Summary Review Documentation for

“Crime Scene Investigation: SMS spam data analysis”

Authors: I. Murynets, R. Jove

Reviewer #1

Summary: This paper investigates the activities of SMS spam accounts in a tier-1 cellular network in the US. In order to remove any false positives, the authors focused on the banned accounts listed by the fraud department of the network operator. When compared to legitimate users, SMS spammers send orders of magnitude larger number of SMSs to a wide range of recipients. They try to mimic the behaviors of legitimate users by sending SMSs during daytime and also showing light mobility patterns.

Overall, the dataset is one-of-a-kind and findings of this paper are very interesting. I learned a lot of new things about SMSs from this paper. Given the rich dataset, authors could pursue interesting future directions on the social network structure of SMS spammers based on the SMS logs or call logs.

Strengths: The dataset is impressive. The paper is well-structured and provides interesting insights about SMS spammers.

Weaknesses: Analysis is not assisted with proper statistics. Observations are made mostly from graph visualization, not from statistical tests (e.g., t-test).

Comments to authors: How are the legitimate users chosen? What are the maximum values in Figure 1(a)? Based on the later result (Fig 8), some legitimate users seem to post up to 20,000 SMSs a day, which is very surprising. Given 86400 seconds a day, this behavior seems suspicious. In Section 3.2: Is opting-out also considered as a valid response? In Section 3.3.2: The mobility pattern of spammers is very interesting. Based on these patterns, are you able to find spammers who carry multiple devices? In Section 3.5.4: To whom these spammers make phone calls? Are they calling to legitimate users or to corporate numbers? Do they actually converse with others? Please show the mean values in all figures (or in text).

Reviewer #2

Summary: This paper presents the first large-scale study of SMS spam based on call data records of a "tier-1 cellular operator" (presumably AT&T’s cellular network). The paper characterizes the methods and sending behavior of spammers versus legitimate senders. The paper finds (perhaps unsurprisingly) that spammers have a higher SMS volume and a lower interarrival time between SMSes (as well as lower entropy of interarrival time) than legitimate senders do.

Strengths: This is the first paper to address the topic of SMS spam in any level of depth. The data sets that the paper presents are comprehensive and interesting, and many of the findings about the behavior of SMS spammers, as well as the infrastructure that they use to send their messages, is very interesting. This is a very interesting first peek into the behavior of SMS spammers.

Weaknesses: Although there is not a lot of related work in SMS spam, there is plenty of related work in email spam—much of it that looks at email sending behavior—that should probably be cited in this paper. In general, the coverage of related work is poor. None of the results are particularly surprising, and, since the data set is unlikely to be released, it will be very difficult for researchers to do any type of follow-up work on this study. Many of the graphs are "mere data reporting" and are light on insights and content. Comments to Authors: The dataset in this paper is truly impressive, and the analysis that the paper conducts is also interesting. I found some of the features of SMS spammer behavior, such as the inter-arrival times, volumes, and the variance in inter-arrival times, to be both intuitive and interesting.

The paper reminds me a lot of the work of Ramachandran et al. from SIGCOMM 2006: "Understanding the Network Level Behavior of Spammers". That paper is on email spam, not on SMS spam, but it also looks at how the email sending behavior of spammers differs fundamentally from that of normal email senders. This paper would be worth citing. Another relevant paper is "#bias: Measuring the Tweeting Behavior of Propagandists", by Lumezanu et al. in ICWSM 2012. That paper essentially looks at the tweet volumes, inter-arrival times, etc. of tweet spam and actually has some very similar observations to those in this paper.

In general, the coverage of related work is poor, and the paper needs a related work section. Although there is certainly not a lot of related work on SMS spam, there is work in other areas—email spam, VoIP spam, etc.—that could certainly be compared to this work. Although the conduit for the spam messages is different in this case, many of the observations about behavior are likely to carry over from one domain to another.

