We collect and study over ***2 million*** phone calls to uncover more than ***1,000 robocalling campaigns*** that are currently operational in the wild.

Who's Calling?

Characterizing Robocalling Campaigns

Sathvik Prasad, Athishay Kiran, Elijah Bouma-Sims, Bradley Reaves

INTRODUCTION

- Automated phone calls, or ***Robocalls***, are a menace with no solution in sight
- We collect data from the real-world telephone network to characterize robocalls
- Our findings highlight the need for stronger and more robust techniques to combat robocalls in modern telephone networks

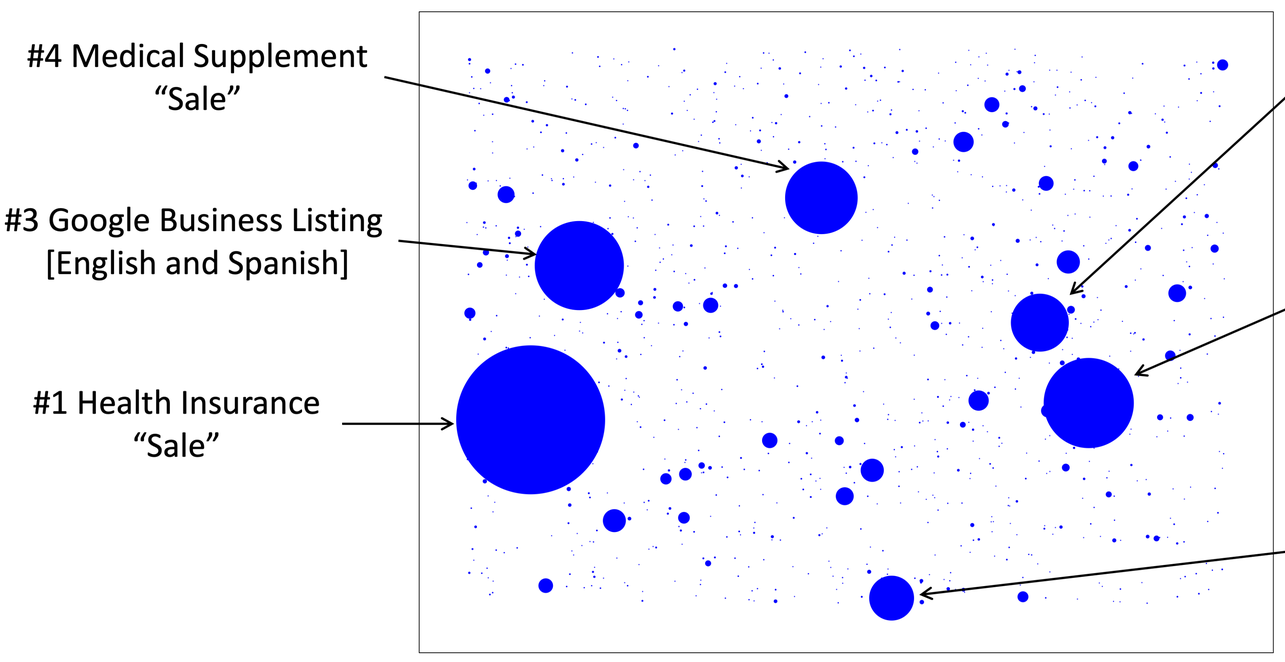
METHODOLOGY

- Using 66,000 inbound telephone lines, we deploy the largest academic telephony honeypot
- We systematically collect call signaling data, Call Detail Records (CDR) and call audio
- We apply robust audio fingerprinting techniques for robocall campaign identification and clustering
- Our approach of campaign clustering using audio fingerprinting is immune to caller ID spoofing, can identify campaigns operating in different languages, and is independent of transcription accuracy

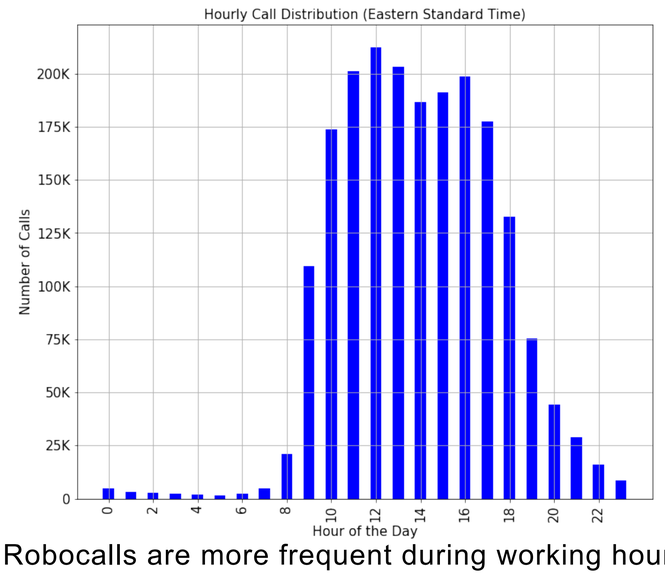
RESULTS

- We identify over 1000 robocalling campaigns that are currently operational
- Our approach identifies campaigns operating in different languages (English, Spanish, Mandarin)
- We uncover a large-scale SSN Fraud campaign, which is the 5th largest campaign in our honeypot
- We observe that caller ID spoofing is rampant among adversarial robocalling campaigns

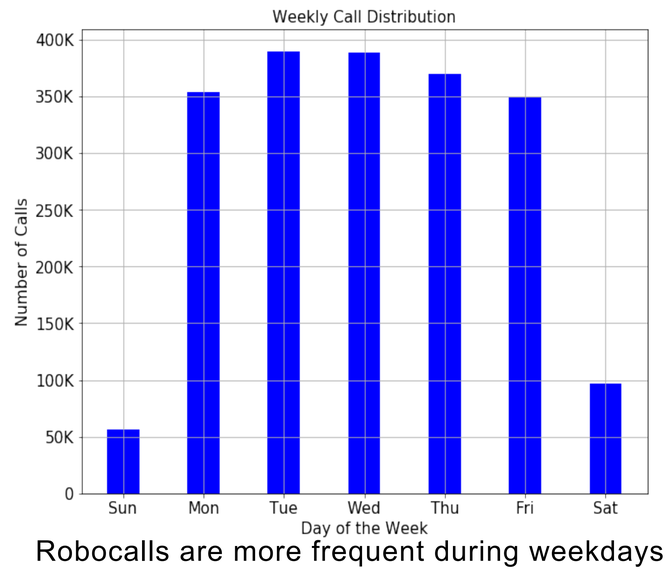
Number of Calls Received	2,003,000 +
Number of Calls Recorded	60,000 +
Duration of Study	10 Months
Number of Inbound Telephone Lines	66,606



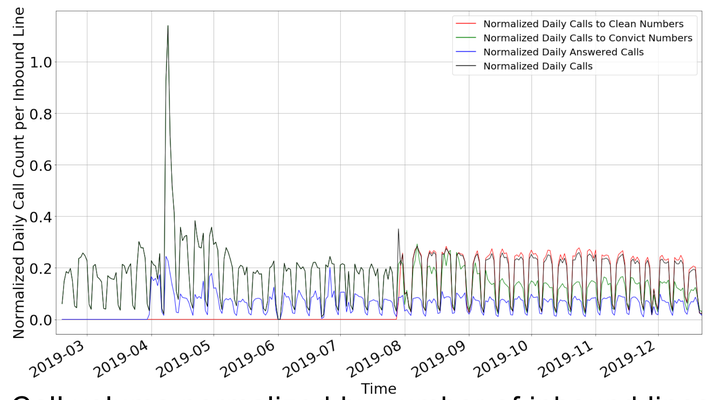
Unsolicited Calling Campaigns (Top 6 Labeled)



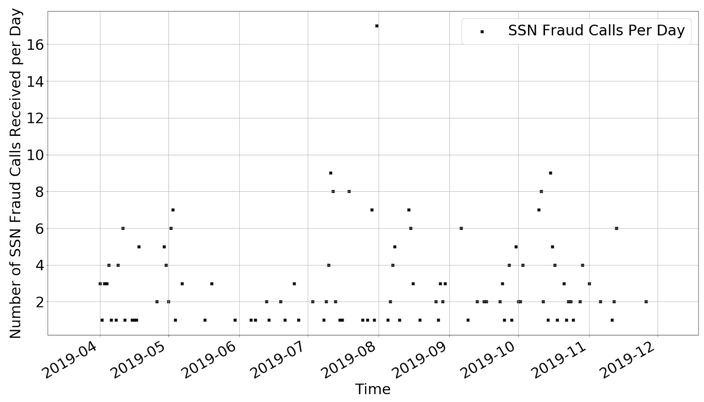
Robocalls are more frequent during working hours



