Diving into Robocall Content with SnorCall

Sathvik Prasad
PhD Candidate @ NC State University

Diving into Robocall Content with SnorCall
Sathvik Prasad, Trevor Dunlap, Alexander Ross, and Bradley Reaves
North Carolina State University

USENIX Security 2023
https://www.usenix.org/conference/usenixsecurity23/presentation/prasad

Everyone Hates Illegal Robocalls

“Hiker lost for 24 hours ignored rescuers' calls because 'they didn’t recognize the number’“ - NBC NEWS

“..people in their twenties reported losing money to fraud at a higher rate than people in their seventies..” - The FTC May 2023

“..The FTC has stopped a pair of student loan debt relief schemes.. ..bilked students out of approximately $12 million…” - May 2023

https://www.nbcnews.com/news/us-news/hiker-lost-24-hours-ignored-rescuers-calls-because-they-didn-n1282381
https://consumer.ftc.gov/consumer-alerts/2023/05/scam-proof-young-people-your-life
https://www.ftc.gov/news-events/news/press-releases/2023/05/ftc-stops-student-loan-debt-relief-schemes-it-says-bilked-students-out-millions

Why are we still struggling to stop illegal robocalls in 2023?

1. Carriers lack the tools and techniques to combat illegal robocalls at scale

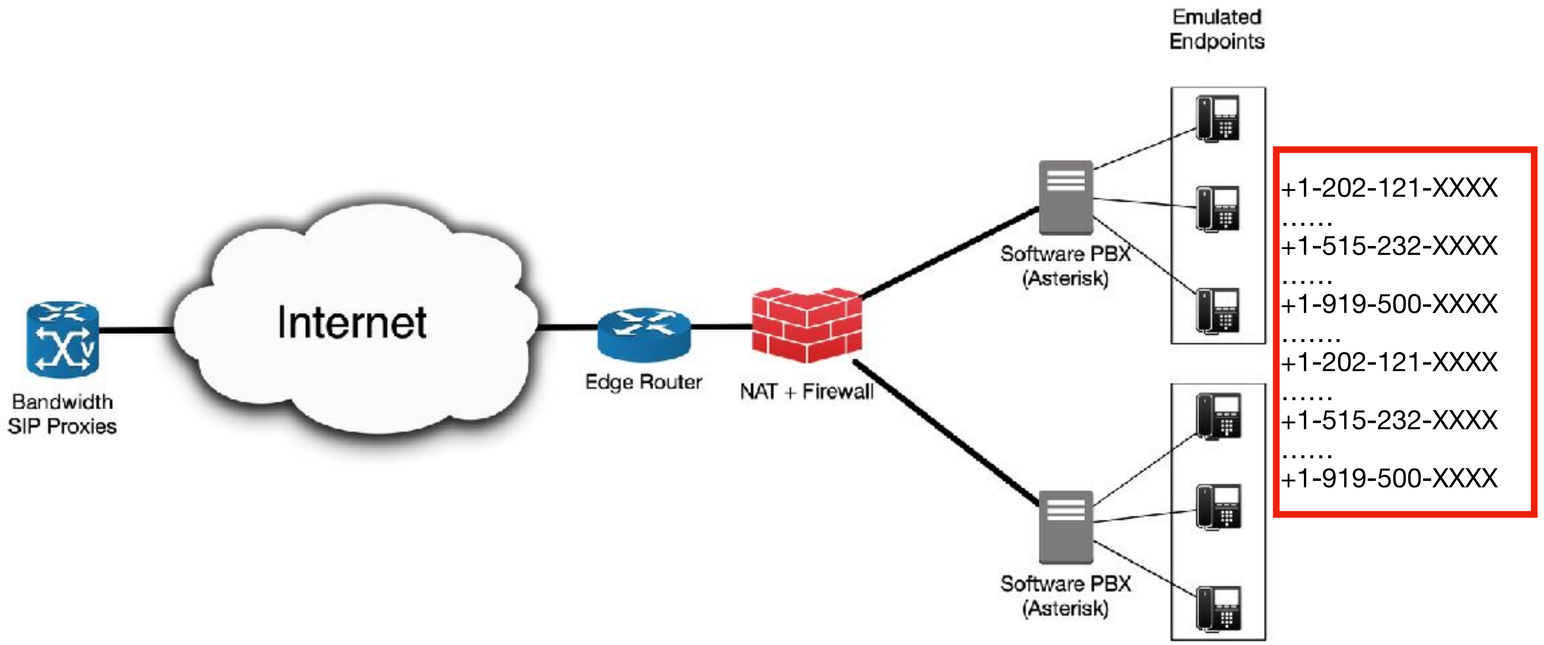
2. Enforcement agencies don't have enough time, human resources, or the tools to take action against every illegal robocalling operation

3. Robocall blocking techniques rely on adversary-controlled metadata

What will you learn from this talk?

1. How Social Security scams and Tech Support scams have changed their tactics?
2. What was the impact of the US Presidential Elections on the robocalling landscape?
3. How do robocallers engage with their targets?

Data Collection using Telephony Honeypot



23 Months

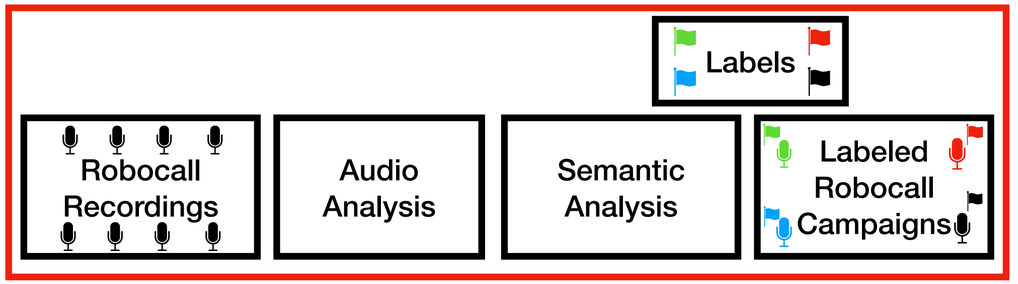
1.3 Million Robocalls

~66k VoIP Phone Numbers

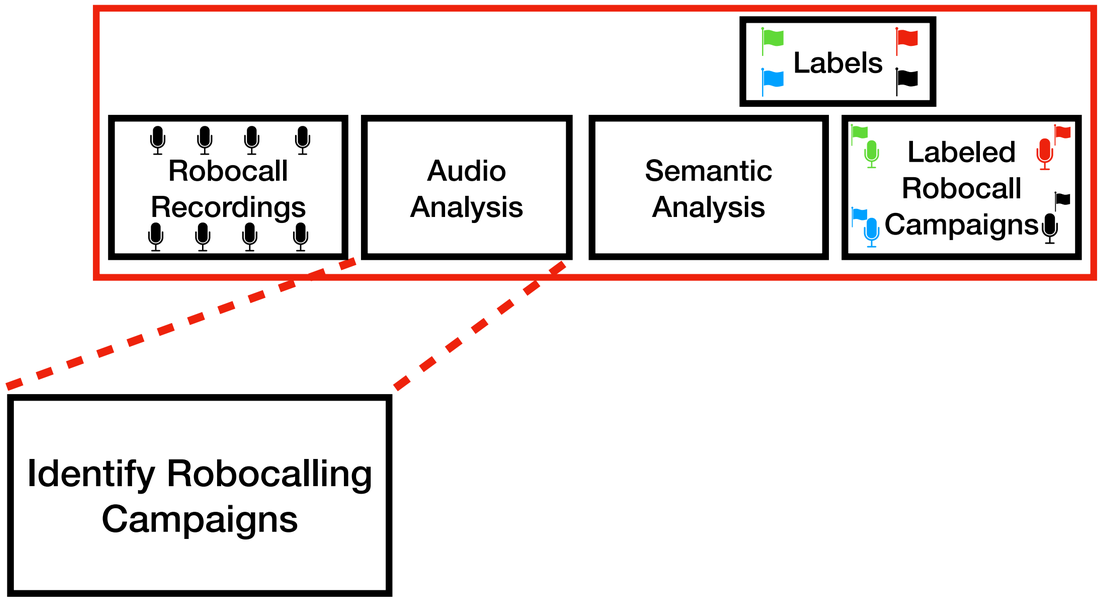
SnorCall - A Robocall Content Analysis Pipeline

Operating a telephony honeypot for data collection

23 Months
1.3 Million Robocalls



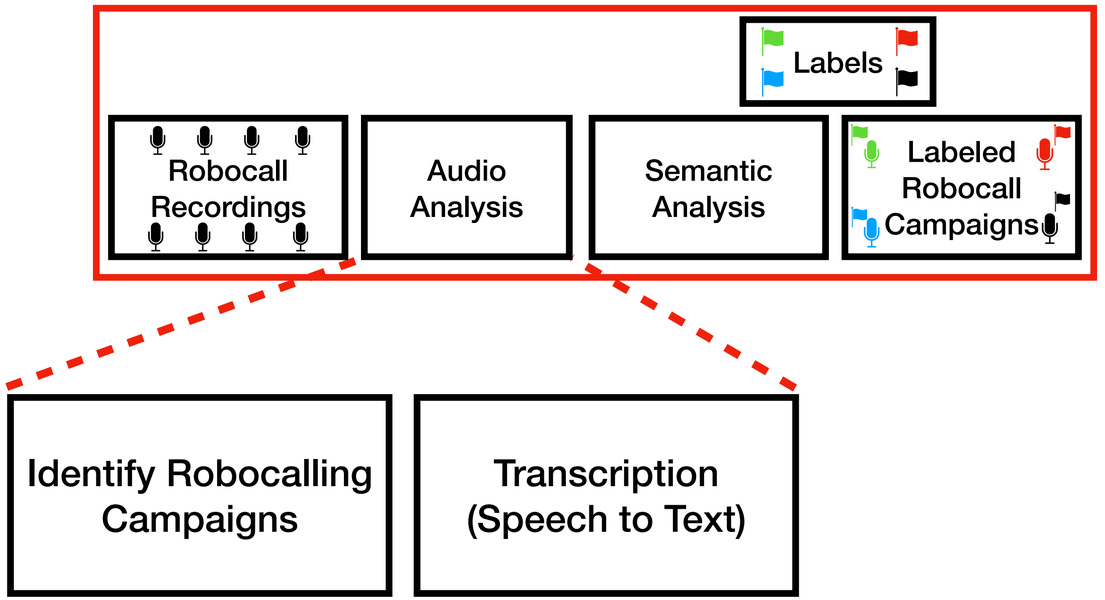
Robocall Audio Analysis



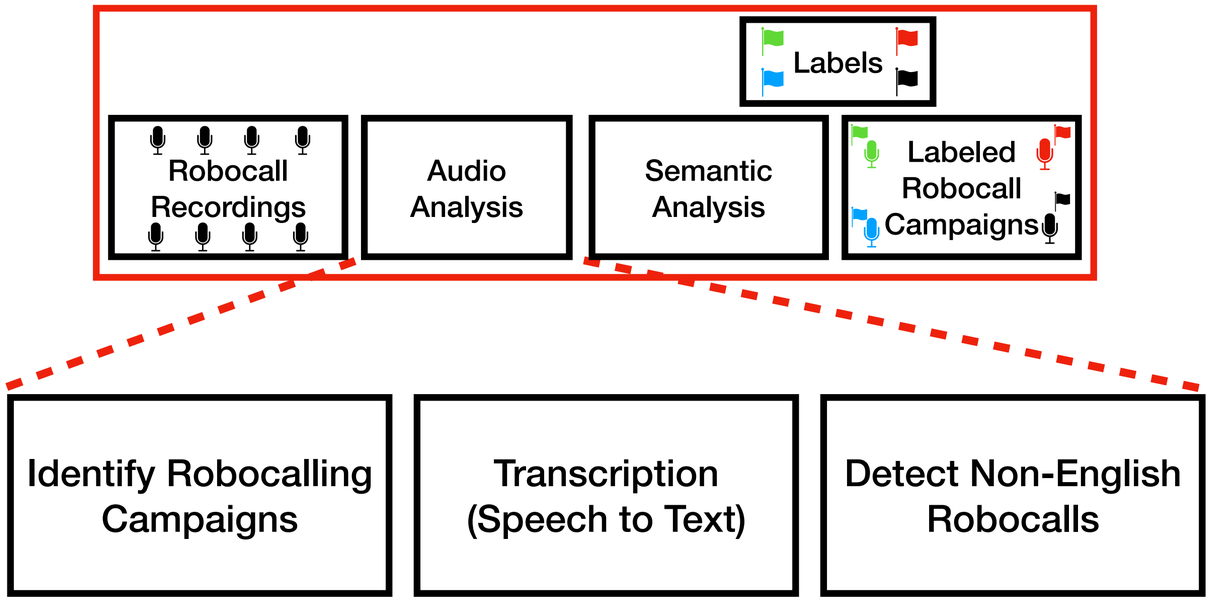
** Robocall Campaign: A set of audio recordings with identical or nearly identical audio content

