4Z-3

マニフェストファイルの分析によるAndroidマルウェア検出法

Analysis of Manifest File for Detecting Android Malware

佐藤 亮† 千葉 大紀‡ 後藤 滋樹†

†早稲田大学 基幹理工学部 情報理工学科
‡早稲田大学 基幹理工学研究科 情報理工学専攻

概 要

Android スマートフォンの急速な普及に伴い，Android を対象とするマルウェアの脅威が拡大している．一方で，急増する Android マルウェアへの従来の対策は十分ではない．そこで本論文は新たな Android マルウェア検出手法を提案する．本手法は，Android アプリケーションの中に必ず含まれるマニフェストファイルのみを分析対象とするのが特徴である．具体的には，マニフェストファイルに含まれるアプリケーションに関する様々な情報を抽出し，良性アプリケーション (以下良性アプリ) とマルウェアを識別する．実データを用いた評価実験の結果，本手法が未知の Android マルウェアに対しても有効であることがわかった．

1 提案手法

本研究では，Android アプリケーションに必ず含まれるマニフェストファイル (AndroidManifest.xml) を利用した Android マルウェアの検出手法を提案する．具体的には，以下の 3 つの手順でマルウェアを検出する．(1) 検体のマニフェストファイルに記述されている特定の情報を抽出し，(2) 本研究で独自に作成した判断基準との比較を行い，(3) 悪性度が高いと判断されたアプリケーションをマルウェアと判定する．以下に提案手法の詳細を述べる．

1.1 抽出する情報の選定

マニフェストファイルが持つ様々な情報の中から，マルウェアの検出に有効な情報を選定する．今回は事前に良性アプリ 30 個とマルウェア 30 個を分析し，良性アプリとマルウェアの識別に実際に有効な情報を選定した．表 1 に利用する情報を示す．ここで，パーミッション名とはアプリケーションが使用する権限のことである．action 名および category 名とは，アプリケーションが応答する Intent (アプリケーション間で連携を取るための仕組み) の種類を規定したものである．プロセス名は，アプリケーションを実行するプロセスの名前である．優先度とは，Intent に対する優先度を整数値で設定したものであり，値が高いほうが高優先度となる．またマルウェアの中には，マニフェストファイル上に同種のパーミッションを複数回宣言しているものが存在した．そこで，パーミッションの再定義回数を情報の一つとして用いる．

表 1: マニフェストファイルから抽出する情報一覧

(1) パーミッション名
(2) action 名 (Intent-Filter)
(3) category 名 (Intent-Filter)
(4) 実行されるプロセス名
(5) 優先度 (Intent-Filter)
(6) パーミッションの再定義回数

1.2 判断基準の作成

本手法では，表 1 に示した 6 種類の情報のうち，(1) パーミッション名，(2) action 名，(3) category 名，(4) 実行プロセス名の 4 種類の情報に対して，それぞれキーワードリストを作成した．キーワードリストとは，マルウェア検出に有効なキーワード (文字列) をまとめたものであり，マルウェアはより多くのキーワードを含む．なお，(5) 優先度 (Intent-Filter) および (6) パーミッションの再定義回数に関しては，数値そのものを判断基準とする．

1.3 良性・悪性の判定

本手法では，上記の判断基準をもとに (1) 作成したキーワードリストとの一致率，(2) 優先度の値，(3) パーミッション再定義回数の 3 つに対して閾値を設定する．良性・悪性の判定は，この閾値との比較により行う．

2 性能評価

2.1 実証実験の概要

提案手法の未知の検体に対する有効性を評価する．実験に利用した検体は，公式および非公式のマーケットから無作為に選定，収集した良性アプリ 235 個と，研究用にマルウェア検体を公開しているサイト [1] より収集した Android マルウェア 130 個である．これらの検体は，第 1.1 節で用いた検体とは重複しない．収集した検体が確かに良性アプリ，あるいはマルウェアであることは，Virus Total [2] を用いて確認した．Virus Total は，複数のアンチウイルスエンジンを利用して検体の検査が行えるフリーの Web サービスである．本実験では，収集した 130 個のマルウェアのうち，Virus Total で解析された日時が 2011 年 9 月以前のものを学習データ，それ以降のものをテストデータとして利用した．これにより，本実験で扱うテストデータは，ハッシュ値を利用するシグネチャベースの既存手法では検出できない未知の検体として扱うことができる．本実験で使用する各検体の個数を表 2 に示す．

表 2: 各データの個数

	良性アプリ	マルウェア
学習データ (個)	60	34
テストデータ (個)	175	96
総数 (個)	235	130

2.2 実験の結果

テストデータとして用意した良性アプリ 175 個，マルウェア 96 個に対する提案手法の検出精度を求めた．精度の評価指標としては，正答率および誤検知率を用いた．

- 正答率：検体を正しく分類できた割合を意味する．具体的には，良性アプリを正しく良性，マルウェアを正しく悪性と判断できた割合を示す．
- 誤検知率：検体を正しく分類できなかった割合を意味する．具体的には，本来良性な検体を悪性，またはマルウェアを良性と判断した割合を示す．

実験の結果を表 3 に示す．

表 3: 提案手法の検出精度

	正答率 (%)	誤検知率 (%)
良性アプリ	91.4	8.6
マルウェア	87.5	12.5
全体	90.0	10.0

2.3 実験の考察

実験結果に対する考察を述べる．まず正答率は，良性アプリに対しては 91.4%，マルウェアに対しては 87.5% であり，提案手法は Android 検体を高精度で分類できることがわかる．本実験において学習データとして用いた検体はすべて，テストデータ用の検体より発見時期が早いものであった．このことから，本手法において使用するマニフェストファイルの情報は検体の発見時期に依存せず，未知のマルウェアも検出可能であることが分かる．また，検出できなかったマルウェアについて調査したところ，本手法ではアドウェアの検出率が低いことがわかった．アドウェアは，広告の過剰表示という動作以外は良性アプリと同等である場合が多い．そのため，マニフェストファイルにもほとんど差異が見られず，結果として誤検知された．

3 まとめ

本研究では，マニフェストファイルから得られる情報のみを利用して Android マルウェアを検出する手法を提案した．マニフェストファイルは，すべてのアプリケーションの中に必ず含まれる．そのため，本手法はすべての検体に対して適用が可能である．また実証実験より，既知のパターンを用いるシグネチャベースの手法では検出できない未知のマルウェアに対しても本手法は有効であることが示せた．マニフェストファイルのみの分析は低コストで行えることから，他の手法との組み合わせも可能である．

4 今後の課題

本研究では評価実験を行ったが，実験に利用できた検体の数が少なかった．今後はさらに多くの検体を収集し，より信頼性の高い評価実験が行えるようにしたい．また，本研究では 6 種類の情報をマニフェストファイルから抽出し，利用した．しかし，この他にも Android マルウェアの検出に有効な情報が含まれている可能性がある．今後は，マニフェストファイルが持つ情報をより詳細に分析し，さらに有効なマルウェア検出手法を提案したい．

参考文献

[1] contagio mobile
http://contagiominidump.blogspot.jp/

[2] Virus Total
https://www.virustotal.com/

[3] Yajin Zhou, Xuxian Jiang, “Dissecting Android Malware: Characterizarion and Evolution”, Security and Privacy (SP), 2012 IEEE Symposium on, pp.95–109, May 2012.