Robocalls are more frequent during weekdays



Call volume normalized by number of inbound lines



Distribution of SSN Fraud calls

NC STATE UNIVERSITY

NSF

WSPR

Poster: Who's Calling? Characterizing Robocalling Campaigns

Sathvik Prasad
snprasad@ncsu.edu
North Carolina State University

Elijah Bouma-Sims
erboumas@ncsu.edu
North Carolina State University

Athishay Kiran
akmylapp@ncsu.edu
North Carolina State University

Bradley Reaves
bgreaves@ncsu.edu
North Carolina State University

I. Extended Abstract

Spam calls and automated phone calls, also called robocalls, are a rampant problem in the United States with no solution in sight. Such a growing problem calls for a thorough understanding of how adversaries operate in the phone network. An in-depth characterization is crucial to develop new and more robust mechanisms to combat adversarial telephone campaigns in the real-world. Unlike the Internet, telephone networks have well-guarded network boundaries which makes it challenging to observe how adversaries operate. As part of our research, we develop and maintain a large-scale academic telephony honeypot [4], [5], [3], [2], [1] to characterize unsolicited phone calls in the modern telephone network. We uncover some of the largest fraudulent campaigns which are currently operational in the real-world and systematically study multiple strategies used by such campaigns to target the users of the phone network.

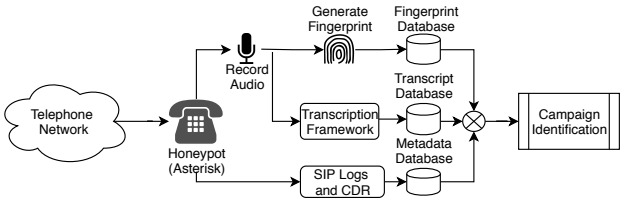


Fig. 1: Data Collection and Data Processing Sequence

As part of this ongoing research project, we deploy a telephony honeypot with over 66,000 phone numbers. A high-level architecture of our honeypot and the associated subsystems are show in Figure 1. Using our honeypot, we collect call signaling information, call audio and Call Detail Records (CDR) of unsolicited calls made to the phone numbers mapped to our honeypot for over 10 months. A study that spanned over multiple months allowed us to observe the temporal characteristics of unsolicited phone calls in the real-world and evaluate the effects of answering such phone calls.

We observe that the call volume of unsolicited calls made to our honeypot is largely stationary. We normalize the number of calls received per day by the number of active inbound lines configured for a specific call treatment – answer a call or reject a call. We track the call volume of various groups of numbers which have different history (new unassigned numbers and previously assigned numbers) separately and the same is shown in Figure 2. We report the scale of potential Voicemail spam or Voicemail injection observed in our honeypot, which is a mechanism for an adversary to reach a user's voicemail with minimal user interaction. We also report the scale of potential Wangiri scams observed in our honeypot.

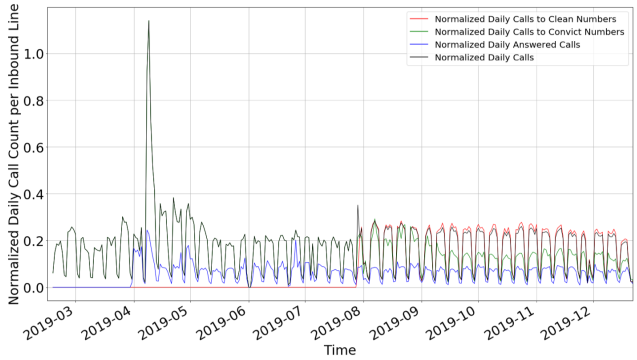


Fig. 2: Stationary Call Volumne Distribution with some Outliers in April 2019

To identify campaigns among the data collected in our honeypot, we develop and apply audio-fingerprinting based clustering techniques. Poor quality audio and audio with background noise are challenging to transcribe accurately. Despite such challenges, our clustering mechanism is capable of identifying campaigns effectively, since we do not depend on the accuracy of transcription. We also identify campaigns which operate in different languages like English, Mandarin and Spanish.

