平成 28 年度 修士論文

世界各国の Android サードパーティー マーケットのセキュリティに関する調査

Understanding the Security Management of Global Third-Party Android Marketplaces

指導教員　森 達哉 准教授

早稲田大学基幹理工学研究科 情報理工・情報通信専攻

学籍番号 5115F007-8

石井 悠太

平成 29 年 1 月 30 日

概要

Android アプリを配布するマーケットには，Google Play 公式マーケットの他にも数多くのサードパーティーマーケットが存在する．サードパーティーマーケットの運営母体は多岐にわたり，それぞれの運用形態も様々である．したがって，それらサードパーティーマーケットの安全性や健全性は明らかではない．このような背景の下，本研究は全世界に偏在するサードパーティーマーケットの安全性や健全性の実態調査を行う．調査のため合計 27 の Android アプリマーケットから約 470 万のアプリデータを収集し，各マーケットの安全性と健全性の評価を行った．また，各マーケットに対して独自に定義したセキュリティインデックスを算出し，安全性と影響度の関係を分析した．その結果，これまであまり研究対象とされていなかったマーケットでは過去にウイルススキャンにかけられた形跡がない検体の割合が高かったこと，中国系マーケットのように特定のマーケット間でのマルチリリースが活発であること，ユーザに与える影響範囲が大きいにもかかわらず安全性の低いマーケットが存在すること等を明らかにした．

目次

図目次

表目次

第 1 章　序論

Android は世界で広く利用されているモバイル端末向けのオープンソース OS である．スマートフォンやタブレットをはじめとする多くのデバイスが Android を搭載している．全世界における Android スマートフォンの出荷台数は 2015 年には 1 億 4000 万台を突破し [7]，公式マーケットである Google Play 上のアプリ数は 2017 年 1 月時点で 260 万を突破した．

一方で，Android は攻撃者からのターゲットとされやすい傾向にある．Android 悪性アプリの挙動は様々であり，ユーザの情報を秘密裏に窃取するマルウェアや，感染させたユーザに対して金銭を要求するランサムウェア [21] 等が存在する．モバイル OS 向けマルウェアのターゲットのうち約 97% が Android 向けであるという報告もある [14]．

悪性アプリが Android に多い理由の一つに，アプリマーケットのオープン性があげられる．Android アプリマーケットには，Google Play 公式マーケットのほかに，Google の関与しないサードパーティーマーケットが多数存在する [24]．ユーザがサードパーティーマーケットを利用するには，端末の「提供元不明のアプリ」のインストールを許可する設定にする必要がある．この設定が許可されていれば，マーケット専用のクライアントアプリを経由するか，Web サイトから直接 APK ファイルをダウンロードすることによって，目的のアプリをインストールすることができる．

サードパーティーマーケットを使用することで，ユーザは公式マーケットでは提供されていないアプリを使用することができたり，ときには本来有料のアプリを無料で入手したりすることができる．また中国やキューバのように，地域によっては Google Play を自由に利用できない国も存在する [23]．そのような国でユーザが任意の Android アプリを利用したい場合は，サードパーティーマーケットに頼るほかない．実際，中国国内で出荷されるスマートフォンには，サードパーティーマーケットの専用クライアントアプリがプリインストールされていることも多い．

しかしながら，こうしたサードパーティーマーケットの安全性をユーザが判断することは難しい．サードパーティーマーケットの運営母体には，著名なインターネットサービスを提供する企業によるもの，スマートフォン等のハードウェアベンダーによるもの等，多様な組織が存在する [5]．マーケットによっては，運営母体についての正確な情報が記載されていないものも少なくない．当然ながら各サードパーティーマーケットの品質や運営ポリシーは多岐に渡る．例えば，Google Play では Bouncer[31] と呼ばれるマルウェア対策機能が存在することが知られ

ているが，各サードパーティーマーケットがこのような対策を実施しているとは限らない．したがって，サードパーティーマーケットの安全性を計測することは容易ではない．

このような背景の下，本研究はサードパーティーマーケットの安全性や健全性の実態調査を行う．これまでにも Android セキュリティの研究では，アプリ解析手法を評価するためにサードパーティーマーケットのデータが利用されてきた．しかし，マーケットそのものを評価の対象とした研究はほとんど例がない．

我々は調査のために，はじめに合計 27 の Android アプリマーケットから約 470 万のアプリデータを収集した．そしてマーケットごとに安全性や健全性の評価を行った．また，各マーケットに対して独自に定義したセキュリティインデックスを算出し，安全性とマーケット影響度の関係を分析した．

その結果，これまであまり研究対象とされていなかったマーケットでは過去にウイルススキャンにかけられた形跡がない検体の割合が高かったこと，中国系マーケットのように特定のマーケット間でのマルチリリースが活発であること，無視できない数の有料アプリがマルチリリースされていること，ユーザに与える影響範囲が大きいにもかかわらず安全性の低いマーケットが存在すること等を明らかにした．

本論文の残りの構成は以下のようになっている．第 2 章では使用するデータセットの説明を行う．第 3 章にて VirusTotal による検体の分析結果を示し，第 4 章で複数マーケットにリリースされたアプリの分析を行う．第 5 章ではマーケットの品質を決定する要因について議論する．第 7 章で関連研究を紹介し，第 6 章では本研究の制限事項と今後の展望を述べ，最後に第 8 章で本研究のまとめとする．

第 2 章　データセット

本章では，使用したデータセットについての説明を行う．幅広いマーケットを含めた調査を行うため，初めに世界各国の Android アプリマーケット状況について事前調査を行った．それをもとに選定した 13 のマーケットについては，独自にクローラを開発することでアプリを収集した．また，Androzoo データセットに含まれる 14 マーケットのデータも使用した．全体でユニークなデータ数は 4,761,283 検体である．

2.1　事前調査

研究対象としてサードパーティーマーケットを幅広く分析するためには，様々な言語圏のマーケットに対して調査を行う必要がある．そこで事前調査として Lancers[19] 等のサービスを利用することで，各言語 (中国語，イタリア語，ドイツ語，スペイン語，トルコ語，ロシア語，アラビア語，ベトナム語，タイ語，ペルシア語等) に精通した人々を雇い，各国の Android マーケット事情の調査を依頼した．その結果，合計 15 ヶ国語圏の国々について調査結果を集めることができた．調査結果をもとに各国で有名なサードパーティーマーケットのリストアップを行い，その中から利用者数が多く，かつアプリの自動クロールが可能なマーケットを選定した．

2.2　独自に収集したマーケット

対象として選定したマーケットは表 2.1 に挙げる上から 13 のマーケットである．本論文では便宜上この表に示した名称を用いるが，マーケットの正式名称とは異なる場合がある．収集は 2016/6〜2016/9 の期間に断続的に行った．マーケットはサイト形式のものやクライアントアプリ形式のものまで様々である．マーケットによって内部の API 仕様は異なるため，一つ一つについてツールを用いてリバースエンジニアリングをすることでクローラを開発した．リバースエンジニアリングの方法として，エミュレータや実機を用いてパケットをキャプチャしたり，`apktool`[6] や `smali`[22]，`jadx`[17] 等のツールを活用したりした．アプリの名称や説明文，ダウンロード数のようなメタデータは取得が可能なものについては取得し，パースし

た．クロールの基本方針として，総合ランキングや各カテゴリのランキング上位から順に取得していった．一部マーケットに対してはその上でさらにランダムクエリ検索を行った．第 6 章でも述べるが，各マーケットに存在するアプリを大部分収集できたかどうかは定かではない．これらの 13 マーケットから，合計で約 80 万検体のアプリを収集することができた．この数にはマーケットをまたいで存在する検体の重複も含まれている．

2.3 Androzoo データセット

その他のマーケットのデータには Androzoo データセット [28] を利用した．Androzoo では Google Play や *Genome* マルウェアデータ，torrent から取得したデータ等を含む 14 マーケットの Android アプリ検体を提供している．本研究では 2016/7/10 時点のリストを使用した．この時点での検体数は約 430 万検体であった．ただしこのデータセットにはマーケット上のメタデータは存在しない．Androzoo の検体は著者らによって VirusTotal にスキャンされており，それらの検知数はデータとして提供されている．Androzoo データセットに含まれる各マーケットの詳細については，文献 [28] も参照されたい．

2.4 検体の内訳

独自にクロールしたものと Androzoo データセットの検体を合計した容量は約 45TB となった．検体数の内訳は表 2.1 に示したとおりである．各検体について MD5 ハッシュ値を計算し，全体でユニークな検体数を調べたところ，4,761,283 検体となった．また，ユニークなパッケージ名の数を調べたところ，2,827,578 種類となった．

