Internet of things paradigm

Technology, Computer

Introduction

According to 2016 statistical forecast, there are almost 4. 77 billion number of mobile phone users in globally and it is expected to pass the five billion by 2019. [1] The main attribute of this significant increasing trend is due to increasing popularity of smartphones. In 2012, about a quarter of all mobile users were smartphone users and this will be doubled by 2018 which mean there are be more than 2. 6 million smartphone users. Of these smartphone users more than quarter are using Samsung and Apple smartphone.

Until 2016, there are 2. 2 million and 2 million of apps in google app store and apple store respectively. Such explosive growth of apps gives potential benefit to developer and also companies. There are about $88. 3 billion revenue for mobile application market.

Prominent exponents of the IT industry estimated that the IoT paradigm will generate $1. 7 trillion in value added to the global economy in 2019. By 2020 the Internet of Things device will more than double the size of the smartphone, PC, tablet, connected car, and the wearable market combined.

Technologies and services belonging to the Internet of Things have generated global revenues in $4. 8 trillion in 2012 and will reach $8. 9 trillion by 2020, growing at a compound annual rate (CAGR) of 7. 9%.

From this impressive market growth, malicious attacks also have been increased dramatically. According to Kaspersky Security Network(KSN) data report, there has been more than 171, 895, 830 malicious attacks from online resources among word wide. In second quarter of 2016, they have detected 3, 626, 458 malicious installation packages which is 1. 7 times more than first quarter of 2016. Type of these attacks are broad such as RiskTool, AdWare, Trojan-SMS, Trojan-Dropper, Trojan, Trojan-Ransom, Trojan-Spy, Trojan-Banker, Trojan-Downloader, Backdoor, etc..

http://resources. infosecinstitute. com/internet-things-much-exposed-cyber-threats/#gref

Unfortunately, the rapid diffusion of the Internet of Things paradigm is not accompanied by a rapid improvement of efficient security solutions for those “ smart objects”, while the criminal ecosystem is exploring the technology as new attack vectors.

Technological solutions belonging to the Internet of Things are forcefully entering our daily life. Let’s think, for example, of wearable devices or the SmartTV. The greatest problem for the development of the paradigm is the low perception of the cyber threats and the possible impact on privacy.

Cybercrime is aware of the difficulties faced by the IT community to define a shared strategy to mitigate cyber threats, and for this reason, it is plausible that the number of cyber attacks against smart devices will rapidly increase.

As long there is money to be made criminals will continue to take advantage of opportunities to pick our pockets. While the battle with cybercriminals can seem daunting, it’s a fight we can win. We only need to break one link in their chain to stop them dead in their tracks. Some tips to success:

Deploy patches quickly

Eliminate unnecessary applications

Run as a non-privileged user

Increase employee awareness

Recognize our weak points

Reducing the threat surface

Currently, both major app store companies, Google and Apple, takes different position to approach spam app detection. One takes an active and the other with passive approach.

There is strong request of malware detection from global

Background (Previous Study)

The paper “ Early Detection of Spam Mobile Apps” was published by dr. Surangs. S with his colleagues at the 2015 International World Wide Web conferences. In this conference, he has been emphasised importance of early detection of malware and also introduced a unique idea of how to detect spam apps. Every market operates with their policies to deleted application from their store and this is done thru continuous human intervention. They want to find reason and pattern from the apps deleted and identified spam apps.

The diagram simply illustrates how they approach the early spam detection using manual labelling.

Data Preparation

New dataset was prepared from previous study [53]. The 94, 782 apps of initial seed were curated from the list of apps obtained from more than 10, 000 smartphone users. Around 5 months, researcher has been collected metadata from Goole Play Store about application name, application description, and application category for all the apps and discarded non-English description app from the metadata.

Sampling and Labelling Process

One of important process of their research was manual labelling which was the first methodology proposed and this allows to identify the reason behind their removal.

Manual labelling was proceeded around 1. 5 month with 3 reviewers at NICTA. Each reviewer labelled by heuristic checkpoint points and majority reason of voting were denoted as following Graph3. They identified 9 key reasons with heuristic checkpoints. These full list checkpoints can be find out from their technical report. (http://qurinet. ucdavis. edu/pubs/conf/www15. pdf)[]

In this report, we only list checkpoints of the reason as spam.

Graph3. Labelled spam data with checkpoint reason.

Checkpoint S1-Does the app description describe the app function clearly and concisely?

100 word bigrams and trigrams were manually conducted from previous studies which describe app functionality. There is high probability of spam

apps not having clear description. Therefore, 100 words of bigrams and trigrams were compared with each description and counted frequency of occurrence.

Checkpoint S2-Does the app description contain too much details, incoherent text, or unrelated text?

literary style, known as Stylometry, was used to map checkpoint2. In study, 16 features were listed in table 2.

Table 2. Features associated with Checkpoint 2

Feature

1

Total number of characters in the description

2

Total number of words in the description

3

Total number of sentences in the description

4

Average word length

5

Average sentence length

6

Percentage of upper case characters

7

Percentage of punctuations

8

Percentage of numeric characters

9

Percentage of common English words

10

Percentage of personal pronouns

11

Percentage of emotional words

12

Percentage of misspelled word

13

Percentage of words with alphabet and numeric characters

14

Automatic readability index(AR)

15

Flesch readability score(FR)

For the characterization, *feature selection of greedy method* [ ] was used with max depth 10 of decision tree classification . The performance was optimized by *asymmetric F-Measure*

[55]

They found that *Feature number 2, 3, 8, 9, and 10* were most discriminativeand spam apps tend to have less wordy app description compare to non-spam apps. About 30% spam app had less than 100 words description.

Checkpoint *S* 3 – Does the app description contain a noticeable repetition of words or key words?

They used vocabulary richness to deduce spam apps.

Vocabulary Richness (VR) =

Researcher expected low VR for spam apps according to repetition of keywords. However, result was opposite to expectation. Surprisingly VR close to 1 was likely to be spam apps and none of non-spam app had high VR result. [ ]

This might be due to terse style of app description among spam apps.

Checkpoint S4 – Does the app description contain unrelated keywords or references?

Common spamming technique is adding unrelated keyword to increase search result of app that topic of keyword can vary significantly. New strategy was proposed for these limitations which is counting the mentioning of popular applications name from app’s description.

In previous research name of top-100 apps were used for counting number of mentioning.

Only 20% spam apps have mentioned the popular apps more than once in their description. Whereas, 40 to 60 % of non-spam had mention more than once. They found that many of top-apps have social media interface and fan pages to keep connection with users. Therefore, theses can be one of identifier to discriminate spam of non-spam apps.

Checkpoint S5 – Does the app description contain excessive references to other applications from the same developer?

Number of times a developer’s other app names appear.

Only 10 spam apps were considered as this checkpoint because the description contained links to the application rather than the app names.

Checkpoint S6 – Does the developer have multiple apps with approximately the same description?

For this checkpoint, 3 features were considered:

The total number of other apps developed by same developer.

The total number of apps that written in English description to measure description similarity.

Have description *Cosine similarity(s)* of over 60%, 70%, 80%, and 90% from the same developer.

Pre-process was required to calculate the *cosine similarity: [ ]*

Firstly, converting the words in lower case and removing punctuation symbols.

Then calibrate each document with word frequency vector.

Cosine similarity equation: