

## 混在するスマートフォンアプリケーションの パケット分類手法

NTTネットワーク基盤技術研究所  
中野雄介, 上山憲昭, 塩本公平  
大阪大学大学院情報科学研究科  
村田正幸, 宮原秀夫  
大阪大学サイバーメディアセンター  
長谷川剛

Copyright©2014 NTT corp. All Rights Reserved.

## 背景

- スマートフォンの普及による、スマートフォンのトラフィックの増加

年	スマートフォンによるトラフィック (%)	モバイルネットワーク全体のトラフィック (%)
2013	12%	88%
2014	10%	90%
2015	9%	91%
2016	7%	93%
2017	5%	95%
2018	4%	96%

モバイルネットワークに対する影響の増大

http://www.cisco.com/web/IP/solution/isp/pngn/literature/white\_paper\_c11-520862.html Copyright©2014 NTT corp. All Rights Reserved. 2

## 課題

- 通信事業者以外の事業者やユーザによるアプリケーション作成
- ネットワークへの負荷を意識しないアプリケーション

**特定アプリケーションによるリソースの専有**

複数のスマートフォンアプリケーションの packets が混在する  
トラフィックデータをアプリケーションごとに解析をする必要がある

Copyright©2014 NTT corp. All Rights Reserved. 3

## 既存技術

- TCPヘッダ, IPヘッダの情報でアプリケーションごとのパケットに分類

スマートフォンアプリケーションにはそのまま使えない

HTTPを用いるアプリケーションが多い

TCP,IPヘッダのみでアプリケーションごとに分類することが困難

Copyright©2014 NTT corp. All Rights Reserved. 4

## 研究の目的

- 様々なスマートフォンアプリケーションの packets が混在したキャプチャデータから、アプリケーションごとの packets に分類

将来的には

**アプリケーションごとのトラフィック解析の実現**

※Androidアプリケーションを対象とする

Copyright©2014 NTT corp. All Rights Reserved. 5

## 提案手法

- 下記2つの解析手法を組み合わせる

**Androidアプリケーションの通信の特徴を抽出**

ソースコードに含まれる通信関連の文字列と宛先ホスト名とを通信の特徴として抽出

**キャプチャデータからのアプリケーションごとの packets 抽出**

抽出された通信の特徴を用い、キャプチャデータから各アプリケーションの通信の特徴と類似する packets をアプリケーションごとに抽出

Copyright©2014 NTT corp. All Rights Reserved. 6

### Androidアプリケーションの通信の特徴抽出

- HTTP関連の特徴抽出
  1. 各アプリケーションのAPKファイルからソースコードを生成
  2. ソースコードから、URL、HTTPヘッダ情報を抽出
- HTTP以外の特徴抽出
  1. 静的解析ツールで通信関連のメソッド名を抽出
  2. 予め作成されたメソッド毎の宛先ホストリストを用い、通信関連のメソッド名から宛先ホスト名に変換

①ソースコード生成 ②ホスト名、URL、HTTPヘッダ情報抽出

通信の特徴抽出: アプリαのバイナリ → アプリαのソースコード → HTTPヘッダ情報(送信先ホスト名)

HTTP以外の特徴抽出: 静的解析ツール※ → 通信関連のメソッド名 → 宛先ホスト名

メソッド毎の宛先ホストリスト

※http://sable.github.io/soot/

Copyright©2014 NTT corp. All Rights Reserved. 7

### Androidアプリケーションの通信の特徴抽出-詳細

- HTTP関連の特徴
  - ソースコードからのキーワードに一致する部分を抽出
    - ソースコード中の記述例
- HTTP以外の特徴
  - メソッド名から宛先ホスト名に変換
    - GoogleCloudMessaging → android.googleapis.com
    - Ads → android.clients.google.com, googleads.g.doubleclick.net, redirector.gvt1.com, r5---sn-3pm7enes.gvt1.com, dl.google.com, api.criticism.com, www.googleapis.com

```

localHttpGet.setHeader("Content-Type", "application/json");
localHttpPost.setHeader("Content-Type", "application/json");

public static final String BASE_URL = "http://officialapi.spikaapp.com";
public static final String INFORMATION_URL = "http://officialapi.spikaapp.com/page/information/";
public static final String LIST_SERVERS_URL = "http://officialapi.spikaapp.com/api/servers";
    
```

Copyright©2014 NTT corp. All Rights Reserved. 8

### アプリケーションごとのパケット抽出

1. スマートフォンアプリケーションのパケットを含む通信をキャプチャし、端末のIPアドレス毎に分類
2. 通信の特徴を用い、キャプチャデータから各アプリの通信の特徴と一致するパケットを発見
3. 発見されたパケットを各アプリケーションのパケットとして抽出

①キャプチャデータ収集、②アプリケーションの通信の特徴でキャプチャデータから発見、③アプリケーションごとのパケット抽出

抽出対象: Spika, Line (インスタントメッセージアプリ)

抽出結果: アプリα (AAA, BBB, CCC), アプリβ (XXX, YYY, ZZZ)

Copyright©2014 NTT corp. All Rights Reserved. 9

### アプリケーションごとのパケット抽出-詳細

- HTTP
  - 各アプリケーション特有のHTTPヘッダと一致するパケットをキャプチャ結果から発見
- HTTP以外
  - 送信先ホスト名に対するDNSクエリを発見し、そのAnswerのIPアドレスを送信先とするパケットをキャプチャ結果から発見