表 2.1　データセットのマーケット一覧

マーケット	検体数	言語	運営母体	備考
Alandroid[3]	9,620	アラビア語	不明	中東向けのマーケット.
Appvn[18]	34,415	ベトナム語	不明	モバイルマルチプラットフォーム向けにアプリを提供.
Aptoide[8]	138,421	多言語	ポルトガルの企業	ユーザが独自にマーケットを作成可能.
Baidu[27]	13,020	中国語	大手ポータルサイト Baidu	中国国内では Google Play は利用不可.
Blackmart[9]	100,127	英語	不明	専用クライアントアプリを利用してアプリをダウンロードする.
Cafebazaar[11]	54,034	ペルシア語・英語	イランの IT 企業	イラン向けマーケット．有料アプリの課金体制あり．イラン国内では Google Play は無料アプリのみ利用可能.
Entumovil[13]	235	ポルトガル語	キューバの企業	キューバ向けマーケット．キューバ国内では Google Play は無料アプリのみ利用可能.
Getjar[16]	38,180	英語	リトアニアの企業	–
Mobogenie[20]	31,547	多言語	アメリカの企業	専用クライアントアプリや PC 向けクライアントソフトが提供されている.
Mobomarket[12]	10,392	インドネシア/タイ/英語	Baidu 傘下の企業	専用クライアントアプリや PC 向けクライアントソフトが提供されている.
Uptodown[10]	59,428	スペイン語/多言語	スペインの企業	専用クライアントアプリが提供されている．Android の他，マルチプラットフォーム向けのアプリマーケット.
Yandex[26]	22,964	ロシア語	ポータルサイト Yandex	専用クライアントアプリを利用してアプリをダウンロードする.
Zhushou360[1]	204,417	中国語	セキュリティベンダー 360	中国国内では Google Play は利用不可.
Play.google.com	3,608,379	多言語	Google	公式マーケット.
Anzhi	736,517	中国語	中国の企業	中国国内では Google Play は利用不可.
Appchina	593,128	中国語	中国の企業	中国国内では Google Play は利用不可.
Mi.com	104,029	中国語	中国大手スマートフォンベンダ Xiaomi	中国国内では Google Play は利用不可.
1mobile	57,525	多言語	不明	–
Angeeks	55,815	中国語	中国の企業	中国国内では Google Play は利用不可.
Slideme	52,448	英語/フランス語	アメリカの企業	–
Torrents	5,294	–	–	BitTorrent を利用して収集されたデータ.
Freewarelovers	4,145	英語	ドイツの企業	–
Proandroid	3,683	ロシア語	不明	–
Hiapk	2,510	中国語	Baidu 傘下の企業	中国国内では Google Play は利用不可.
Fdroid	2,023	英語	イングランドの企業	オープンソース・ソフトウェアのみを集めたマーケット.
Genome	1,247	–	–	Zhou ら [38] によって収集されたマルウェアデータセット.
Apk_bang	363	不明	不明	すでに閉鎖.

第 3 章　アプリのセキュリティ分析

本章では，収集した検体について VirusTotal のレポートを取得し，マーケットごとに集計した分析を行う．その結果，これまであまり研究者の目にさらされていなかったマーケットが存在することや，そのことが必ずしもマーケットの品質に関係するわけではないということがわかった．また，中国系マーケットのマルウェア率が全体的に高いことも確認できた．

3.1　Dragnet 調査

マーケットに含まれるアプリが過去にセキュリティ調査の対象となったかを判定する分析を，dragnet (捜査網) 調査と呼ぶこととする．この分析は，より多くの調査対象となっているアプリほど安全であるべきだという直感に基づくものである．本研究ではこの dragnet の指標として，アプリが過去にオンラインスキャンサービスによってスキャンされているかどうかを調査した．昨今，それらのスキャンサービスは研究者のみならず一般のユーザによっても Web インタフェースを通じて利用されている．さらにスキャン結果は一般に公開されていることが多いため，検索エンジンを通じてアクセスでき，アップロードされた検体はいくつかの組織間で共有されている．オンラインスキャンサービスとして，我々は VirusTotal[25] (以下 VT) を利用した．VT は複数のアンチウイルスベンダーによるスキャンを行うことのできるサービスである．我々は独自にクロールした検体のハッシュ値を VT で検索し，スキャンレポートを取得した．各マーケットについて，この時点でまだ VT によるスキャンが行われていなかった検体の割合を調べた．Androzoo 由来の検体は基本的にすべてスキャン済みであったため，ここでは分析の対象外としている．結果を表 3.1 に示す．

他のマーケットと比較して，*Cafebazaar* の未スキャン率が高いことがわかる．このマーケットについては第 5 章でも述べる．逆に，Google Play と重複率の高いマーケット (第 4 章で後述) は，当然ながらスキャン済みの割合が高い結果となった．

表 3.1　各マーケットの VirusTotal 未スキャン率

マーケット	未スキャン率 (%)
Cafebazaar	75.6
Yandex	20.6
Mobomarket	20.5
Baidu	19
Getjar	15.5
Appvn	14.1
Zhushou360	13.1
Alandroid	6.9
Aptoide	4.6
Mobogenie	2.2
Entumovil	1.7
Blackmart	1.7
Uptodown	1.2

3.2　良性/悪性アプリの内訳

図 3.1 に，VT におけるマーケットごとの悪性アプリ検知率を示す．前節で未スキャンであった検体については，我々が新たにサブミットすることでレポートを取得した．また，検体の容量超過が原因でスキャンできなかった数検体は除外している．ここでは検知数が 10 以上のもののうち，検知ラベルにアドウェアが一つでも当てはまった場合はアドウェア，そうでないものをマルウェアとした．簡単のため，検知ラベルに「adware」もしくは「addisplay」という文字列が含まれていた場合をアドウェアの検知条件とした．ただし Androzoo は検知ラベルではなく検知数のみしか提供していないので，このデータからのみではマルウェアとアドウェアの両者の区別はできない．そこで，便宜上図 3.1 ではすべてマルウェアとして扱っている．改めて我々がスキャンレポートを取得すれば可能であるが，処理コストの都合により今回は割愛した．

最上部の *Genome* はマルウェアデータセットのため，マルウェア率が 100% となっているのはもっともである．また全体的に，*Appchina* や *Anzhi*，*Baidu* 等，中国系マーケットの悪性率が高いことがわかる．Google Play は良性率の上位 8 位となった．また，未スキャン率の最も高かった *Cafebazaar* が，最も悪性率の低い結果となったことは興味深い．このことから，研究者コミュニティからあまり注目されてこなかったマーケットが，必ずしも悪性率が高いわけではないということが判明した．*Fdroid* のようなオープンソースコミュニティのマーケットも良性率が高いことが確認できる．以降の分析では特に断りのない限り，検知数 10 以上のものをアド

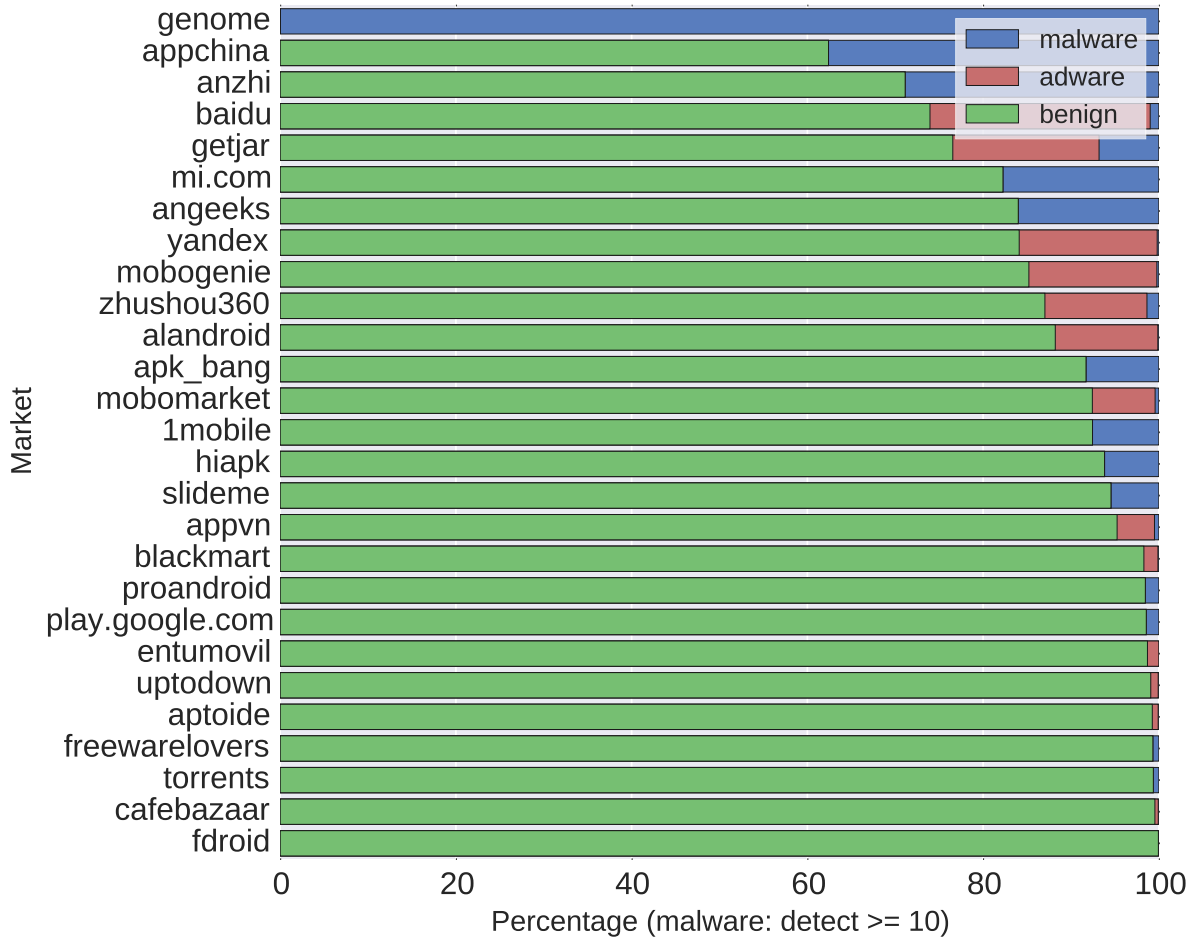

図 3.1　マーケットごとの悪性アプリ検知率の内訳

ウェアとの区別を行わずにすべてマルウェアとして扱う.

第 4 章　マーケットにおけるアプリの流通

同一のアプリが複数のマーケットにリリースされている場合がある．本研究ではこのようなアプリのことをマルチリリースアプリと呼ぶ．本章では，パッケージ名一致と MD5 ハッシュ値一致の 2 つの観点からマルチリリースアプリの特徴や関係性を調べた．その結果，全体の約 10〜20% がマルチリリースされていること，マルチリリースされるマーケットの間には特徴的な関係性がみられること，有料アプリのマルチリリースが無視できない数存在すること等を明らかにした．

4.1　マルチリリースアプリ

同一のアプリが複数のマーケットにリリースされているアプリのことを，本研究ではマルチリリースアプリと呼ぶ．マルチリリースのパターンとして，パッケージ名が一致する場合と MD5 ハッシュ値が一致する場合の 2 つを想定する．パッケージ名はアプリの内部識別子であり，Android 端末内ではパッケージ名の重複は許されない．同一パッケージ名かつ同一署名の場合，アップデートとしてインストールされることになる．したがって，パッケージ名一致でみたときのマルチリリースの場合，同じ作者によるバージョン違いも同一ものとしてまとめることになる．ただし，偶然パッケージ名が重複する場合や，パッケージ名の変更なしにリパッケージングされた場合も含まれてしまうことに留意する必要がある．

またパッケージ名・MD5 いずれの場合も，マルチリリースがその開発者本人によるものなのかどうかまでは判断できない．目的は様々であろうが，あるマーケットで配布されていたアプリを第三者がそのまま別マーケットに流通させることも可能である．パッケージ名の異なるリパッケージングの可能性はここでは考慮しない．また，マーケット内に同名パッケージのアップロードが可能かどうかは，そのマーケットによって異なっている．例えば Google Play ではパッケージ名はユニーク識別子だが，全マーケットでそのような仕様とは限らない．こうした前提のもと，以降では具体的にマルチリリースの実態を探っていく．

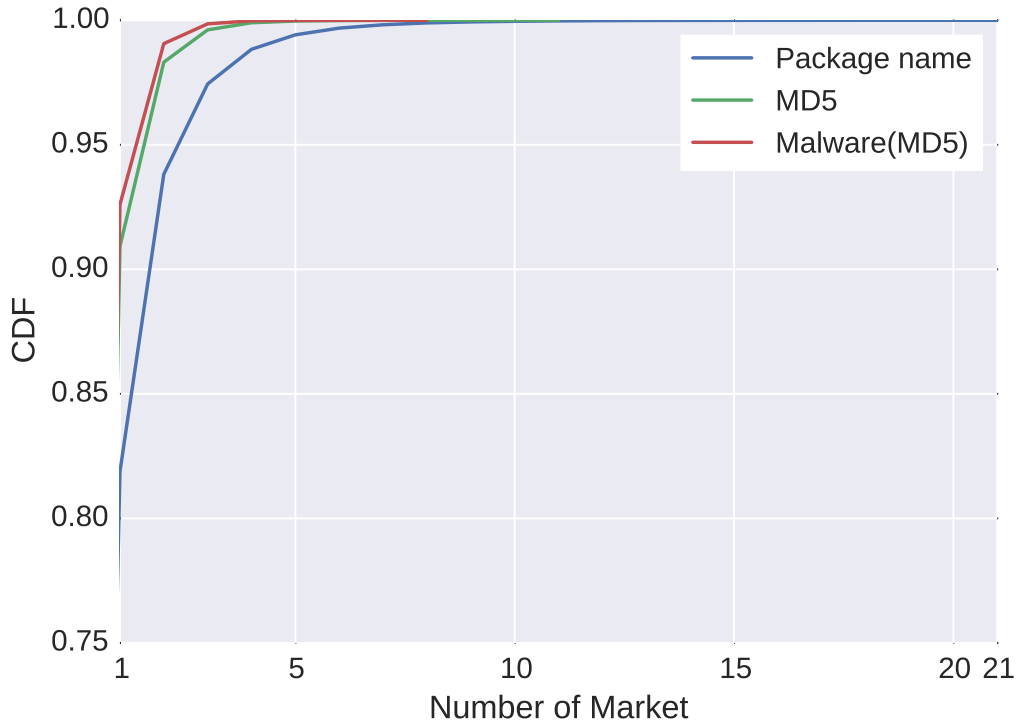

図 4.1　マルチリリースされているマーケット数の CDF

4.2　マルチリリースマーケット数

図 4.1 にアプリごとにマルチリリースされているマーケット数の CDF を示す．パッケージ名よりも MD5 のほうがより厳密な一致判定であることが図からも確認できる．後述するマルウェアのマルチリリースも同図に描いている．これを見ると，全体の約 1〜2 割のアプリが少なくとも 2 つ以上のマーケットにリリースされていることがわかる．すなわち，MD5 一致の条件では 476 万*10% = 47 万，パッケージ名一致の条件では 282 万*20% = 56 万 がマルチリリースであると言うことができる．パッケージ名で見たときは最大で 21 のマーケットにマルチリリースがされており，具体的には com.estrongs.android.pop (ES File Explorer File Manager) であった．次点で 19 リリースの com.google.zxing.client.android (Barcode Scanner)，com.shazam.android (Shazam)，xcxin.filexpert (File Expert) が並んだ．また，マルウェアがマルチリリースされている割合のほうが，そうでないものよりも僅かに低い．

4.3　マーケット間の共起関係

あるマーケットに含まれるアプリが別のマーケットに含まれる割合を示したのが図 4.2 である．ここではパッケージ名を一致の条件とした．縦軸のマーケットを M_a，横軸のマーケットを M_b とし，それぞれのマーケットに含まれるアプリの集合を $App(M_a)$，$App(M_b)$ のように

