WWW2015 FLORENCE, ITALY
24TH INTERNATIONAL WORLD WIDE WEB CONFERENCE
Florence 18 - 22 May 2015

Early Detection of Spam Mobile Apps

Suranga Seneviratney[1,2], Aruna Seneviratney[1,2], Mohamed Ali Kaafary[1], Anirban Mahantiy[1], Prasant Mohapatraz[3]

[1]NICTA (National ICT Australia)

[2] School of EET, UNSW (The University of New South Wales)

[3] Department of CS, University of California, Davis

Outline

- Background
- Motivation
- Early Detection of Spam Mobile Apps
- Evaluation
- Conclusion

Background

- As of mid-2014, Google Play Store and Apple App Store, each hosted approximately 1.2 million apps, with around 20,000 new apps being published each month in both of these app markets.

- Apps can be "spammy" in multiple ways including not having a specific functionality, unrelated app description or unrelated keywords and publishing similar apps several times and across diverse categories.

- Google and Apple take different approaches to the spam detection problem.

Outline

- Background
- Motivation
- Detection of Spam Mobile Apps
- Evaluation
- Conclusion

Motivation

- Spam apps vitiate the app market experience and its usefulness.

- With the ever increasing number of apps being submitted daily for approval, the app market operators need to be able to detect spam apps quickly and accurately.

Contributions

- The authors develop a manual app classification methodology based on a set of heuristic checkpoints that can be used to identify reasons behind an app's removal. They found that approximately 35% of the apps that were removed are spam apps.

- The authors present a mapping of their proposed spam checkpoints to one or more quantifiable features that can be used to train a learning model.

- The authors build an Adaptive Boost classifier for early detection of spam apps and show that their classifier can achieve an accuracy over 95% at a precision between 85%-95% and a recall between 38%-98%.

- The authors applied our classier to over 180,000 apps available in Google Play Store and show that approximately 2.7% of them are potentially spam apps.

Outline

- Background
- Motivation
- Early Detection of Spam Mobile Apps
- Evaluation
- Conclusion

Dataset

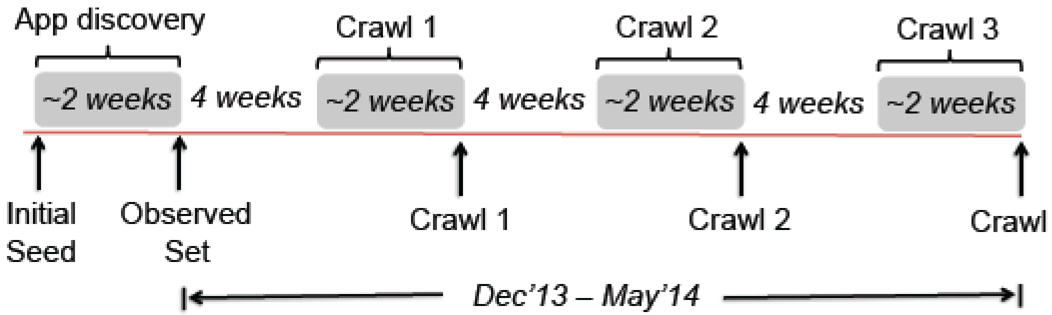


Table 1: Summary of the dataset

Set	Number of apps
Observed set ($\mathbb{O}$)	232,906
Crawl 1 ($\mathbb{C}_1$)	6,566
Crawl 2 ($\mathbb{C}_2$)	9,184
Crawl 3 ($\mathbb{C}_3$)	18,897

- The initial seed contained *94,782* apps and was curated from the lists of apps obtained from approximately *10,000* smartphone users (*Appbrain*).
  - Functionally similar apps
  - Other apps by the same developer
- The subset of apps that were unavailable at the time of this second crawl is referred to as ***Crawl 1* - *C_1***. This process was repeated two times, ***Crawl 2 - C_2*** and ***Crawl 3 - C_3***.
- Only *85 (0.13%)* apps identified as removed in ***Crawl 1*** reappeared in ***Crawl 2***. Similarly, only *153 (0.02%)* apps identified as removed in ***Crawl 2*** reappeared in ***Crawl 3***.

App Labelling Process

- They identified *9 key reasons.*
- They formulated *a set of heuristic checkpoints.*
- They took a random sample of *1500* apps.
- They asked *3* independent reviewers to identify the highest likely reason.
- The reviewers were *Android app developers* and worked full time for *1.5 months at NICTA* for this task. The manual labelling processing took approximately *20 working days* (7 hours per day).