More details in our 2020 paper: Who's Calling? Characterizing Robocalls through Audio and Metadata Analysis

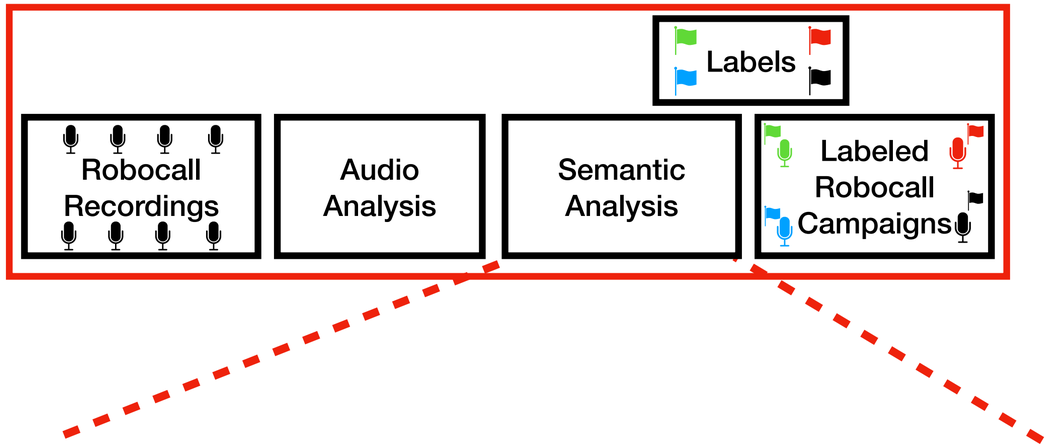
Robocall Audio Analysis



Robocall Audio Analysis



Robocall Campaign Labeling



Categorize robocalls based on the call semantics

Intuition behind Labeling a Social Security Scam call

Social Security Administration calling regarding a lawsuit which has been filed on your name and your social before we **suspend your social number** and issued your warrant of arrest against your name **call Social Security immediately** on 208-203-XXXX. I repeat 208-203-XXXX. Thank you. **

Social Security Administration ORG calling regarding a lawsuit which has been filed on your name and your social before we suspend your social number and issued your warrant of arrest against your name call Social Security ORG immediately on 208-203-XXXX. I repeat 208-203-XXXX. Thank you.

```
NER_Tag_Presence (Type = “ORG”)
Keyword_Presence (keywords = [“Social Security number”, “SSN”])
```

** Robocall transcripts inspired by similar calls observed in our honeypot. NER annotation using SpaCy

Key Insight: Named Entities and Keywords can help in Labeling Robocall Campaigns

We used a semi-supervised ML framework called "Snorkel" to translate human intuition into functions that can label data.