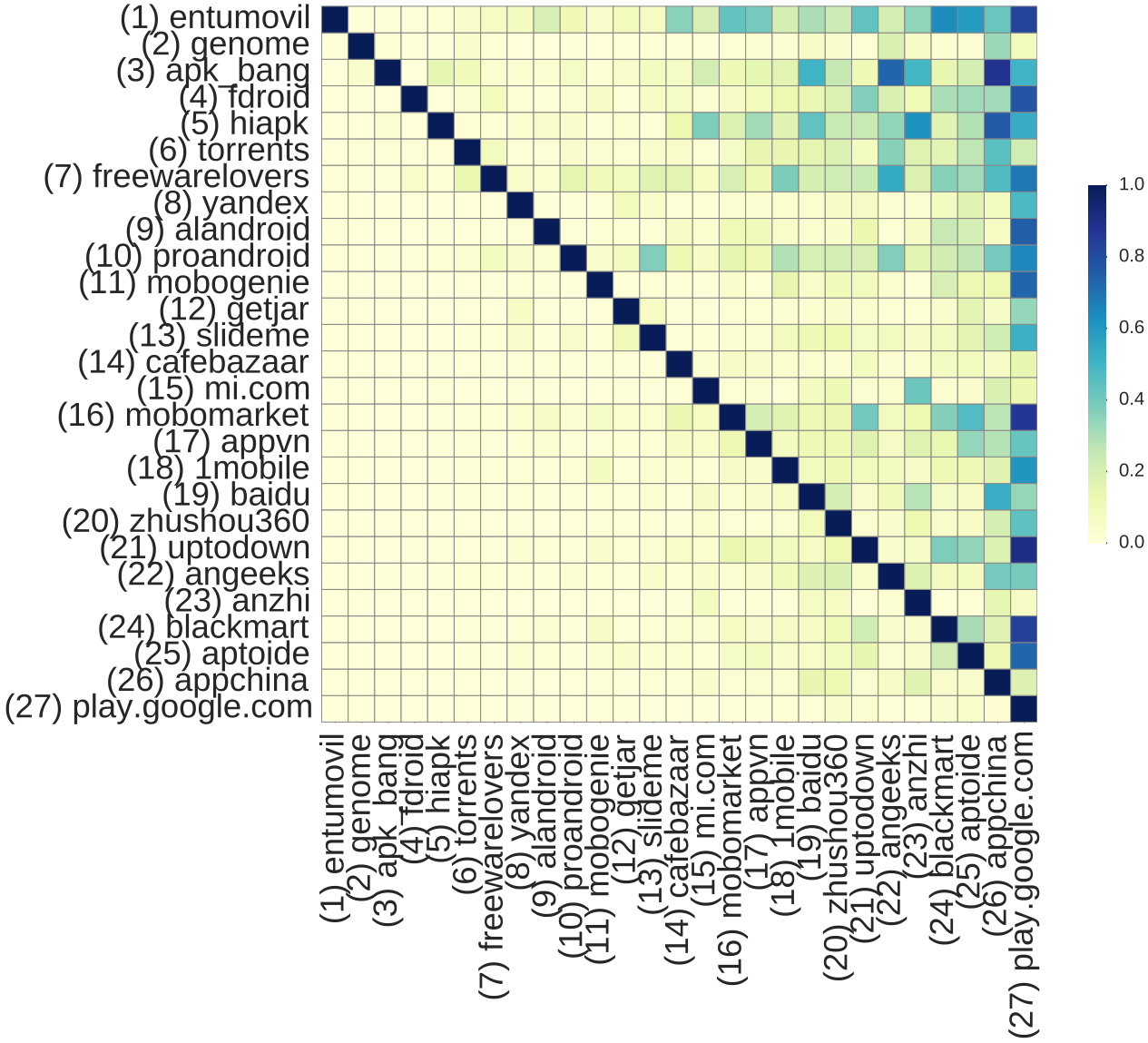

図 4.2　パッケージ名で見たマルチリリースのヒートマップ

表すと，

$$\frac{\mid App(M_a) \cap App(M_b) \mid}{\mid App(M_a) \mid} \tag{4.1}$$

の値を色の濃さで表している．この値が大きいほど，マーケット M_a に含まれるアプリはマーケット M_b にも存在する割合が高い．

$b = 27$ に注目すると，Google Play と共通のアプリをもつマーケットが多いことが見て取れる．特別な事情のない限り，アプリ開発者としても最も利用者の多いであろう公式マーケットにまずリリースすることは理にかなっている．一方でマルウェアデータセットの *Genome* やイランマーケットの *Cafebazaar*，中国マーケットの *Mi.com*，*Anzhi* 等 ($a = 2, 14, 15, 23$) を見ると，Google Play と共通アプリをあまり持たない傾向にある．*Genome* は既知のマルウェアで構成された有名なデータセットであるため，ここに含まれるアプリは多くのマーケットで削除さ

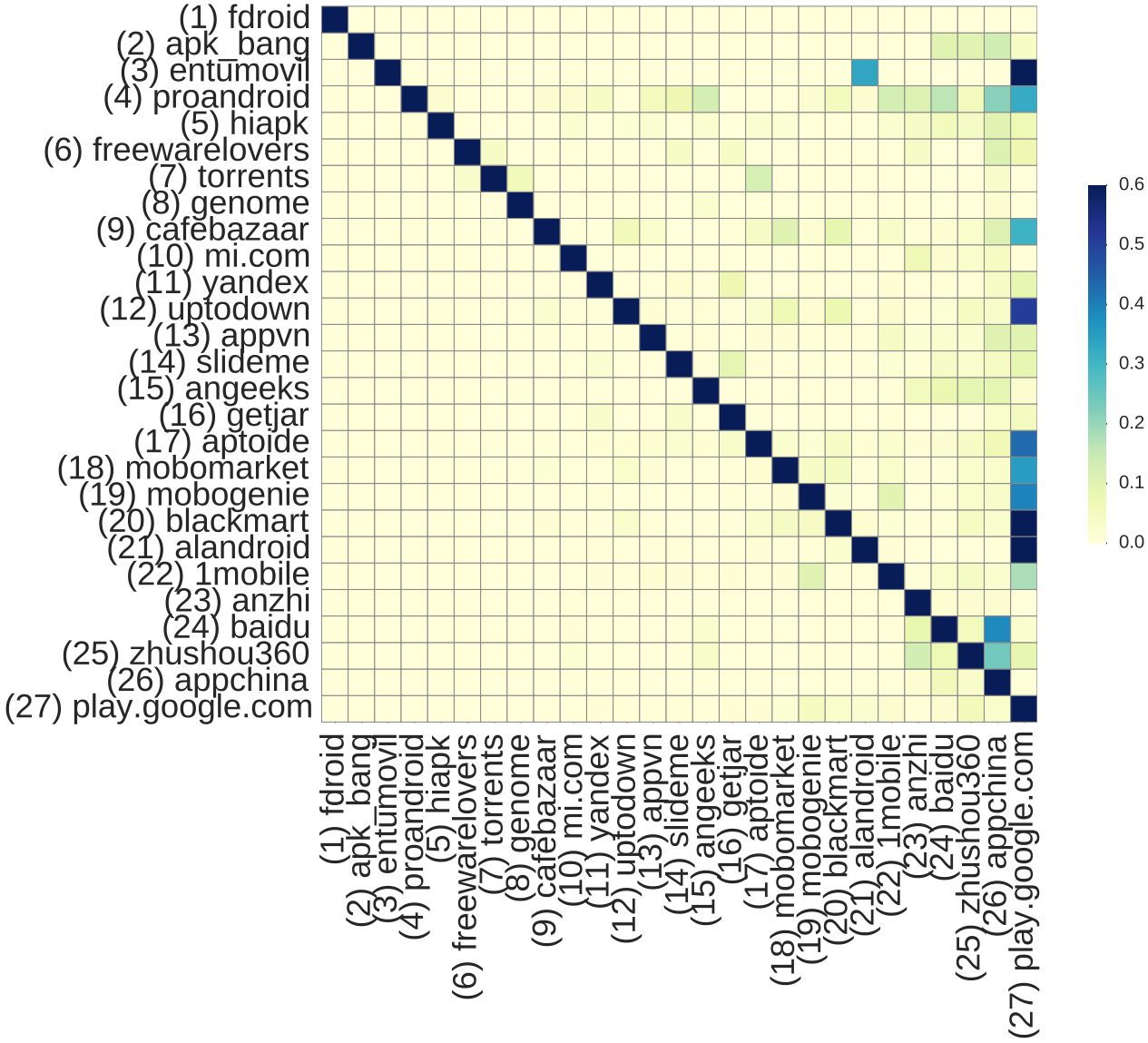

図 4.3　マルウェアのマルチリリースのヒートマップ

れているからであると推測する．中国に関しては Google Play が利用できないことに加え，イランでも第 5 章で述べるような独自性が確認できる．また，*Mi.com* や *Anzhi*，*Hiapk*，*Appchina*，*Zhushou360* のような中国マーケット同士も色の濃い組み合わせが多く，これらの間でのマルチリリースが行われていることが伺える．

4.4　マルウェアのマルチリリース

図 4.2 と同様の方法で，マルウェアに絞って分析したものが図 4.3 である．こちらはマルウェア検体そのものの流通を見る必要があるので，パッケージ名ではなく MD5 一致の条件ででみている．視認性の向上のため，ヒートマップの色の上限値を 0.6 とした．

$a = b = 23, 24, 25, 26$ の部分に注目すると，中国系マーケットの塊がみえる．これらはもともと互いにマルチリリース数が高い傾向にあったが，マルウェアに関しても同様のことが言える．また，Google Play と共通のマルウェアをもつ割合高いマーケットも多い．アドウェアとの区別を行っていないことには留意する必要があるが，少なくともクロールを行った時点では公式マーケットでもこれらが流通していたということができる．

4.5 マルチリリースの過程

4.5.1 リリース順序

あるアプリがどのような過程で各マーケットにマルチリリースされたのかがわかれば，サードパーティーマーケット全体のエコシステムの把握に役立てることができる．本項ではその一つの手掛かりとして，アプリがマーケットにリリースされる順序に着目した．アプリがリリースされた日を比較するには，マーケットのメタ情報を利用することができる．我々が本研究で収集したデータセットに含まれるメタデータのうち，リリース日が記載されているマーケットは *Alandroid*, *Appvn*, *Aptoide*, *Blackmart*, *Mobogenie*, *Uptodown*, *Zhushou360* の 7 つであった．また Androzoo データセットにはメタデータが含まれていないため，これに該当するマーケットもまた本分析の対象外としている．Google Play もその一つである．分析対象の 7 つのマーケットに該当する検体数は 51,410 であった．

リリース順序の分析では同一バージョンのアプリ同士で比較する必要があるため，パッケージ名ではなく MD5 一致のマルチリリースを対象とした．また，すでに述べたように正規の開発者によってマルチリリースされたかどうかまでは判断できないことにも留意する必要がある．当然ながら，リリースされるマーケットの順番はアプリによって異なる．そこで，アルゴリズム 1 に示す方法で順位を分析した．今回の場合 $market\_num$ の値は 7 となる．例としてアプリ x がマーケット *Uptodown* と *Mobogenie* でマルチリリースされている状況を考える．*Mobogenie* よりも *Uptodown* のリリース日のほうが早ければ，$\mathbf{M}(x)$ は，[*Uptodown*, *Mobogenie*] のようなリストとなる．分析の結果，*Aptoide*, *Blackmart*, *Uptodown*, *Zhushou360*, *Mobogenie*, *Appvn*, *Alandroid* の順となった．しかしながら，本アルゴリズムは一つの傾向を示したにすぎない．例えばマーケットごとにデータセットの検体数は異なるため，比較の際に偏りが生じている可能性がある．リリース順の正確な実態を把握するためには，より詳細な分析を行う必要があると考える．

4.5.2 リリース日の近いマーケット

アプリによっては，あるマーケット A でリリースされた同日に別のマーケット B にリリースされることもあれば，長い月日が経った後にマーケット C にリリースされるという事例がある．マーケット同士のリリース日が互いに近い場合，開発者自身が同時に複数マーケットにリ

