラスの出現頻度に関して，PUA と正規アプリそれぞれ平均を計算した．出現頻度平均の比が大きかったクラスを 5.1 節にまとめた．

調査 2: API 毎の呼出頻度差の調査
1717 種の API の出現頻度に関して，PUA と正規アプリでそれぞれ平均を計算した．出現頻度平均の比が大きかった API を 5.2 節にまとめた．

調査 3: 最頻呼出 API 上位 50 種の比較
1717 種の API から出現頻度の平均が高い API 50 種を PUA と正規アプリでそれぞれ求め，どのような差が見られるか比較調査を行った．その結果を 5.3 節にまとめた．

API コールログの長さは表 1 に示すように様々であり，各 API の呼出回数の平均値を元にした分析を行うと，ログの短いアプリケーションの特徴が現れにくく，また，ログの長いアプリケーションの特徴が平均値に大きく影響することとなる．そこで本研究では，各アプリケーションの数値を均等に考慮するため，ログ全体に対する API の呼出頻

表 2: 呼出頻度が高い API クラスの比較
Table 2 The Comparison of Frequently Called API Classes.

(a) 正規アプリが多く呼び出すクラスの頻度
Frequency of API Classes Proper Apps Call Frequently

クラス名	正規アプリ (%)	PUA (%)	比
animation	1.95	0.88	2.2
hardware	0.64	0.32	2.0
app.usage	0.25	0.16	1.5
graphics.drawable	5.18	3.80	1.3
database	3.25	2.44	1.3

(b) PUA が多く呼び出すクラスの頻度
Frequency of API Classes PUAs Call Frequently

クラス名	正規アプリ (%)	PUA (%)	比
webkit	0.17	0.38	2.3
telephony	0.11	0.25	2.3
content.pm	2.61	4.80	1.8
net	1.73	3.13	1.8
view.inputmethod	0.19	0.27	1.4

度を用いて分析を行った．なお本稿では，Android API の API 名に関して，共通の接頭語 ‘android.’ は省略して表記を行う．

5. 調査結果

5.1 API クラス毎の呼出頻度差の調査

Android API 93 クラスのうち，正規アプリと PUA の双方で呼び出された 53 クラスについて呼出頻度を調査した．正規アプリでの呼出頻度が 0.2%以上の API クラスについて PUA における呼出頻度との比が大きい上位 5 クラスを表 2(a) に示す．同様に PUA での呼出頻度が 0.2%以上の API クラスについて正規アプリにおける呼出頻度との比が大きい上位 5 クラスを表 2(b) に示す．表 2 から分かるように，正規アプリでは `animation` や `graphics.drawable` といった UI に関わる API クラスが多用されている．一方 PUA では，`telephony` や `net` といった通信状況の調査に用いることができる API クラスが呼び出されていた．しかしながら，API クラスの中には用途の異なる API が同一クラスに含まれているものもあるため，個々の API による呼出状況を調査することにした．

5.2 API 毎の呼出頻度差の調査

正規アプリと PUA でそれぞれ頻繁に利用される 1717 種の API の個別の呼出頻度を調査した．正規アプリと PUA の間で呼出頻度が 5 倍以上異なるクラスに着目し調査した．なお，正規アプリのみで呼び出された API が 109 種，PUA のみで呼び出された API が 41 種存在したが，本節の分析対象から除外した．API 呼出ログを確認した結果，これらは，2 種類以上のアプリケーションから呼び出された物が少なかったり，1 つのアプリケーションが特に多種の API を呼出ている事例が確認されたりしたため，収集したアプリケーションの偏りによりたまたま少数回呼び出された API であると考えたためである．結果，正規アプリの呼出頻度が PUA における呼出頻度の 5 倍以上となる API は 24 種，PUA の呼出頻度が正規アプリにおける呼出頻度の 5 倍以上となる API は 24 種となった．

表 3: 正規アプリが PUA の 5 倍以上の頻度で呼び出す API
Table 3 APIs Called Freqently by Proper Apps 5 Times or More.

API 名	機能	比
graphics.Camera	UI	33.7
graphics.Point	UI	5.2
graphics.drawable.InsetDrawable	UI	6.1
hardware.SensorEvent	調査	5.3
location.Location	調査	8.6
media.MediaScannerConnection	調査	10.3
net.SSLCertificateSocketFactory	OS	89.3
os.PersistableBundle	OS	6.0
text.SpannableString	UI	16.5
text.method.TextKeyListener	UI	5.5
text.style.CharacterStyle	UI	18.0
text.style.ClickableSpan	UI	7.7
text.style.ForegroudColorSpan	UI	8.0
text.style.URLSpan	UI	7.7
util.Range	他	17.9
util.Size	UI	98.5
util.SizeF	UI	12.9
view.animation.RotateAnimation	UI	9.2
widget.ActionMenuView	UI	8.7
widget.ArrayAdapter	UI	13.0
widget.HorizontalScrollView	UI	9.5
widget.ListView	UI	5.9
widget.Spinner	UI	15.7
widget.Toolbar.LayoutParams	UI	11.3

呼出頻度が 5 倍以上異なる API とその比を表 3 と表 4 に示す．なお，各 API について，端末やネットワークの内部調査に関わる API（調査），OS に関わる API（OS），UI に関わる API（UI），その他の 4 種類の用途に分類した．結果，UI などアプリケーションの体裁に関わる API が，正規アプリで多く使われていたのは 18 種，PUA で多く使われていた API は 7 種となり，正規アプリでは体裁を考慮した API が呼び出されていることが明らかになった．また，PUA より正規アプリに多く呼び出される API の中には，文字の表示形式に関わる API やリンク付テキストに関わる API を含んでいる点でも，UI の体裁を整える API が多用されていると言える．

また，ネットワークや端末の情報および個人情報の取得に用いる API が，PUA に比べて正規アプリで頻繁に呼び出されてた API が 3 種，正規アプリに比べて PUA で頻繁

表 4: PUA が正規アプリの 5 倍以上の頻度で呼び出す API
Table 4 APIs Called Freqently by PUAs 5 Times or More.

API 名	機能	比
accounts.AccountManager	OS	7.5
app.ActivityManager.MemoryInfo	調査	5.2
app.ActivityManager...	調査	5.2
content.ContentProviderClient	調査	25.9
database.MatricCursor	調査	6.1
graphics.RadialGradient	UI	2729.0
graphics.Shader	UI	6.2
media.SoundPool	他	8.7
net.wifi.WifiInfo	調査	6.0
net.wifi.WifiManager	調査	29.2
os.CountDownTimer	OS	7.2
os.ParcelFileDescription...	OS	9.4
os.Vibrator	OS	8.8
telephony.CellLocation	調査	21.4
telephony.CellSignalStrength	調査	32.0
telephony.ServiceState	調査	10.1
telephony.gsm.GsmCellLocation	調査	21.4
text.StaticLayout	UI	51.8
text.StaticLayout.Builder	UI	5.7
text.method.DialerKeyListner	UI	9.3
view.inputmethod.InputMethodInfo	調査	7.1
view.inputmethod.InputMethodSubtype	調査	17.3
widget.Gallery.LayoutParams	UI	8.1
widget.GridLayout.Spec	UI	10.9

に呼び出された API は 9 種となり，PUA は頻繁に端末の状態の調査活動を行っていることが明らかになった．

また，SSL 通信において，SSL の証明書およびホストの検証や SSL のハンドシェイクを扱う net.SSLCertificateSocketFactory が，PUA に比べて正規アプリで頻繁に呼び出されることが明らかになった．

5.3 最頻呼出 API 上位 50 種の比較

正規アプリにおいて使用頻度の平均が高かった API 上位 50 種と，PUA において使用頻度が高かった API 上位 50 種を比較したところ，42 種の API が共通であった．異なった 8 種の API を表 5 に示す．

正規アプリで頻繁に使われている API には UI や画面描画に関わる API が並ぶ中，PUA ではアプリケーション情報取得関連の API や SQL 関連の API が含まれている．PUA は正規アプリほどアプリの体裁に関わる API が利用されないことが明らかになった．

5.4 結果総括と考察

5.2 節と 5.3 節で示した，正規アプリと PUA による呼出傾向に特徴がある API について用途分類した結果を表 6 に示した．その結果，正規アプリでは UI 関連の API が，PUA では無線 LAN の状態やインストールされたアプリケーションなど Android 内部の調査を行う API が多く呼び出されていた．これは，個人情報の窃取を企てる PUA の目的を果たすための送信準備の様子や窃取活動による必要外のデータベースや無線 LAN の状態へのアクセスによるものであると考えられる．また，PUA は UI が簡素なものであり，必要外の情報窃取に重点を置いたアプリケーションであることが推察される．