Table 2: Key reasons for removal of apps

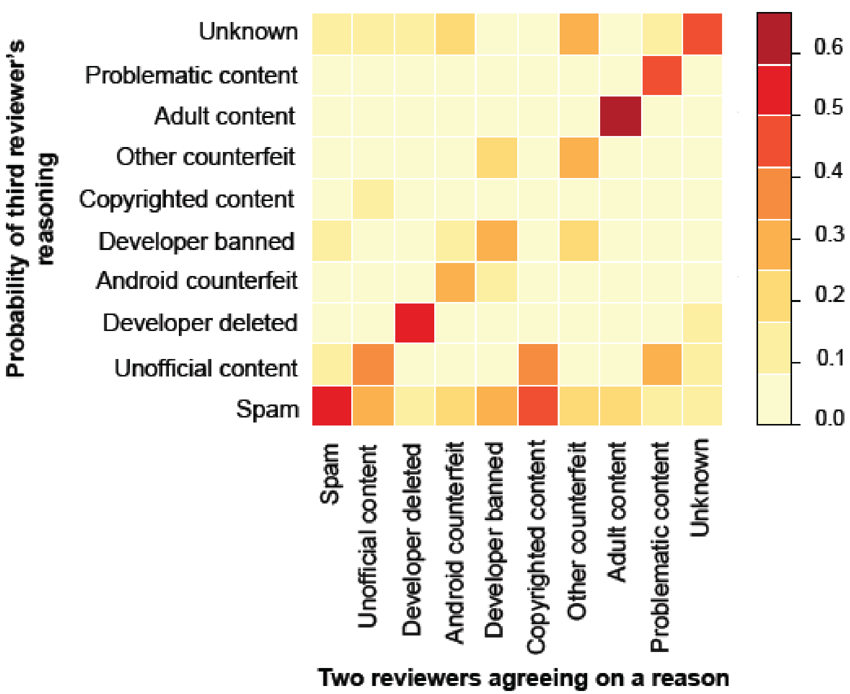
Reason	Description
Spam	These apps often have characteristics such as unrelated description, keyword misuse, and multiple instances of the same app. Section 4 presents details on spam app characteristics.
Unofficial content	Apps that provide unofficial interfaces to popular websites or services (E.g., an app providing an interface to a popular online shopping site without any official affiliation).
Copyrighted content	Apps illegally distributing copyrighted content.
Adult content	Apps with explicit sexual content.
Problematic content	Apps with illegal or problematic content. E.g., Hate speech and drug related.
Android counterfeit	Apps pretending to be another popular app in the Google Play Store.
Other counterfeit	A counterfeit app, for which the original app comes from a different source than Google Play Store (E.g., Apple App Store)
Developer deleted	Apps that were removed by the developer.
Developer banned	Developer's other apps were removed due to various reasons and Google decides to ban the developer. Thus all of his apps get removed.

Agreement Among the Reviewers

- For approximately *40% (601 out of 1500)* of labelled apps, the three reviewers reached a consensus on the reason for removal.
- For *90% (1350 out of 1500)* of the apps majority of the reviewers agreed on the same reason.
- Spam is the main reason for app removal, accounting for *approximately 37%* of the removals.

Table 3: Reviewer agreement in labelling reason for removal

Reason	3 reviewers agreed	2 reviewers agreed	Total	Percent. (%)
Spam	292	259	551	36.73%
Unofficial content	65	127	192	12.80%
Developer deleted	68	56	124	8.27%
Android counterfeit	27	61	88	5.87%
Developer banned	24	54	78	5.20%
Copyrighted content	2	34	36	2.40%
Other counterfeit	11	23	34	2.27%
Adult content	8	4	12	0.80%
Problematic content	3	4	7	0.47%
Unknown	101	127	228	15.20%
Sub total	**601**	**749**	**1350**	**90.00%**
Reviewer disagreement	NA	NA	150	10.00%
Total Labelled	**NA**	**NA**	**1500**	**100.00%**



- There is over *50%* probability of the third reviewer's judgment of an app being spam, when two reviewers already judged the app to be spam.
- Other reasons showing such high probability are *developer deleted* and *adult content* apps.

Spam Checkpoints (I)

Attribute	ID	Description and Examples
Description	S_1	Does the app description describe the app function clearly and concisely? ***E.g. Signature Capture App (Non - spam) - Description is clear on the functionality of the application*** *"SignIt app allows a user to sign and take notes which can be saved and shared instantly in email or social media."* ***E.g. Manchester (Spam) - Description contains details about Manchester United Football Club without any detail on the functionality of the app.*** *"Manchester United Football Club is an English professional football club, based in Old Trafford, Greater Manchester, that plays in the Premier League. ..... In 1998-99, the club won a continental treble of the Premier League, the FA Cup and the UEFA Champions League, an unprecedented feat for an English club."*
	S_2	Does the app description contain too much details / incoherent text / unrelated text for an app description? ***E.g. SpeedMoto (Non - spam) - Description is clear and concise about the functionality and usage.*** *"SpeedMoto is a 3d moto racing game with simple control and excellent graphic effect. Just swap your phone to control moto direction. Tap the screen to accelerate the moto. In this game you can ride the motorcycle shuttle in outskirts, forest, snow mountain,bridge. More and More maps and motos will coming soon"* ***E.g. Ferrari Wallpapers HD (Spam) - Description starts mentioning app as a wallpaper. However, then it goes into to details about Ferrari.*** *"*HD WALLPAPERS *EASY TO SAVE *EASY TO SET WALLPAPER *INCLUDES ADS FOR ESTABLISHING THIS APP FREE TO YOU THANKS FOR UNDERSTANDING AND DOWNLOADING =) Ferrari S.p.A. is an Italian sports car manufacturer based in Maranello, Italy. Founded by Enzo Ferrari in 1929, as Scuderia Ferrari, the company sponsored drivers and manufactured race cars before moving into production of street-legal ....."*
	S_3	Does the app description contain a noticeable repetition of words or keywords? ***E.g. English Chinese Dictionary - Keywords do not have excessive repetition.*** *"Keywords: ec, dict, translator, learn, translate, lesson, course, grammar, phrases, vocabulary, translation, dict"* ***E.g. Best live HD TV no ads (Spam) - Excessive repetition of words.*** *"Keywords: live tv for free mobile tv tuner tv mobile android tv on line windows mobile tv verizon mobile tv tv streaming live tv for mobile phone mobile tv stream mobile tv phone mobile phone tv rogers mobile tv live mobile tv channels sony mobile tv free download mobile tv dstv mobile tv mobile tv....."*
	S_4	Does the app description contain unrelated keywords or references? ***E.g. FD Call Assistant Free (Non - spam) - All the keywords are related to the fire department.*** *"Keywords: firefighter, fire department, emergency, police, ems, mapping, dispatch, 911"* ***E.g. Diamond Eggs (Spam) - Reference to popular games Bejeweled Blitz and Diamond Blast without any reason.*** *"Keywords : bejeweled, bejeweled blitz, gems twist, enjoy games, brain games, diamond, diamond blast, diamond cash, diamond gems, Eggs, jewels, jewels star"*

Spam Checkpoints (Ⅱ)

	S_5	Does the app description contain excessive references for other applications from the same developer? ***E.g. Kids Puzzles (Non - spam) - Description does not contain references to developer's other apps.*** *"This kids game has 55 puzzles. Easy to build puzzles. Shapes Animals Nature and more... With sound telling your child what the puzzle is. Will be adding new puzzles very soon. Keywords: kids, puzzle, puzzles, toddler, fun"* ***E.g. Diamond Snaker (Spam) - Excessive references to developer's other applications.*** *"If you like it, you can try our other apps (Money Exchange, Color Blocks, Chinese Chess Puzzel, PTT Web ....."*
	S_6	Does the developer have multiple apps with approximately the same description? ***The developer "Universal App" has 16 apps having the following description, with each time XXXX term is replaced with some other term.*** *"Get XXXX live wallpaper on your devices! Download the free XXXX live wallpaper featuring amazing animation. Now with "Water Droplet", "Photo Cube", "3D Photo Gallery" effect! Touch or tap the screen to add water drops on your home screen! Touch the top right corner of the screen to customise the wallpaper <>. To Use: Home -> Menu -> Wallpaper -> Live Wallpaper -> XXXX 3D Live Wallpaper To develop more free great live wallpapers, we have implemented some ads in settings. Advertisement can support our develop more free great live wallpapers. This live wallpaper has been tested on latest devices such as Samsung Galaxy S3 and Galaxy Nexus. Please contact us if your device is not supported. Note: If your wallpaper resets to default after reboot, you will need put the app on phone instead of SD card. "*
Identifier	S_7	Does the app identifier make sense and have some relevance to the functionality of the application or does it look like auto generated? ***E.g. Angry Birds Seasons & Candy Crush Saga (Non - spam) - Identifier give an idea about the app.*** *"com.rovio.angrybirdsseasons", "com.king.candycrushsaga"* ***E.g. Game of Thrones FREE Fan App & How To Draw Graffiti (Spam) - Identifiers appear to be auto generated.*** *"com.a121828451851a959009786c1a.a10023846a", "com.a13106102865265262e503a24a.a13796080a"*
Reviews	S_8	Do users complain about app being spammy in reviews? ***E.g. Yoga Trainer & Fitness & Google Play History Cleaner (Spam) - Users complain about app being spammy.*** *"Junk spam app Avoid", "More like a spam trojan! Download if you like, but this is straight garbage!!"*
Adware	S_9	Do the online APK checking tools highlight app having excessive advertising? ***E.g. Biceps & Triceps Workouts*** *"AVG threat labs" [25] gives a warning about inclusion of malware causing excessive advertising.*

Reviewer Agreement on Spam Checkpoints

- The table also suggests that checkpoints *S1, S2, S3* and *S6* are the most dominant checkpoints.