I found Figure 3 somewhat puzzling: shouldn’t the message ratios for spammers be much more imbalanced than those for legitimate senders? I didn’t understand why SMS spammers would be receiving so many messages. Also, I did not get much out of these time series plots: there appears to be no cycles, monthly patterns, diurnal trends, etc., and the patterns certainly don’t differ much from those of the legitimate senders.

Figure 5 is interesting, but it is just an example. Would it be possible to represent this trend more generally, for a larger number of senders? Perhaps there is a way to characterize the burstiness of each time series and then compare the general burstiness of spammer behavior to that of legitimate senders.

I wasn’t sure what conclusions I was supposed to draw from the location analysis of the SMS spammers. Why is this interesting,
and why does it matter? Also, the Google maps don’t really show much interesting. This appears to be "mere data reporting", without any real insights.

**Reviewer #3**

**Summary:** The paper presents a measurement-based analysis of SMS spam traffic from a large US cellular carrier (AT&T). The authors compare the communication patterns with those of non-spammers, and report some interesting findings. For instance, the authors find that the communication patterns of spammers are similar to those of "M2M" systems. The authors also analyze the main geographical sources of spammers and their mobility.

**Strengths:** The paper presents one of the first studies of SMS spam in large cellular networks, and presents some useful and interesting findings. I particularly like the analysis and findings of the geographical locations of spammers and their mobility.

**Weaknesses:** Besides presenting various analyses of the SMS spam traffic, the paper lacks a well-defined and coherent measurement methodology. It reads like a collection of analysis results. While some results are interesting and useful, others are expected. The analysis is often somewhat light-weight, lacks depth. For instance, the authors state (see the conclusion) that "[we] confirm common intuitions about spammers such as the large number of text messages sent per day to a wide target list." However, there is no further investigation as to how the spammers select their "wide target list". Does "wide" imply that their targets are all across the US, even though the spammers tend to be concentrated in certain large states or metro areas?

**Comments to authors:** The analysis of the paper is based on the CDRs of 9000 spammer accounts. However, I could not find any description as to how these spammers are identified in the first place. Are all SMS messages from these accounts spam, or a majority of them? As the authors show that SMS spammers are also receiving many messages, are these messages due to spam targeted users responding to spam, or something else? In fact, as the example in Fig.2 show that one would expect that the spam targeted users would in general be solicited to respond to a different phone number (than those of the spammers) or some websites, etc. This would likely be the case given that many spammers use multiple devices simultaneously to wage the spam campaign.

For comparative analysis, the authors state that they used the CDRs of almost 17000 from "randomly selected legitimate accounts" from the one year period between March 2011 and February 2012. How are the "legitimate" users determined? How are these "legitimate" users "randomly" selected? All these accounts select from all US subscribers? Also, does the time period selected match those of the spammers? Since you discovered that many of the spammers come from certain states or metro areas, should it be more advisable to select "legitimate" users with similar geographical distributions? One would expect that users from different regions (e.g., rural vs. urban) may have different usage patterns and behavior, in particular given that the cellular coverage (esp. data coverage) can vary from region to region.

The authors use the term (embedded) "M2M" systems/appliances throughout the paper, but fail to provide more specific definition/description of these M2M accounts. For instance, how are these "M2M" accounts classified and identified? What kind of services are running on them? For instance, are e-reader devices considered "M2M" appliances? Smart-grid meters? What do they use SMS messages for? For instance, are they used for transmitting control messages or telemetry data? Do these accounts mostly talk to other M2M accounts, or to "regular" users? This is important since one of the main findings of the paper is that "M2M systems exhibit communication profiles similar to spammers." I wish that the authors have expanded on this finding further and perform more in-depth investigation. As the authors have correctly stated, this has important implications in how we design filter algorithms to distinguish spam vs. non-spam messages (at the cellular network provider’s side).

I also wish that the authors performed some analysis of the spam content; again this has implications in terms of spam message filtering. The authors only provide a sample of the spam messages in Figure 2.