抽出対象: アプリαの特徴

- HTTPヘッダ: HTTP, CONTENT-TYPE = application/json, HOST = officialapi.spikaapp.com
- DNS (Answer): HTTP以外, 送信先ホスト: android.googleapis.com
- 送信先IP
- 送信先IP

Copyright©2014 NTT corp. All Rights Reserved. 10

### 提案手法の評価

- 評価手法
  1. Spika, Line (インスタントメッセージアプリ)が混在したキャプチャデータ取得
    1. Spika:メッセージ送信 Line:アイドル状態
  2. キャプチャデータから提案手法を用い、Spikaのみのパケット抽出
  3. スマートフォン内でSpikaのみのパケットをキャプチャしたものと比較し、抽出結果の適合率、再現率算出

抽出対象: Spika (インスタントメッセージアプリ), Nexus5 (スマートフォン), Network Log (キャプチャツール)

キャプチャ用PC: Wireshark

インターネット

混在したキャプチャデータ

提案手法

抽出結果: Spikaのみ抽出結果(正解), Spikaのみ抽出結果(評価対象)

適合率, 再現率

Copyright©2014 NTT corp. All Rights Reserved. 11

### 提案手法の評価

- 評価結果

適合率	再現率
0.98075	1.0

- 適合率: 抽出されたSpikaのパケット数/抽出されたパケット数
- 再現率: 抽出されたSpikaのパケット数/抽出するべきSpikaの全パケット数

- 考察
  - 誤って抽出したと判定された理由(適合率)
    - Network LogはTLSのパケットを抽出対象としていなかった。一方、提案手法はTLSを抽出したため ⇒提案手法が誤って抽出したわけではない。

今後、更に多くのAndroidアプリケーションを評価対象とする

Copyright©2014 NTT corp. All Rights Reserved. 12

### 今後の課題

複数のアプリケーションが同一のホストと通信する場合、アプリケーションごとの特徴の違いが無く、パケットの分類が困難

複数アプリがGoogle Cloud Messagingと通信する例

アプリα

特徴  
android.googleapis.com  
と通信

両方と一致

アプリβ

特徴  
android.googleapis.com  
と通信

pcap

DNS android.googleapis.com  
Answer XX.XX.XX.XX

XX.XX.XX.XX

XX.XX.XX.XX

通信の順番の考慮が必要

Copyright©2014 NTT corp. All Rights Reserved. 13

### 今後の課題

#### -通信の順番を考慮した特徴の抽出-

- 各アプリケーションのAPKファイルから静的解析ツール(Soot)によって制御フローグラフを生成
- 制御フローグラフから通信関連のメソッドのみの呼び出し順番を生成
- 各メソッドに対する通信の特徴(後述)と、通信関連のメソッドのみの制御フローグラフから通信の順番を生成

①制御フローグラフ抽出
②通信関連メソッドの呼び出し順番生成
③アプリケーションの通信の順番を生成

アプリαのバイナリ → 静的解析ツール ※ → HTTPリクエスト ↓ Google Cloud Messaging ↓ HTTPリクエスト

各メソッドに対する通信の特徴 → 通信の順番

※http://sable.github.io/soot/ Copyright©2014 NTT corp. All Rights Reserved. 14

### 今後の課題

#### -各メソッドに対する通信の特徴の抽出-

- 通信関連のメソッドを実行する実験用アプリケーションを作成
- 実験用アプリケーションを実行することで、各メソッドに対する通信の特徴を抽出

通信関連のメソッドを実行する実験用アプリ → 生成 →

メソッド	methodA	methodB	methodX	methodY	...
パケットの特徴	AAA	BBB	XXX	YYYY	...

パケットの特徴の例

GCM	DNSサーバ	android.googleapis.com
-----	--------	------------------------

抽出対象

各パケットの宛先ホスト名、プロトコル、サイズ、含まれる文字列等

Copyright©2014 NTT corp. All Rights Reserved. 15

### 今後の課題

#### -通信の順番を考慮したパケット抽出-

- スマートフォンアプリケーションのパケットを含む通信をキャプチャし、端末のIPアドレス毎に分類
- 通信の順番を用い、キャプチャデータを走査し、各アプリの順番と一致するパケットを発見
- 発見されたパケットを各アプリケーションのパケットとして抽出

①キャプチャデータ収集、②アプリケーションの通信の順番でキャプチャデータを走査、③アプリケーション端末ごとに分類 ことのパケット抽出

NW → AAA BBB → 1.AAA 2.BBB 3.CCC → アプリα (AAA BBB CCC)

XXX YYY ZZZ → アプリβ (XXX YYY ZZZ)

Copyright©2014 NTT corp. All Rights Reserved. 16

### まとめ

- 混在するスマートフォンアプリケーションのパケット分類手法を提案
  - アプリケーションごとに、ソースコードに含まれる通信関連の文字列を通信の特徴として抽出
  - 抽出された通信の特徴を用い、キャプチャデータから各アプリケーションの通信の特徴と類似するパケットを抽出
  - Spikaのパケットのみを分類できていることを確認
- 今後は、
  - 通信の順番を考慮することで、共通のホストと通信する複数のアプリケーションのパケットを分類できる手法を実現
  - 評価対象のアプリケーションの増加

Copyright©2014 NTT corp. All Rights Reserved. 17