Table 5: Checkpoint-wise reviewer agreement for spam

Checkpoint	S1	S2	S3	S4	S5	S6	S7	S8	S9
3 reviewers agreed	22	63	20	0	4	89	11	3	3
2 reviewers agreed	52	81	75	3	6	115	11	26	15
Disagreed	45								

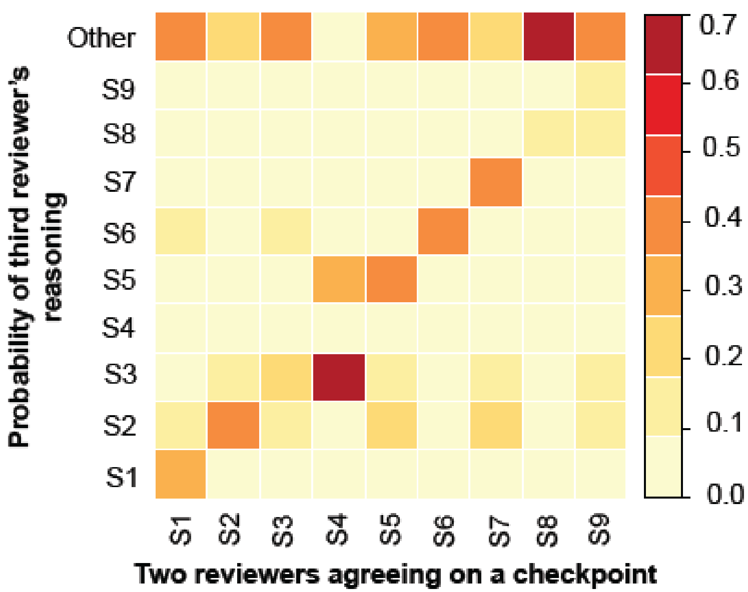


Figure 3: Probability of a third reviewer's judgement when two reviewers already agreed on a checkpoint

- We observe that for checkpoints *S1, S2, S5, S6, S7 and S8*, there is a high probability that the third reviewer indicates the same checkpoint when two of the reviewers already agreed on a checkpoint.

- There is, however, a noticeable anomaly for checkpoint *S4.*

Feature Mapping

- They assumed that top apps are quite likely to be ***non-spam***.
- For non-spam apps, they selected top ***k*** times the number of labelled spam apps (551) from the set ***O***, except all removed apps.
- They varied ***k*** logarithmically between ***1 and 32***, (i.e., 1x, 2x, 4x,..., 32x) to obtain ***6*** datasets of non-spam apps.

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

- They manually read the descriptions of *top 100 non-spam apps* and identified *100 word*
at describe app

one or more
whereas *80% of*
ore of these bi-

ıram and word-
ɔ frequency of
occurrence as the value of the feature.