また，正規アプリでは SSL 通信を実現するための，SSL の証明書およびホストの検証や SSL 通信のハンドシェイクを扱う API が呼び出されていた．正規アプリでは通信の真正性を保証した信頼性の高い通信を行っている様子が伺える．

UI の描画関連の API は PUA でも少なからず呼び出され，正規アプリよりも多く呼び出される API も存在したため，PUA 検知に用いる特徴としては不向きであると考える．一方，端末情報の調査活動に使われる API の過剰な呼出に注目することで，PUA の検知を実現できると考えられる．

表 5: 正規アプリおよび PUA 呼出頻度上位 50 種で相異なる API
Table 5 Different APIs Frequently Called 50 APIs between Proper Apps and PUAs

(a) 正規アプリの上位 50 API にのみ含む API
APIs included in Frequently Called 50 APIs of Proper Apps

API 名	機能
animation.KeyFrame	UI
graphics.Canvas	UI
graphics.ColorFilter	UI
graphics.Point	UI
graphics.PointF	UI
graphics.PorterDuffColorFilter	UI
widget.LinearLayout.LayoutParams	UI
widget.TextView	UI

(b) PUA の上位 50 API にのみ含む API
APIs included in Frequently Called 50 APIs of PUAs

API 名	機能
content.pm.PackageInfo	調査
database.sqlite.SQLiteStatement	調査
gesture.GesturePoint	UI
graphics.RadialGradient	UI
graphics.Shader	UI
net.Uri.Builder	OS
os.UserHandle	OS
text.SpannableStringBuilder	UI

表 6: 調査 2, 3 の分析で得られた API の用途分類
Table 6 Classification of API Usage of in Survey 2 and 3

	正規アプリ		PUA	
	全 API	TOP50	全 API	TOP50
UI	18	8	7	4
内部調査	3	0	12	2
OS	2	0	4	2
その他	1	0	1	0

6. 結論

本稿では，Android の正規アプリと PUA が使用する API の呼出頻度に着目し，呼出頻度に差がある API クラスと API を分析した．また，呼出頻度の高い API 上位 50 種を比較した．その結果，正規アプリでは UI などアプリケーションの体裁に関わるものや，SSL の証明書およびホストの検証や SSL 通信のハンドシェイクを扱う API が呼び出されていることが明らかになった．一方，PUA では，無線 LAN や端末の情報確認に関わる API が頻繁に呼び出されていることが明らかになった．

今後の課題として，サンプルとして用いるアプリケーション数が少ないため，数を増やして同様の結果となるか確認する必要がある．また，PUA の端末調査関連 API の呼出頻度が高いことに着目し，正規アプリと区別するアルゴリズムの検討を行う予定である．

参考文献

[1] 総務省. 数字で見たスマホの爆発的普及（5 年間の量的拡大）. http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc111110.html.
[2] International Data Corporation. Smartphone OS Marcket Share. https://www.idc.com/promo/smartphone-market-share/os.
[3] Information technology Promotion Agency. 2016 年度情報セキュリティの脅威に対する意識調査. http://www.ipa.go.jp/files/000056568.pdf.
[4] 畑田充弘, 森達哉. DNS クエリ分析に基づく Android PUA の識別と亜種分類. コンピュータセキュリティシンポジウム, pp. 1068–1075, 2017 年 9 月.
[5] 韓燦洙, 松本晋一, 川本淳平, 櫻井幸一. Android マルウェアの API 呼出し記録による分類及び個人情報漏えい検知. 情報処理学会九州支部 火の国シンポジウム, pp. 1–8, 2016 年 9 月.
[6] 細谷竜平, 角田裕太, 森達哉, 齋藤孝道. モバイルアプリケーションが取得しているプライバシー情報の調査. コンピュータセキュリティシンポジウム, pp. 553–560, 2017 年 9 月.
[7] 三村隆夫, 巻島和雄, 岩本一樹. ソーシャルネットワークで共有される android アプリケーションの実態調査. コンピュータセキュリティシンポジウム, pp. 113–120, 2018 年 9 月.
[8] Katsutaka Ito, Hirokazu Hasegawa, Yukiko Yamaguchi, and Hajime Shimada. Detecting Privacy Information Abuse by Android Apps from API Call Logs. *International Workshop on Security, IWSEC 2018*, September 2018.
[9] Google Developers. Google Play のインストールの統計情報. https://developer.android.com/about/dashboards/.
[10] Google Developers. Package Index. https://developer.android.com/reference/packages.