All in all, I find that the analysis is often somewhat light-weight, lacks depth. For instance, the authors state (see the conclusion) that "[we] confirm common intuitions about spammers such as the large number of text messages sent per day to a wide target list." However, there is no further investigation as to how the spammers select their "wide target list". Does "wide" imply that their targets are "randomly selected" all across the US, even though the spammers tend to be concentrated in certain large states or metro areas? Understanding the target selection strategies of spammers is important, in particular in light of the authors’ findings that "M2M systems exhibit communication profiles similar to spammers."

**Reviewer #4**

**Summary:** The paper analyzed real SMS spam traffic from tier-1 ISP and compared their characteristics to legitimate SMS messages from cell phone users & M2M connected appliances.

**Strengths:** Valuable set of data containing ground truth for SMS traffic and attributes/information of the SMS spammer (e.g., devices used to launch SPAM and geographical locations).

**Weaknesses:** The paper merely presents statistics on various characteristics (SMS volume, destination, location, etc.) of SMS spammers, as compared to legitimate users & M2M appliances. It’s clear that none of the observations alone would be distinctive enough feature for spam detection, as expected. As a result, no true new insights are gained. Graphs are impossible to read with the overlapping histograms.

**Comments to authors:** While it’s useful to confirm common conjectures about how spammer traffic would be different, the paper doesn’t offer much new insights beyond what we already know from the wire-line spam literature. There are a few interesting nuggets, e.g., devices used to launch spam, and geographical locations, but they do not constitute enough technical contributions.

It would have been more interesting if a spam detector is developed and evaluated. For instance, can all the different features shown here be combined in machine learning to classify/identify spammer? In particular, Figure 12 (b) shows a clear separation between SMS spam and normal traffic in terms of average number of call vs. average number of SMS recipients. Can this ratio be used to detect spammer and how effective is it?
Graphs are hard to read – even though different colors are used, it’s hard to differentiate the different histograms when they overlap. For Figure 4(b) and 8(b), you might want to discuss what values of entropy is considered ‘high’.

Reviewer #5

Summary: The authors analyze a large amount of SMS data from a big cellular operator to characterize the nature of SMS spam: when and how it’s sent, from whom, and to whom. They find evidence that such spam is predominantly sent from only a few locations and with a small number of distinct types of devices.

Strengths: A novel topic tackled with juicy-scale data. The findings about prevalence, location, and patterns of activity (such as the spamming devices also place voice calls) are quite interesting.

Weaknesses: It’s hard to tell just how sound the results are. There are some potential biases (particularly with regard to ground truth) that the paper doesn’t identify or explore, as I note below. In addition, the presentation is frustrating. The color figures don’t work in black-and-white, and some of the discussion comes across as unnecessary detail, such as section 3.5.2 and the accompanying figure 13.

Comments to authors: While I realize this might be tricky, it’s really important to explain something about how the fraud department flags spammers, because that goes to the basic nature of the ground truth data. For example, suppose that for some reason the department primarily focuses on particular regions around the country - that could explain the location results in 3.3.1 as being due to the nature of the detection rather than something fundamental about the spammers themselves. And/or if account age is part of what the department uses, then that would provide an alternative interpretation for the finding in 3.1 that the average age of an illegitimate account is just a few days.

The analysis should be careful to take into account the nature of random sampling. Due to skew in cellular use, a random sample of legitimate accounts provides an unbiased look at how legitimate accounts behave, but does not illuminate rarely-seen-but-eligible accounts that might behave a lot like spammers do. For example, if only 1 in 50,000 legitimate accounts is such a peculiar high-volume account, then it might not occur at all in your random sample.

Please clarify just what sort of devices and usage the M2M traffic reflects. This is discussed only later in the paper. Also, does an "approved M2M device" mean that the hardware is of a type that the operator has approved, or that the user/account is a known M2M service that the operator condones?