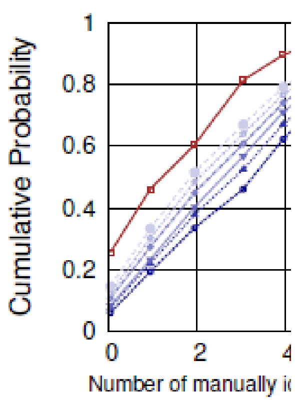


Figure 4: CDF of the
identified word-bigran

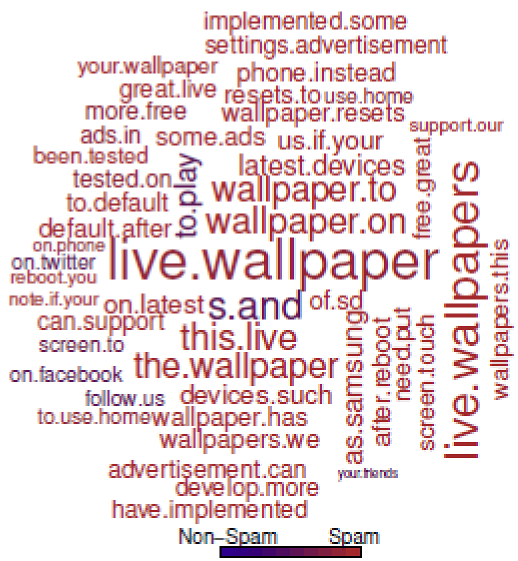


Figure 6: Top-50 bigrams differentiating Spam and Top-1x

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

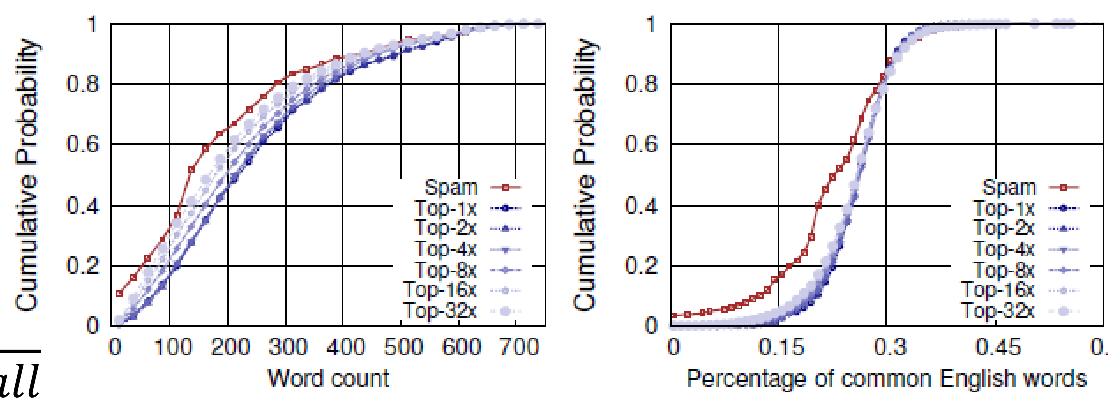
Table 6: Features associated with **Checkpoint** S_2

	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	Percentages of numeric characters
9	Percentage of non-alphabet characters
10	Percentage of common English words [8]
11	Percentage of personal pronouns [8]
12	Percentage of emotional words [39]
13	Percentage of misspelled words [58]
14	Percentage of words with alphabet and numeric characters
15	Automatic readability index (AR) [54]
16	Flesch readability score (FR) [17]

- Nearly ***30% of the spam apps*** have less than 100 words whereas approximately ***only 15% top-1x apps and top-2x apps*** have less than 100 words.

- Spam apps typically use ***fewer common English words*** compared to non-spam apps

$$F_\beta = (1+\beta^2)\,\frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$$



(a) (b)

Figure 5: Example features associated with **Checkpoint** S_2

Checkpoint S3 - Does the app description contain a noticeable repetition of words or keywords?

$$VR = \frac{Number\ of\ unique\ words}{Total\ number\ of\ words}$$

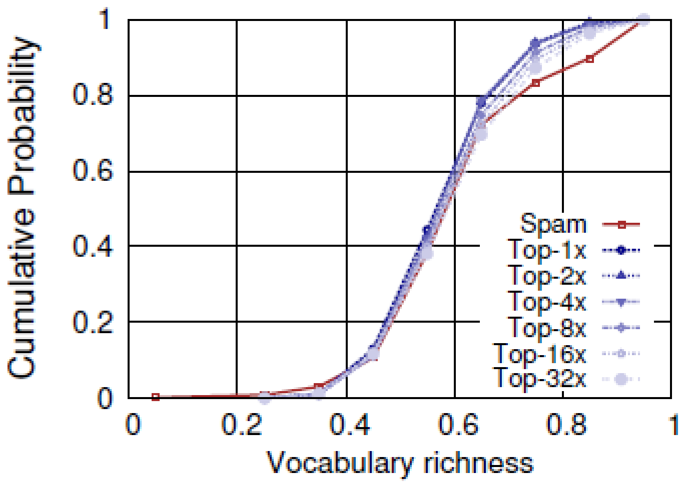


Figure 7: Vocabulary Richness

- *They expected spam apps to have low VR* due to repetition of keywords. However, we observe this only in a limited number of cases.
- If *VR is less than 0.3*, an app is only marginally more likely to be spam. Perhaps the most surprising finding is that the apps with *VR close to 1* are more likely to be spam.
- *10% of the spam apps had VR over 0.9* and none of the non-spam apps had such high VR values.