Algorithm 1: リリース順の分析アルゴリズム

```
market_num;                                          /* 対象マーケットの数 */
A;                                                   /* 分析対象の全アプリの集合 */
M(x);                                                /* アプリ x がリリースされたマーケットの順番 */
flow ← [];                                           /* 出力されるマーケット順位のリスト */
while sizeOf(flow) < market_num do
    counts ← dictionary();
    for ∀a ∈ A do
        add 1 into counts[M(a)[0]];                  /* アプリ a が最も早期にリリースされたマーケットを加算 */
    earliest ← getMaxValueOf(counts);                /* その順位における最頻出のマーケットを取得する */
    append earliest to flow;
    for ∀a ∈ A do
        if M(a)[0] == earliest then
            delete M(a)[0];
```

表 4.1　リリース日の近いマーケット

マーケットのペア	検体数
Blackmart & *Aptoide*	1898
Appvn & *Aptoide*	1665
Uptodown & *Aptoide*	465
Uptodown & *Blackmart*	162
Uptodown & *Appvn*	158
Blackmart & *Appvn*	124
Zhushou360 & *Mobogenie*	52

リースを行っているという可能性がある．あるいはマーケットが別のマーケットを監視し，新着アプリを発見した場合は自動でコピーしているという可能性もある．こうした傾向を捉えるため，各マーケットのペアについて，同じアプリが 2 日以内に両マーケットにリリースされた数を調査した．上位 6 つのペアについて結果を表 4.1 に示す．最も頻度の高かった *Blackmart* と *Aptoide* が，前節のリリース順における上位 2 つのマーケットと一致したことは興味深い．ただし本分析も前項と同様の手法の制限があるため，詳細な分析は今後の課題とする．

表 4.2　有料アプリの一致検体数

一致方法	一致数	一致率 (%)
MD5	1,907	12.8
パッケージ名	4,251	28.5

表 4.3　有料アプリと価格・ダウンロード数との関係

	価格 (円)	ダウンロード数
全有料アプリ	334.6	10,553.7
パッケージ名一致	335.7	33,761.3
MD5 一致	293.3	27,973.0

4.6　有料アプリのマルチリリース

4.6.1　有料アプリ

今回収集したデータの他に，我々が過去に独自にクロールを行なった Google Play 有料アプリが 14,906 検体存在する．この有料アプリと，本研究のために収集したデータセットに含まれる検体との一致率を調べた結果を表 4.2 に示す．ここで，今回収集したデータセットはすべて無料でクロール可能な検体のみで構成されていたことに注意されたい．つまりパッケージ名でみると，データに含まれる約 3 割もの有料アプリと同一のアプリが，マルチリリース先の各マーケットで無料で入手できるといえる．ただし期間限定で無料にしているものや，開発者の方針でサードパーティーマーケットでは無料で配布しているものが含まれる可能性には留意する必要がある．次に，有料マルチリリース先に多いマーケットを調べた．その結果，上位 4 つは *Appvn*, *Appchina*, *Aptoide*, *Anzhi* となった．しかしながら，これらのマーケットに有料マルチリリースが数多く存在した正確な理由はわからない．

4.6.2　有料マルチリリースアプリの特徴

マルチリリースされた有料アプリについて，Google Play 上の価格やダウンロード数との関係を表 4.3 に示す．各値は平均値である．これを見ると，マルチリリースされやすいアプリと価格との間に相関はみられない．ダウンロード数に関しては，マルチリリースされたアプリの値はマルチリリースされていないものに比べて約 3 倍となっている．このことから，ランキング上位のアプリがマルチリリースされやすい傾向にあると考える．

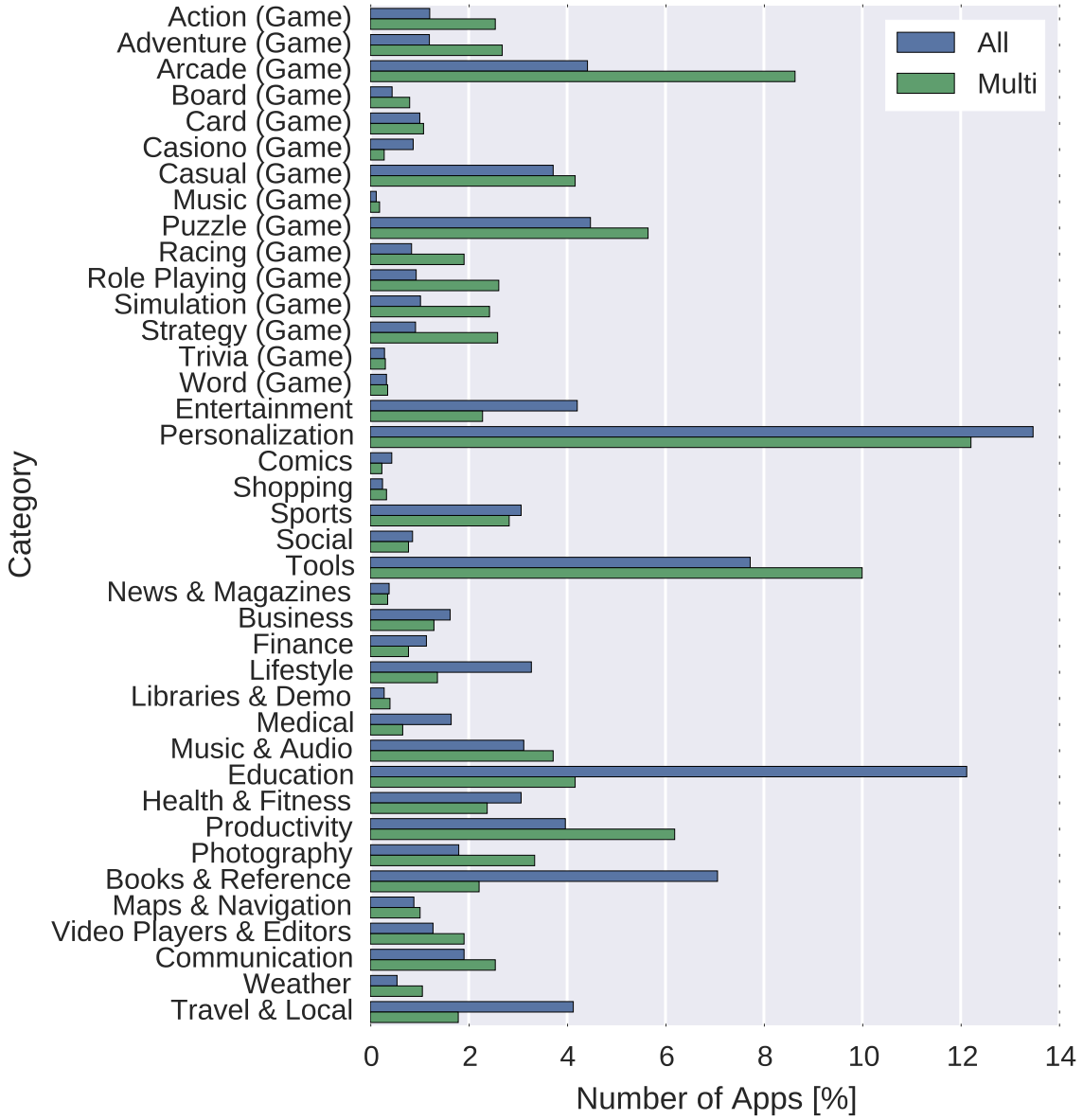

図 4.4　マルチリリースされた有料アプリのカテゴリ内訳

図 4.4 に，マルチリリースされた有料アプリの，マーケットにおけるカテゴリ内訳を示す．有料アプリ全体の内訳と比較を行う．この図から，アプリ全体で大部分を占めるカテゴリは，マルチリリースの中においても大部分を占める傾向にあると言える．また，ゲーム系カテゴリのマルチリリース率が全体的に高い．Tool や Productivity カテゴリもマルチリリースされやすいことがわかる．

4.6.3　LVL 実装

マルチリリースされた有料アプリについて，License Verification Library (LVL)[2] が実装されているかどうかを調査した．LVL とは，正規のユーザが購入したかどうかをチェックするライセンス機構である．LVL が正しく実装されたアプリは，そのユーザの購入状況を，起動時等にラ

表 4.4　有料アプリと LVL 実装率との関係

	実装数	検体数	割合 (%)
非マルチリリース	1,803	10,655	17
マルチリリース	1,739	4,251	41

イセンスサーバに問い合わせる．正規購入したユーザだとわかればそのまま正しく動作し，そうでない場合はアプリを終了する等の動作が可能である．LVL の動作には CHECK_LICENSE パーミッションが必要である．

そこで，パッケージ名でみたときのマルチリリースされた有料アプリについて，このパーミッションの有無を調査した．その結果を表 4.4 に示す．この結果からわかるように，マルチリリースされている有料アプリのほうが，そうでない有料アプリよりも LVL 実装率は高い．つまり LVL を実装していることが，マルチリリース自体への抑止力にはなっていないといえる．また，マルチリリースされているアプリのうち，LVL の実装されていない 59% はそのまま起動できてしまう可能性がある．パーミッションがあったとしても LVL を正しく実装できているとは限らないため，実際には不正利用されうる割合はより高いと考える．ただし，CHECK_LICENSE パーミッションは APK 拡張ファイルの実装にも使用されることがあるため，必ずしも LVL 実装がされているとは限らないことには留意する必要がある．