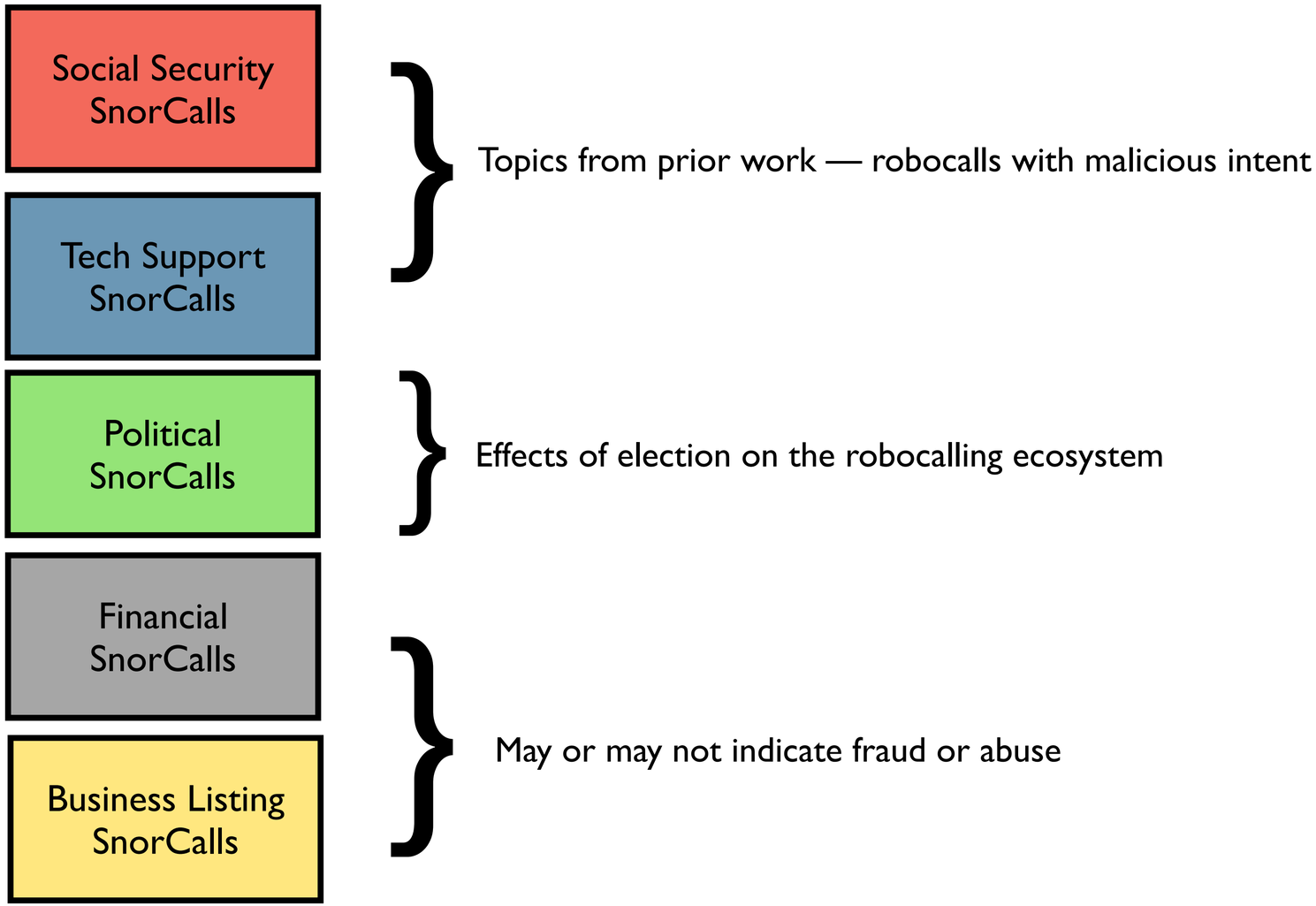
Snorkel framework enables domain experts to label large amount of data (text, image, etc.) by developing simple labeling functions

** Alexander Ratner, et. al. Data programming: Creating large training sets, quickly. Published at NIPS 2016.

** Alexander Ratner, et. al. Training complex models with multi-task weak supervision. Published at AAAI Conference on Artificial Intelligence 2019.

** Alexander Ratner, et. al. Snorkel: Rapid Training Data Creation with Weak Supervision. Published at VLDB 2020.

Robocall Categories Explored in Our Work



What did we learn about the robocalling landscape by processing our honeypot data using SnorCall?

Social Security Scams Have Evolved Substantially

Old Tactics (well-known)

1. Impersonate Social Security Administration employees
2. Threaten the victim with dire consequences
3. False sense of authority and urgency
4. Increase credibility by referencing other government entities: FBI, DEA, etc.

"This call is regarding to your social security number. We found some fraudulent activities under your name and **...**

arrest warrant has been issued and your Social Security would be suspended soon…

Please press one to talk with officer right away. I repeat, please press one to talk with officer right away. Thank you."

Social Security Scams Have Evolved Substantially

Old Tactics (well-known)

1. Impersonate Social Security Administration employees
2. Threaten the victim with dire consequences
3. False sense of authority and urgency
4. Increase credibility by referencing other government entities: FBI, DEA, etc.

"This call is regarding to your social security number. We found some fraudulent activities under your name and ...

arrest warrant has been issued and your Social Security would be suspended soon…

Please press one to talk with officer right away. I repeat, please press one to talk with officer right away. Thank you."

New Tactics

1. Impersonate Social Security Disability advisors
2. Target the disabled and elderly*: "..eligibility for Social Security disability benefits.."
3. Non-intimidating and seemly well-intended
4. Government impersonation: "…disability advisor with National Disability"

"Hello, my name is Amy and I'm a social security disability advisor advisors on a recorded line.

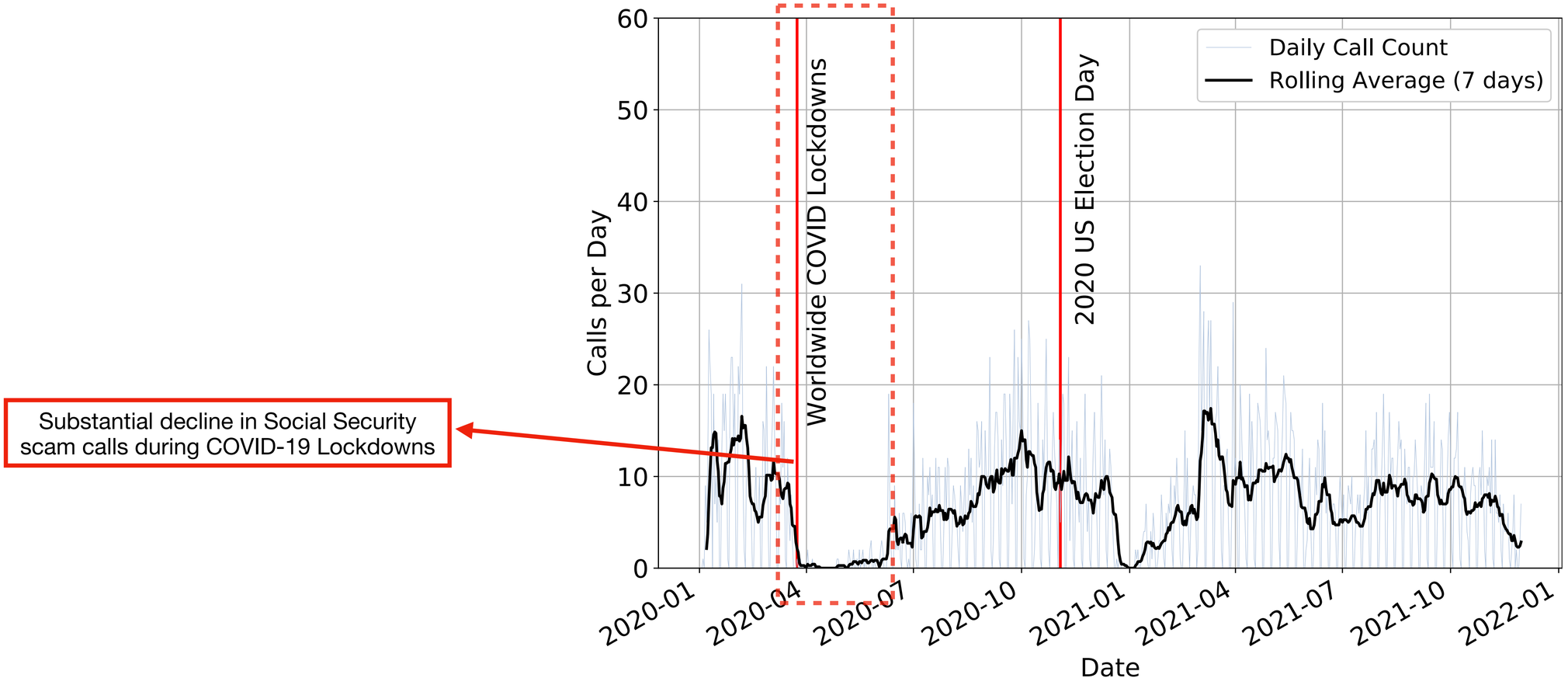
And my call back number is 866-201-XXXX.

Now I show here that you were recently inquired about your eligibility for Social Security disability benefits.

Can you hear me? Okay?"

* "Who can get Social Security disability benefits?": https://www.ssa.gov/pubs/EN-05-10029.pdf