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

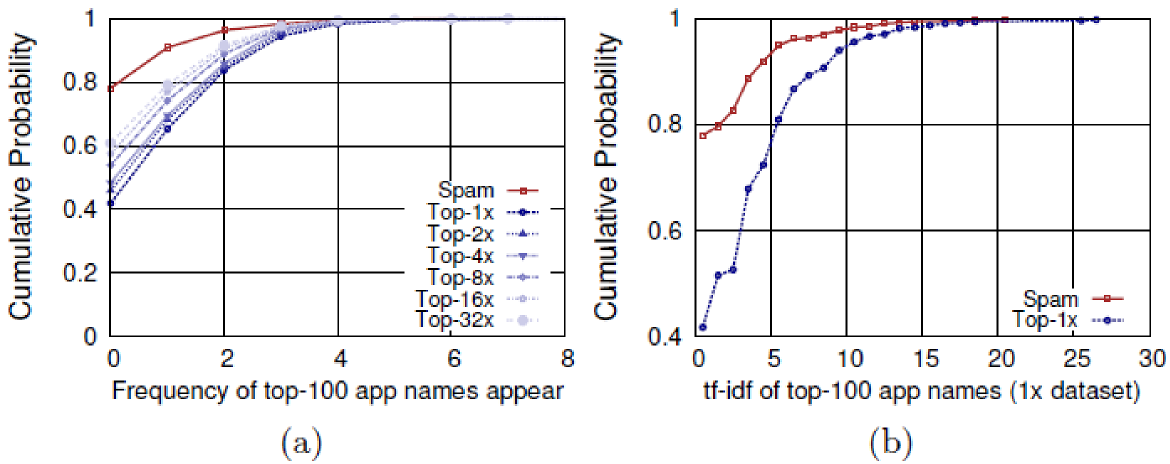


Figure 8: Mentioning popular app names

- For each app they calculated the number of mentioning of the ***top 100 popular app names*** in the app's description.

- They found that ***only 20% of the spam apps*** had more than one mention of popular apps, whereas ***40%-60% of the top-kx apps*** had more than a single mention of popular apps.

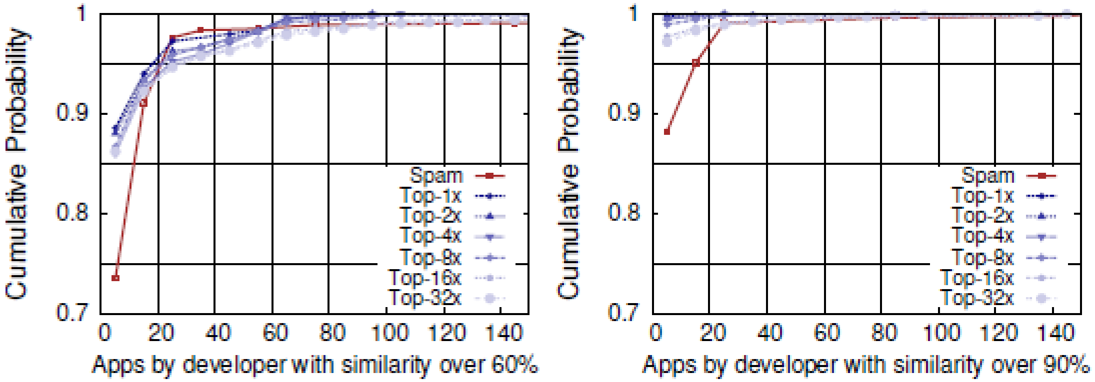
- They used the sum of ***tf-idf*** weights as a feature.

- ***If popular app names are found in the app description, the app tends to be non spam rather than spam.***

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

- They use *the number of times a developer's other app names* appear as the feature corresponding to this checkpoint.
- However, none of the cases marked by the reviewers as matching checkpoint S5 satisfied this feature because the description contained links to the applications rather than the app names and only 10 spam apps satisfied this feature.
- They *do not use checkpoint S5* in their classifier.