Using a combination of call meta-data, call audio and call signaling information, we identify over 1,000 call campaigns which range from benign advertisements to outright fraud. Our audio-fingerprinting based clustering approach allows us to group similar call audio into a cluster of calls. We further study the operational characteristics of these campaigns and quantify the scale of caller ID spoofing among these campaigns. We observe that such campaigns predominantly operate during working days and working hours, with the intention of tar-

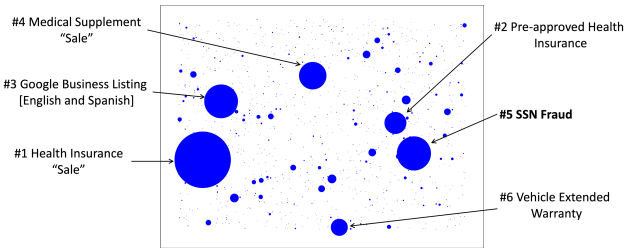


Fig. 3: Relative Size of Phone Campaigns
Top 6 Campaigns Labeled with the Campaign's Intent

geting mobile phone users and numbers owned by businesses. Such behavior is evident in the hourly and weekly call volume distribution shown in Figure 4 and Figure 5. We quantify neighbor-spoofing and other techniques used by adversaries to deceive the end-users. Using statistical techniques, we estimate the number of robocalling campaigns operating in the wild.

Campaign clustering allows us to quantify various techniques used by individual campaigns. Figure 3 shows the relative scale of various campaigns with the top six campaigns highlighted based on the campaign's intent. We study caller ID spoofing and neighbor spoofing techniques employed by these campaigns. After identifying active campaigns in the phone network, we delve into certain campaigns as case studies. Specifically, we look at a large-scale Social Security Number (SSN) fraud campaign and study it's caller ID selection behavior and highlight the temporal characteristics. We also discuss campaigns that use Spanish and Mandarin languages to target specific segments of users. We discuss the effectiveness of existing mitigation techniques like call blacklists, Do Not Call registry, STIR/SHAKEN etc.

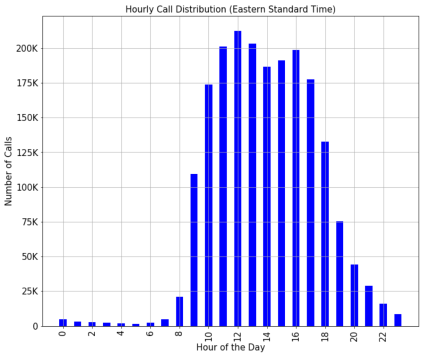


Fig. 4: Over 83% of Unsolicited Calls were Received During Working Hours (Easter Standard Time)

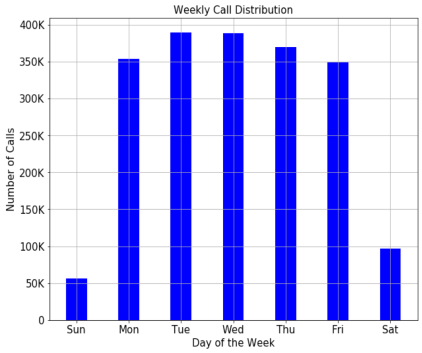


Fig. 5: Over 92% of Unsolicited Calls were Received During Weekdays (Monday-Friday)

REFERENCES

[1] M. Balduzzi, P. Gupta, L. Gu, D. Gao, and M. Ahamad, "Mobipot: Understanding mobile telephony threats with honeycards," in *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, ser. ASIA CCS '16. New York, NY, USA: ACM, 2016, pp. 723–734. [Online]. Available: http://doi.acm.org/10.1145/2897845.2897890

[2] P. Gupta, B. Srinivasan, V. Balasubramaniyan, and M. Ahamad, "Phoneypot: Data-driven understanding of telephony threats," in *22nd Annual Network and Distributed System Security Symposium, NDSS 2015, San Diego, California, USA, February 8-11, 2015*, 2015. [Online]. Available: https://www.ndss-symposium.org/ndss2015/phoneypot-data-driven-understanding-telephony-threats

[3] H. Li, X. Xu, C. Liu, T. Ren, K. Wu, X. Cao, W. Zhang, Y. Yu, and D. Song, "A machine learning approach to prevent malicious calls over telephony networks," in *2018 IEEE Symposium on Security and Privacy (SP)*, May 2018, pp. 53–69.

[4] A. Marzuoli, H. A. Kingravi, D. Dewey, and R. Pienta, "Uncovering the landscape of fraud and spam in the telephony channel," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec 2016, pp. 853–858.

[5] S. Pandit, R. Perdisci, M. Ahamad, and P. Gupta, "Towards measuring the effectiveness of telephony blacklists." Annual Network and Distributed System Security Symposium, NDSS, 01 2018.