I was surprised that there wasn’t discussion of companies that occasionally disseminate large numbers of texts in a short period of time, for example an airline announcing a bunch of canceled flights. Does this show up in the data, and, if so, how do you avoid identifying it as spam? Similarly, if there are Twitter-to-SMS gateways then I’d expect some popular tweeters to seem to appear as spammers, unless these are specially flagged.

I find the title misleading. The phrase "crime scene investigation" suggests a forensic analysis, not a measurement characterization. What proportion of non-spammer accounts is prepaid? I wasn’t clear on how spammers avoid being tracked based on IMEI.

The caption in figure 3 should explicitly mention the normalization; otherwise the figures look implausible (legitimate customers issuing way fewer than one SMS per day). In figure 3, does "cellphone user" include M2M? How representative is the behavior in figure 5? Given it’s a single random sample; I can’t tell whether there might be a wide variation in the patterns across different spammers. In figure 8(d), over what time period is the average computed?

Clarify whether references to entropy equate to the usual definition of H or something else. Clarify what "lacci" refers to in the figure 9 X axis label. It would be nice if the most frequently used devices could be explicitly identified or at least we could be told why they aren’t. The text discusses figure 12a as having the average number of call destinations on the X axis, but instead it’s the average number of calls per day. Also, it’s a log-log plot, not log-linear.

You could analyze the relative timing of replies to spammers to get a sense of how many of them likely reflect some sort of response to a received spam, and in general what sort of delays occur in doing so. In particular, any text received prior to someone being spammed can’t be due to the spam.

Response from the Authors

We would like to start by thanking our anonymous reviewers for their insightful comments which have assisted us in improving the paper for the camera-ready version. As a result, the paper is much better in terms of clarity and content.

We have included a Related Work section (Section 5) discussing existing work on SMS spam detection as well as spam data analysis in the context of YouTube comments, tweets and the blogosphere.

With regard to Section 2 (Data Set), some of the reviewers asked us to provide a better explanation on how each data set is obtained. We have rewritten that section to explain how each set (legitimate customers, M2M and spammers) was obtained. Specifically, paragraphs 4 and 5 in this section provide a detailed description of each data set. Note that the spammer set contains exclusively SMS spam accounts that have been positively identified, checked and canceled. The M2M set is obtained from the list of approved M2M embedded devices of the cellular operator. Details on this list and the kinds of devices and systems included in the system can be found in the 5th paragraph of the section. Unfortunately, the specific procedure followed by the Fraud Department to cancel accounts based on messaging abuse activities cannot be disclosed.

Reviewer 4 observes that it would have been interesting to see a spam detector developed and evaluated. An accurate SMS spam detection engine has been indeed designed and evaluated. Its description, details and results are, however, out of the scope of this paper. Nevertheless, we have introduced a small section (Stopping the crime, Section 4) with some brief details on this algorithm. The actual algorithm, system description and results will be presented in the near future.

Another reviewer comments that it would be very interesting to have the most frequently used devices explicitly identified. Unfortunately, given that all these devices are legitimate devices (in some cases rather well known popular feature phones) we choose to not disclose such information. This way we avoid misconceptions that may argue that these are “spamming devices”. Pictures of the devices will be included in our presentation, though.
Some of the reviewers mention that the figures are difficult to understand in black and white. We have explored some alternatives but the current plots are the most effective ones that preserve the color coding for each category (green—legitimate, blue-M2M and red-spammer) throughout the paper. We understand that IMC papers often contain figures that need to be seen in color for a proper visualization.

Finally, we have made several extra minor adjustments to address the comments from the reviewers. For example, we have clarified section 3.2.1 (Response ratio) to give insights on why spammers receive a large number of text messages replying to the spam they generate. We also included a brief comment on how most spam detection engines access white-list data feeds to avoid wrongly flagging sources of large loads of text messages (i.e., Twitter SMS, American Idol alerts, etc). All the typos and minor mistakes pointed out by the reviewers have been addressed as well.