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



(a) Over 60% similarity (b) Over 90% similarity

Figure 9: Similarity with developer's other apps

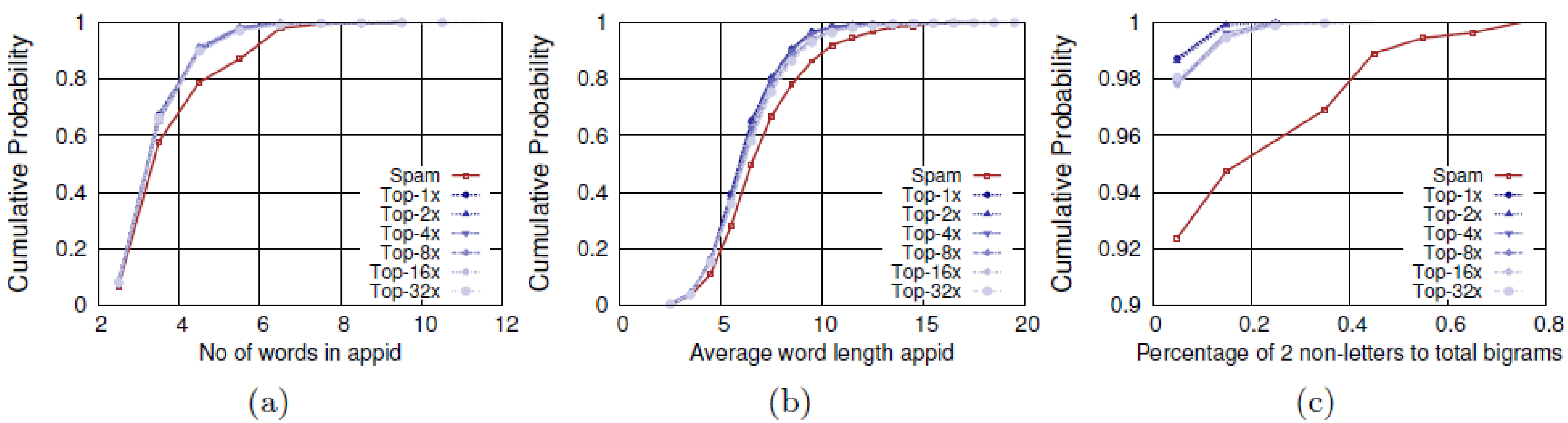
- They considered the following features:
  - *The total number of other apps the developer has.*
  - *The total number of apps with an English language description* which can be used to measure descriptions similarity.
  - *The number of other apps from the same developer* having a description cosine similarity, of over 60%, 70%, 80% and 90%.
- Only about *10%-15% of the non-spam apps* have more than 5 other apps from the same developer with over 60% of description similarity. However, *approximately 27% of the spam apps* have more than 5 apps with over 60% of description similarity.
- *Spam apps tend to have multiple clones with similar app descriptions.*

Checkpoint S7 - Does the app identifier (appid) make sense and have some relevance to the functionality of the application or does it appear to be auto generated?

- ***15% of the spam apps*** had more than 5 words in the appid where as ***only 5% of the non-spam*** had the same.
- For ***10% of the spam apps*** the average word length is higher than ***10*** and it was so ***only for 2%-3% of the non-spam apps***.
- ***None of the non-spam apps*** had more 20% of non-letter bigrams in the appid, whereas ***about 5% of the spam apps*** had more than 20% of non-letter bigrams.

Table 7: Features associated with **Checkpoint** S_7

	Feature
1	Number of characters
2	Number of words
3	Average word length
4	Percentage of of non-letter characters to total characters
5	Percentage of upper case characters to total letter characters
6	Presence of parts of *app name* in *appid*
7	Percentage of bigrams with 1 non-letter to total bigrams
8	Percentage of bigrams with 2 non-letters to total bigrams
9	Percentage of bigrams with 1 or 2 non-letters to total bigrams
10	Percentage of trigrams with 1 non-letter to total trigrams
11	Percentage of trigrams with 2 non-letters to total trigrams
12	Percentage of trigrams with 3 non-letter to total trigrams
13	Percentage of trigrams with 1, 2 or 3 non-letters to total trigrams



(a) (b) (c)

Figure 10: Example features associated with **Checkpoint** S_7

Other Metadata

- They note that *approximately 42%* of the spam apps belong to the categories *Personalisation* and *Entertainment.*
- *40% of the spam apps* had more than 25 characters in the app name. *Only 20% of the non-spam apps* had more than 25 characters in their app names.
- *30% of the top-kx apps* were less than 100KB in size and the corresponding percentage of *spam apps is almost zero*. Almost *all the spam apps* were having sizes less than 30MB whereas *10%-15% of the top-kx apps* were more than 30MB in size.
- If *a link to a developer web site or a privacy policy* is given the app is more likely to be non-spam.

Table 8: Features associated with other app metadata

	Feature		Feature
1	App category	5	Developer's website available
2	Price	6	Developer's website reachable
3	Length of app name	7	Developer's email available
4	Size of the app (KB)	8	Privacy policy available

Table 9: Availability of developer's external information

	Spam	top 1x	top 2x	top 4x	top 8x	top 16x	top 32x
Website availa.	57%	93%	94%	93%	91%	89%	86%
Website reacha.	93%	98%	97%	97%	96%	96%	95%
Email availa.	99%	84%	89%	91%	93%	94%	95%
Priv. policy availa.	9%	56%	50%	48%	38%	32%	26%