LVL 実装の仕方次第では，知識のある技術者によって除去されるおそれがある．そこで，Google Play 上で LVL 実装のあるパッケージ名一致の有料マルチリリースアプリについて，他マーケットで LVL 実装が外されているかどうかを調査した．その結果，1,739 中 153 の検体について，同名のパッケージ名を持ちつつも CHECK_LICENSE パーミッションの外れている検体が存在した．検体数こそ多くはないが，ライセンスチェック機構を除去して意図せずマルチリリースされることは，開発者にとっては直接の不利益を被ることになり，脅威と言える．また，LVL 実装自体を削除して CHECK_LICENSE パーミッションはそのまま残している可能性もあるため，実際の検体数はより多いものと予想する．

4.7　マルチリリースアプリの証明書

パッケージ名一致のマルチリリースアプリそれぞれについて，APK に含まれる開発者証明書を調査した．openssl コマンドで抜き出した文字列情報の比較を行っている．証明書が同一の場合，同一の開発者によって作られたアプリである可能性が高い．その開発者自身が複数マーケットに投稿したかどうかまでは定かではないが，少なくとも第三者による改変が行われていないことは保証される．逆に同一パッケージ名にもかかわらず証明書が異なる場合，何らかの改変が行われている可能性がある．

調査の結果，証明書の抽出に成功したパッケージ名一致のマルチリリースアプリ 512,577

種類のうち，478,653 種類はすべて同一の証明書を有していた．しかし残りの 33,924 種類のアプリに関しては，少なくとも 2 種類以上の証明書を持っていた．マーケットによって証明書を変更する開発者も存在するだろうが，そうした例は多くはないと考える．上記はパッケージ名一致という条件で抽出したものであるため，パッケージ名を改変して不正にリパッケージングされたアプリの数はさらに増えると予想される．com.dropbox.android (Dropbox) や com.estrongs.android.taskmanager (ES Task Manager)，com.evernote (Evernote)，wp.wattpad (Wattpad Free Books) のように，マルチリリースされているマーケット数が多い (いずれも 17 マーケット以上) にもかかわらず証明書の種類が一つのものもあった．これは開発者がアプリの改ざんに対して厳格な対処を行っているためと考えられる．開発者自身が各マーケットを監視し，アプリが不正に改ざんされて配布されていないかどうかをチェックする体制が整っている可能性がある．また，Google Play にリリースされている検体の証明書よりも，異なる証明書で他マーケットにリリースされている検体のほうが，VT 検知数が全体的に多い傾向もみてとれた．

第 5 章　マーケットのセキュリティ管理状況

本章では，マーケットのセキュリティ品質について議論する．各マーケットについてその品質を決定するいくつかの要因を調査し，それらをもとにセキュリティインデックスという指標を定義した．また，算出したセキュリティインデックスとそのマーケットのもつ影響度との関係の可視化を行った．

5.1　マーケットの品質

マーケットの品質を決定する要因は複数考えられる．例えばユーザがアプリをダウンロードする手段を考えると，Web サイトを経由するか専用のクライアントアプリを使用するかどうかはマーケットによって異なる．ユーザにとっての使い勝手という側面で見た場合には，サイトやクライアントアプリのユーザインタフェースが利便性を左右する一つの要因となりうる．マーケットがどの言語で提供されているかどうかもユーザにとって重要な点である．英語や特定の国の言語だけでなく，様々な言語でマーケットが提供されていれば，より多くのユーザを集めることができる．アプリ開発者の視点で見ると，例えば自身の開発したアプリの管理を容易にする仕組みが整っているかどうかはマーケットによって異なると考える．また開発者登録のような，マーケットにアプリを登録する手順の煩雑さや，審査の難易度といった要素もマーケットの品質に関わる．本研究では，こうしたマーケットの品質の中でも，特にユーザにとってのセキュリティに関するものに着目する．

5.2　セキュリティインデックス

例えば，マルウェアの数が多くても，それをユーザに容易にダウンロードさせない工夫がマーケット側でなされていれば，ある程度は安全と捉えることができる．我々は各マーケットの品質を探るため，表 5.1 に示す指標をマーケットごとに調査した．良性アプリの割合は 0〜1 の連続値で，その他は 0 か 1 の二値である．また，それぞれの指標は個々のアプリについてではな

表 5.1　マーケット品質の指標

指標	重み	説明
良性アプリの割合	5.0	図 3.1 に示した良性アプリの割合 (0〜1)
レビューシステム	0.5	ユーザによるレビュー機能があれば 1
パーミッション説明	0.5	アプリの持つパーミッション説明があれば 1
通報システム	0.5	ユーザによる不適切アプリの通報機能があれば 1
セーフティバッジ	0.5	AV スキャン済みマーク等の安全性を示す印があれば 1
HTTPS 対応	0.5	サイトが HTTPS 対応であれば 1

く，各マーケットとしてその項目を満たしているかどうかという点で判断している．

マーケットごとに，各指標に重みをかけて足し合わせたものをそのマーケットのセキュリティインデックスと定義した．セキュリティインデックスは 0.0 から 7.5 までの値を取る．すなわちマーケット品質の各指標の集合を $X = \{x_1, x_2, ..., x_n\}$ とし，それぞれに対する重みの集合を $W = \{w_1, w_2, ..., w_n\}$ とすると，セキュリティインデックスは以下の式で表される．

$$SecurityIndex = \sum_{i=1}^{n} w_i x_i \tag{5.1}$$

算出したセキュリティインデックスを持つそれぞれのマーケットが有する影響力を可視化するため，Alexa[4] のランキングを用いた．Alexa Internet が提供するこのランキングは，独自の基準でウェブサイトをドメイン単位で順位付けしたものである．ドメイン単位での集計のためマーケットそのもののランキングではないことに留意する必要はあるが，本研究ではこの順位が高いほどユーザへの影響力が高いという仮定に基づいて分析を行った．図 5.1 に，セキュリティインデックスと Alexa ランキングとの関係をプロットしたものを示す．なお，厳密にはマーケットと言えない *Genome* や *Torrent*，すでに閉鎖した *Apk_bang* は含めていない．

Baidu や *Yandex*，*Mi.com* のように，Alexa ランキングが高いにも関わらずインデックスが比較的低いものについては，マーケットとして改善の余地があると考える．これらの多くはマルウェア率の高さが寄与している．また，ランキング 10,000 位から 100,000 位の間に存在するマーケットのセキュリティインデックスは，高いものから低いものまで様々である．Google Play は全体で見るとセキュリティインデックスは高いが，セーフティバッジが無いため 7.0 を下回った．

Google Play と同程度のインデックスを持つ *Cafebazaar* はイラン国内で著名なマーケットである．第 2 章で述べたように，イランではもともと Google Play が利用できなかった．現在でも利用できるものは無料アプリのみである．このような背景から，*Cafebazaar* のようなサードパーティーマーケットの需要は高いと予想できる．第 3 章でも述べたように，VT の未スキャン率は調査したマーケットの中で最も高かった．実際にクロールしたアプリをいくつか見ると，