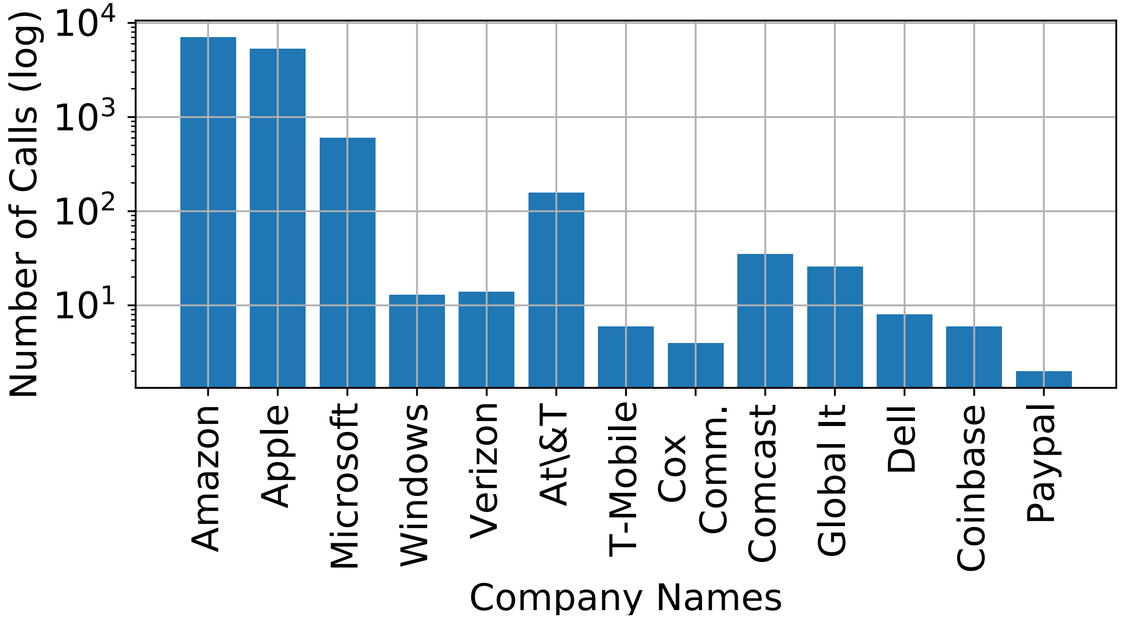
Social Security Scams Operate Like Regular Businesses

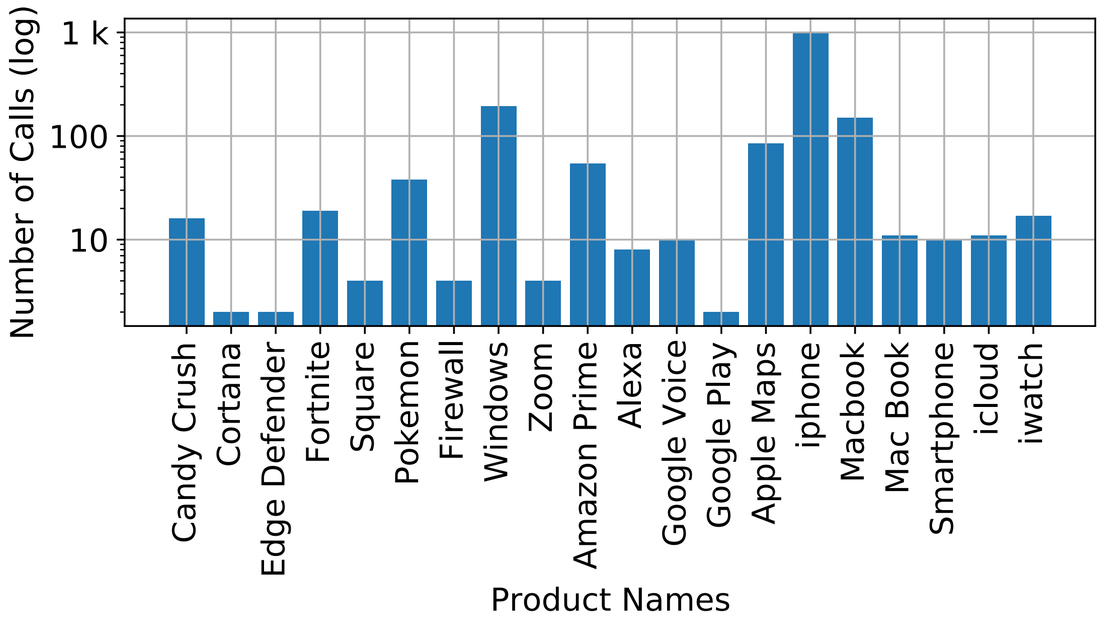


Social Security scam operations function like regular businesses, and were impacted by world-wide COVID-19 lockdown restrictions

What's the "Tech" in Tech Support Scams?

Tactics: Apart from impersonating well-known tech companies, Tech Support robocalls also impersonate well-known telecom carriers and consumer electronics companies





What's the "Tech" in Tech Support Scams?