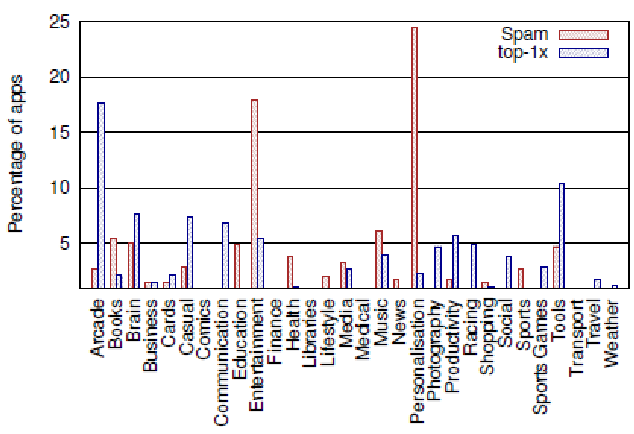
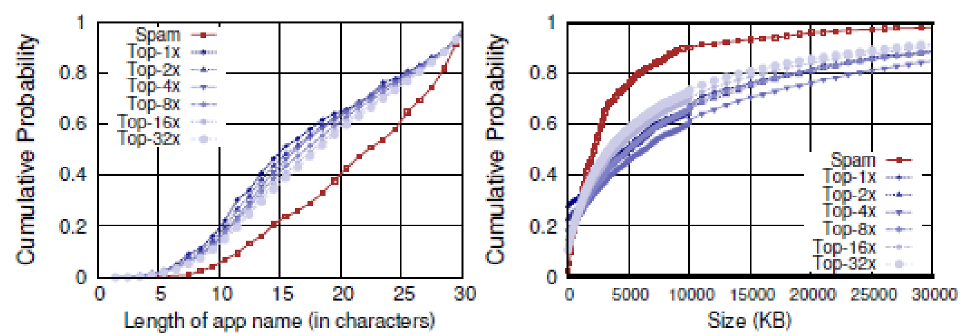


Figure 11: App category



(a) (b)

Figure 12: Features associated with other app metadata

Outline

- Background
- Motivation
- Early Detection of Spam Mobile Apps
- Evaluation
- Conclusion

Evaluation

- The ***Adaptive Boost classifier*** is trained using ***80%*** of the data and the remaining ***20%*** of the data is used for testing.
- The classifiers, while varying the value of ***k***, have ***precision over 85%*** with ***recall varying between 38%-98%***. Notably, when k is small (e.g., when the total number of non spam apps represents 2x the number of spam apps) the classier achieves up to ***95%*** accuracy.
- Table 11 show the ***k = 2*** classifier's performance in the higher order datasets. As can be seen, this classifier identifies ***nearly 90% of the spam***.
- More aggressive classier (k = 2) predicted around ***70%*** of the removed apps and ***55%*** of the other apps to be spam. The conservative classier (k = 32) predicted ***6%-12%*** of the removed apps and ***approximately 2.7%*** of the other apps as spam.

Table 10: Classifier Performance

k	Precision	Recall	Accuracy	$F_{0.5}$
1	0.9310	0.9818	0.9545	0.9408
2	**0.9533**	**0.9273**	**0.9606**	**0.9480**
4	0.9126	0.8545	0.9545	0.9004
8	0.9405	0.7182	0.9636	0.8857
16	0.8833	0.4818	0.9658	0.7571
32	0.8571	0.3818	0.9793	0.6863

Table 11: Classifier Performance: $k = 2$ model

k	Precision	Recall	Accuracy	$F_{0.5}$
4	0.8080	0.9182	0.9400	0.8279
8	0.5549	0.9182	0.9091	0.6026
16	0.2730	0.9182	0.8513	0.3176
32	0.1164	0.9182	0.7862	0.1410

Table 12: Predictions on spam apps in Google Play Store

Dataset	Size	$k = 2$	$k = 32$
Crawl 1 ($\mathbb{C}_1$)	6,566	70.37 %	12.89 %
Crawl 2 ($\mathbb{C}_2$)	9,184	73.14 %	6.57 %
Crawl 3 ($\mathbb{C}_3$)	18,897	72.99 %	6.49 %
Others	180,627	54.59 %	2.69 %

Outline

- Background
- Motivation
- Early Detection of Spam Mobile Apps
- Evaluation
- Conclusion

Conclusion

- The authors develop a manual app classification methodology based on a set of heuristic checkpoints that can be used to identify reasons behind an app's removal. They found that approximately 35% of the apps that were removed are spam apps.

- The authors present a mapping of their proposed spam checkpoints to one or more quantifiable features that can be used to train a learning model.

- The authors build an Adaptive Boost classifier for early detection of spam apps and show that their classifier can achieve an accuracy over 95% at a precision between 85%-95% and a recall between 38%-98%.

- The authors applied our classier to over 180,000 apps available in Google Play Store and show that approximately 2.7% of them are potentially spam apps.

Thank you!