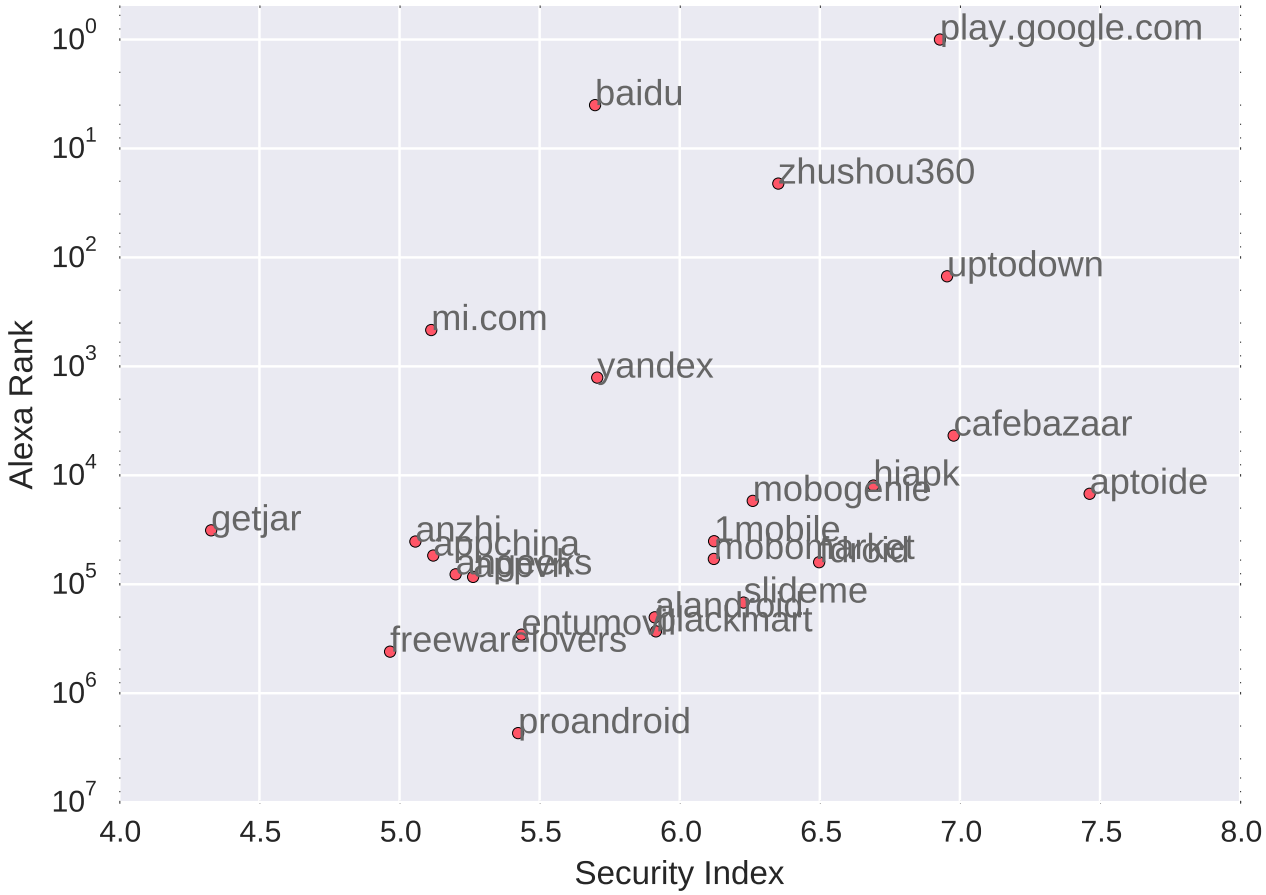

図 5.1　セキュリティインデックスと Alexa ランキングとの関係

ペルシャ語のローカルなアプリも数多く展開されていた．しかし実際のマルウェア率はかなり低い結果となった．有料アプリの課金システムも整っており，運営体制の品質の高さが伺える．

第 6 章　議論

本章では，本研究で用いた手法の制限事項，及びそれらから考えられる今後の展望について述べる．

6.1　クロール方法

研究対象として選定したマーケットがサードパーティーマーケットの全てではない．すべてを網羅することは事実上不可能だが，可能な限り多くの国を含むように選定を行った．また，各マーケットの全アプリをクロールしたわけではない．基本方針としてランキングを巡回することで取得した．マーケットによってはランキング外のアプリが存在することもあり，本手法では取得できていない可能性がある．したがって，取得したアプリに偏りが生じている可能性は否定できない．解決方法として，マーケットのアプリ検索機能に対して様々な検索語句を試行することで，ランキング外アプリを取得する方法がある．また，関連研究 [30] のように継続してクロールを行っていないため，アプリのアップデートや削除の遷移までは捉えきれないことも，本手法の制限である．

6.2　分析手法

本研究では，詳細なコードレベルの品質でアプリの分析を行っていない．例えばマルチリリースアプリについてその前後で変更されたコードの比較が考えられるが，スケーラビリティを考慮して今回は割愛した．

また今回はマルチリリースの基準として，パッケージ名の一致に着目した．実際はに同じマーケット上で同一パッケージ名のアプリは投稿できないことが多いため，第三者がそのようなマーケットにマルチリリースを行う際はパッケージ名を変更する必要が生じる．特にリパッケージングアプリに関してはパッケージ名の変更は常套手段であり，ランダムな文字列にしたり，元のパッケージ名の語尾に別の文字を付加たりする事例が多い．したがって，そうしたアプリのマルチリリースの検知には本手法では対応できない．そうした事例に対応するためには，APPraiser[29] のようなリパッケージングアプリの検出が必要になる．

第 5 章ではマーケットの品質について議論を行ったが，完全な定量化は難しい．例えば，指標として挙げた「通報システム」はその通報内容がセキュリティに関するもののみとは限らない．著作権の侵害や，年齢規制すべきものへの通報が行われる場合もありうる．また「セーフティバッジ」の有無も，マーケットによって基準は異なる．このほかにも開発者がアプリを投稿する難易度を例にあげると，開発者登録に電話番号が必要かどうかの有無や，審査の厳しさ等はマーケットごとに異なり，これらすべてを客観的に数値化することは非常に困難である．本研究では一つの目安として，ユーザの視点で認知できる指標を選定した．

6.3 メタデータの不足

本研究では一部に Androzoo を活用したため，Google Play のメタデータが存在しない．Google Play のメタデータを利用することで，4.5 節で述べたようなマルチリリースの遷移を投稿日から分析したり，開発者情報を各マーケットと比較したりするという分析が可能となる．例えば公式マーケットにリリースされたアプリが，別のマーケットにマルチリリースされるまでにどの程度の期間があるのかを把握するのに役立てることができる．我々が新たに Google Play のメタデータをクロールすることも可能だが，過去バージョンのデータをクロールすることはできない．また検体数が多いため，すべてを網羅することは難しいだろう．

第 7 章　関連研究

本章では，本研究に関連する既存研究を紹介する．はじめに Android アプリのリパッケージングに関連した研究について述べ，続いてサードパーティーマーケットを扱った研究について述べる．

アプリのリパッケージングは Android セキュリティにおいて大きな関心事である．リパッケージングとは Android アプリの実行ファイルである APK をディスアセンブルし，悪性コードや広告を挿入し改変したうえで再ビルドする手法である [15]．様々なマーケットを通じて流通しているこれらのリパッケージングアプリを検出する手法は，これまでにも数多く提案されてきた [37, 36, 35, 32]．

以下に，マーケット自体を分析対象とした研究を示す．2011 年に Vidas ら [33] は 194 マーケットを対象に合計 41,057 検体を収集してリパッケージングアプリの分析を行った．不正なリパッケージングを防ぐため，彼らは AppIntegrity と呼ばれるアプリの認証プロトコルを提案した．これはパッケージ名に含まれるドメイン名を利用して，開発者のサーバに問い合わせを行うものである．

Lindorfer らは 2013 年，事前調査として 8 マーケットを対象にクロールした結果をもとに，Andradar を開発した [30]．Andradar は 16 マーケットを対象にマルウェアの流通をリアルタイムにスキャンし追跡する．マルウェアがマーケットを跨いでリリースされ，時間経過とともにマーケットから削除される過程も観測しており興味深い．Andradar 自体は対象をマルウェアに絞っており，各マーケットのアプリを大規模にクロールしたデータを使っているわけではない点が我々の研究とは異なる．

Viennot らは 2014 年に PlayDrone[34] と呼ばれる Google Play クローラを開発し，収集したデータを用いて様々な分析を行った．レビュー評価とダウンロード数との相関や類似アプリの分析，アプリに含まれる認証トークンの脆弱性調査等を通じて，公式マーケットの実態を明らかにした．我々の研究では，Google Play だけでなくサードパーティーマーケットについてを対象とし，マーケットの品質を決める要因について考察している点が異なる．

第 8 章　まとめ

本研究では Android サードパーティーマーケットの実態を調査するため，合計 27 のマーケットから約 470 万のアプリデータを収集し，分析を行った．本章では分析から得られた結果のハイライトを紹介し，そこから導かれる考察を述べる．最後に，それらを踏まえて Android アプリを取り巻く各ステークホルダーが実施すべき対策の一例を紹介する．

8.1　分析結果

本研究では Android サードパーティーマーケットの実態を調査するため，合計 27 のマーケットから約 470 万のアプリデータを収集した．収集したデータを分析した結果，これまであまり研究対象とされていなかったマーケットでは過去にウイルススキャンにかけられた形跡がない検体の割合が高かったこと，中国系マーケットのように特定のマーケット間でのマルチリリースが活発であること，無視できない数の有料アプリがマルチリリースされていること，ユーザに与える影響範囲が大きいにもかかわらず安全性の低いマーケットが存在すること等を明らかにした．

8.2　考察

本節では分析結果をもとに考察を行う．中国系のマーケットにおいて，悪性アプリを含めた互いのマルチリリースが活発に行われていた．この要因の一つとしては，それらが同一の言語圏のマーケットであるということが考えられる．中国語のアプリは中国語を主として用いるユーザによって利用されることが期待される．したがって，これらの間でマルチリリースが行われるのはもっともである．また，中国では Google Play 公式マーケットが自由に利用できないという事情から，これらのサードパーティーマーケットの需要が高いということも要因だと考える．しかしマーケットの質の観点でみると，第 5 章で述べたようにセキュリティインデックスが低い結果となった．これらのマーケットは悪性アプリの割合が高く，運営による適切な処理が行われてない，もしくは追いついていないと判断できる．また，中国では人口の多さゆ

え Android ユーザ数も多いため，攻撃者がターゲットにしやすいという事情があるのだと推測する．

一方で *Cafebazaar* のように，サードパーティーマーケットの需要が高く，かつ質の高いマーケットも存在した．サイトやクライアントアプリは丁寧に作られており，開発者に対する情報提供も豊富である．アプリの安全性は運営による審査によって担保されていると推測する．また，中国ほどのユーザ数は見込めないため，攻撃者によるターゲットにされにくいということも，悪性アプリの割合が低い要因の一つであると考える．

また，多くの有料アプリが無料で他のマーケットにマルチリリースされていた．第三者が有料アプリをマルチリリースすることによる直接の利点は定かではない．しかしマーケット運営者にとって，有料アプリが無料で手に入るということは，ユーザをそのマーケットに引きつける好材料となりうる．ユーザ数が多ければ多いほど，マーケットはより多くの広告収入等の機会を得ることができる．そのため，マーケット側が有料アプリの不正なマルチリリースを黙認している可能性は否定できない．しかし，当然ながらこれはオリジナルのアプリ開発者にとっての機会損失となる．次節で述べるように，意図せずマルチリリースが行われたときのための対策を講じるべきである．

8.3 対策

最後に，ユーザ向け，開発者向け，マーケット向けにそれぞれの観点から実施すべき対応策を述べる．

ユーザ

各マーケットに記載されている様々な指標を吟味し，アプリの安全性を見極める．またアプリの提供元が信頼できない場合，ウイルスチェッカーにかけて安全性を確認することが望ましい．

開発者

自らが開発したアプリがサードパーティーマーケットで不正に出回っていないかを監視する．不正に出回っていることが発覚した場合は，迅速に削除申請等の措置をとる．リパッケージングアプリの流通を防ぐため，アプリに改ざん対策を施す．有料アプリについては上記に加えて LVL を実装する．適切に LVL が実装されていれば，たとえマルチリリースされていても起動を防ぐことができる．

マーケット

一部のマーケットでは，アプリをリリースするために Google Play からインポートできる機能が存在する．こうしたマーケットでは，インポートの際に Google Play に登録されている開発者のメールアドレスに通知を行うことで，本人の認証を行っていた．このように，正規の開発者による投稿であるか否かを承認する仕組みがあれば，第三者による不正なマルチリリースをある程度は防ぐことができる．インポート機能がない場合で

も，投稿されたアプリと同名のパッケージ名が Google Play に存在するかどうかを確認し，存在する場合はその開発者に通知するという仕組みが考えられる．また，ユーザや開発者に対して，マーケットの運営母体についての情報や運営ポリシーを適切に開示することも大切である．

研究業績

学部・専攻在籍中の研究業績を下記に示す.

ジャーナル論文誌

1. Y. Ishii, T. Watanabe, M. Akiyama, and T. Mori, "APPraiser: A large scale analysis of Android clone apps," IEICE Transactions on Information Systems, Vol. XX, No. XX, pp. XX-XX, 2017 (条件付採録)

国際会議 (査読付き)

1. Y. Ishii, T. Watanabe, M. Akiyama, and T. Mori, "Clone or Relative?: Understanding the Origins of Similar Android Apps," Proceedings of the ACM International Workshop on Security And Privacy Analytics (IWSPA 2016), pp. 25-32, Mar 2016

国際ポスター発表 (査読付き)

1. Y. Ishii, T. Watanabe, M. Akiyama, and T. Mori, "Understanding the Origins of Similar Android Apps," The 18th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2015)

国内研究会

1. 石井 悠太，渡邉 卓弥，金井 文宏，高田 雄太，塩治 榮太朗，秋山 満昭，八木 毅，森 達哉，"Android サードパーティーマーケットの大規模調査"，2017 年暗号と情報セキュリティシンポジウム
2. 石井 悠太，渡邉 卓弥，秋山 満昭，森 達哉，"Android クローンアプリの大規模分析"，コンピュータセキュリティシンポジウム 2015 論文集，vol. 2015，No. 3，pp. 207-214，2015 年 10 月
3. 石井 悠太，渡邉 卓弥，秋山 満昭，森 達哉，"正規アプリに類似した Android アプリの実態解明"， 信学技報，vol. 114，no. 489，ICSS2014-94，pp. 187-192，2015 年 3 月

受賞

1. MWS2015 学生論文賞，2015 年 10 月
2. ICSS 研究会 2014 年度研究賞，2015 年 3 月

謝辞

本研究を進めるにあたり，多くの議論や指導をしてくださった森達哉准教授に感謝いたします．また，ゼミをはじめ日頃より議論に応じてくださった森研究室の皆様に感謝いたします．このほか，各国のマーケット事情の調査に協力をしていただいた方々に心から感謝申し上げます．本研究の一部は JSPS 科研費 JP16H02832 の助成を受けたものです．

参考文献

[1] 360 手机助手. http://zhushou.360.cn/.

[2] Adding licensing to your app | android developers. https://developer.android.com/google/play/licensing/adding-licensing.html.

[3] alandroidnet.com. https://www.alandroidnet.com/.

[4] Alexa - actionable analytics for the web. http://www.alexa.com/.

[5] Android セキュリティ技術の最前線 (4): Android アプリマーケットを守る「不正アプリ抽出技術」総解説 - @ it. http://www.atmarkit.co.jp/ait/articles/1606/07/news005.html.

[6] apktool. https://code.google.com/p/android-apktool/.

[7] Apple, huawei, and xiaomi finish 2015 with above average year-over-year growth, as worldwide smartphone shipments surpass 1.4 billion for the year, according to idc. http://www.idc.com/getdoc.jsp?containerId=prUS40980416.

[8] Aptoide - android apps store. https://www.aptoide.com/.

[9] Blackmart alpha | download android market: Blackmart apk. http://www.blackmart.us/.

[10] Descarga de apps para android - descarga, descubre, comparte en uptodown. https://www.uptodown.com/android/.

[11] Download | install android apps | cafe bazaar. https://cafebazaar.ir/.

[12] Download free android games & apps on mobomarket android market. http://www.mobomarket.net/.

[13] entumovil. http://www.entumovil.cu/downloads/apps.

[14] F-secure: Android accounted for 97% of all mobile malware in 2013, but only 0.1% of those were on google play. http://thenextweb.com/google/2014/03/04/f-secure-android-accounted-97-mobile-malware-2013-0-1-google-play/.

[15] Fake apps: Feigning legitimacy. http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-fake-apps.pdf.

[16] Getjar - download free apps, games and themes apk. http://www.getjar.com/.

[17] jadx. https://github.com/skylot/jadx.

[18] Kho android appvn. http://appvn.com/android.

[19] Lancers. http://www.lancers.jp/.

[20] Mobogenie - download android apps and games for free. http://www.mobogenie.com/.

[21] New ransomware "charger" found on the google play store | android-headlines.com. http://www.androidheadlines.com/2017/01/new-ransomware-charger-found-on-the-google-play-store.html.

[22] smali. https://code.google.com/p/smali/.

[23] Supported locations for distribution to google play users. https://support.google.com/googleplay/android-developer/table/3541286.

[24] The ultimate app store list | business of apps. http://www.businessofapps.com/the-ultimate-app-store-list/.

[25] Virustotal. https://www.virustotal.com/.

[26] Yandex.store is the app store for your android phone. https://m.store.yandex.com/.

[27] 百度手机. http://shouji.baidu.com/.

[28] Kevin Allix, Tegawendé F Bissyandé, Jacques Klein, and Yves Le Traon. Androzoo: collecting millions of android apps for the research community. In *Proceedings of the 13th International Workshop on Mining Software Repositories*, pages 468–471. ACM, 2016.

[29] Yuta Ishii, Takuya Watanabe, Mitsuaki Akiyama, and Tatsuya Mori. Clone or relative?: Understanding the origins of similar android apps. In *Proceedings of the 2016 ACM on International Workshop on Security And Privacy Analytics*, pages 25–32. ACM, 2016.

[30] Martina Lindorfer, Stamatis Volanis, Alessandro Sisto, Matthias Neugschwandtner, Elias Athanasopoulos, Federico Maggi, Christian Platzer, Stefano Zanero, and Sotiris Ioannidis. Andradar: fast discovery of android applications in alternative markets. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 51–71. Springer, 2014.

[31] J. Oberheide and C. Miller. Dissecting the android bouncer. SummerCon, Brooklyn, NY., 2012.

[32] Shao, Yuru, Luo, Xiapu, Qian, Chenxiong, Zhu, Pengfei, Zhang, and Lei. Towards a scalable resource-driven approach for detecting repackaged android applications. In *Proceedings of the 30th Annual Computer Security Applications Conference*, pages 56–65. ACM, 2014.

[33] Timothy Vidas and Nicolas Christin. Sweetening android lemon markets: measuring and combating malware in application marketplaces. In *Proceedings of the third ACM conference on Data and application security and privacy*, pages 197–208. ACM, 2013.

[34] Nicolas Viennot, Edward Garcia, and Jason Nieh. A measurement study of google play. Proc.

of ACM SIGMETRICS 2014, June 2014.

[35] Zhauniarovich, Yury, Gadyatskaya, Olga, Crispo, Bruno, La Spina, Francesco, Moser, and Ermanno. Fsquadra: Fast detection of repackaged applications. Proc. of IFIP DBSec '14, pages 131–146, 2014.

[36] Wu Zhou, Yajin Zhou, Michael Grace, Xuxian Jiang, and Shihong Zou. Fast, scalable detection of "piggybacked" mobile applications. In *Proc. of the third ACM CODASPY 2013*, pages 185–196.

[37] Wu Zhou, Yajin Zhou, Xuxian Jiang, and Peng Ning. Detecting repackaged smartphone applications in third-party android marketplaces. In *Proc. of the second ACM CODASPY 2012*, pages 317–326.

[38] Yajin Zhou and Xuxian Jiang. Dissecting android malware: Characterization and evolution. In *2012 IEEE Symposium on Security and Privacy*, pages 95–109. IEEE, 2012.