Tactics: Apart from impersonating well-known tech companies, Tech Support robocalls also impersonate well-known telecom carriers and consumer electronics companies

Pricing Strategy: A median Tech Support scam call attempts to defraud a victim by about $400

What's the "Tech" in Tech Support Scams?

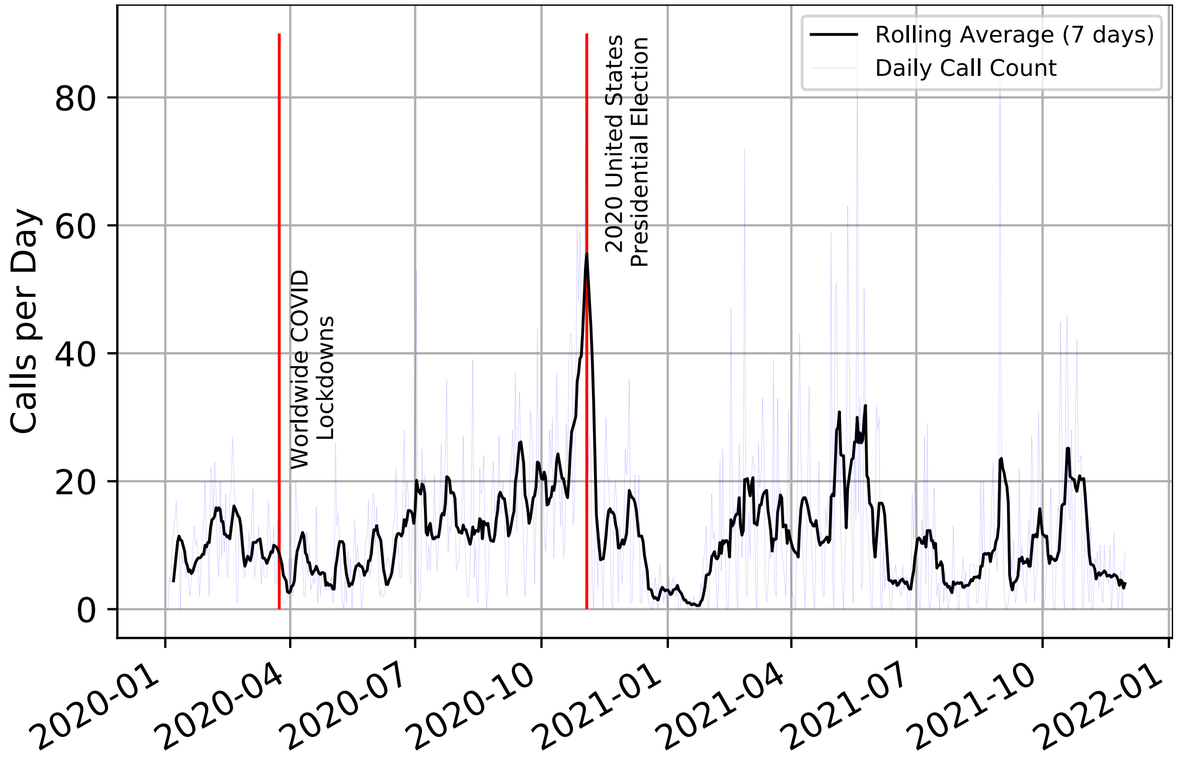
Tactics: Apart from impersonating well-known tech companies, Tech Support robocalls also impersonate well-known telecom carriers and consumer electronics companies

Pricing Strategy: A median Tech Support scam call attempts to defraud a victim by about $400

Cryptocurrency "Tech Support": We uncovered a handful of calls impersonating Coinbase support agents offering assistance to recover locked Bitcoins

The 2020 Presidential Elections and the Robocalling Landscape

Politics as usual: Political robocalls gradually increased towards the 2020 US Presidential election, peaked on the election day (3rd Nov 2020) and declined steeply right after



*We did not perform a partisan analysis of political calls observed in the honeypot, i.e. attempt to classify political calls into either Democratic, Republican or Independent.

The 2020 Presidential Elections and the Robocalling Landscape

Anomalies: We studied the cause of increase in political calls due to large outlier campaigns between August to November 2021

- These outlier campaigns claimed to represent a non-existent "Economic Impact Student Loan Forgiveness Program recently put into effect by the Biden Administration" and encouraged callers to enroll

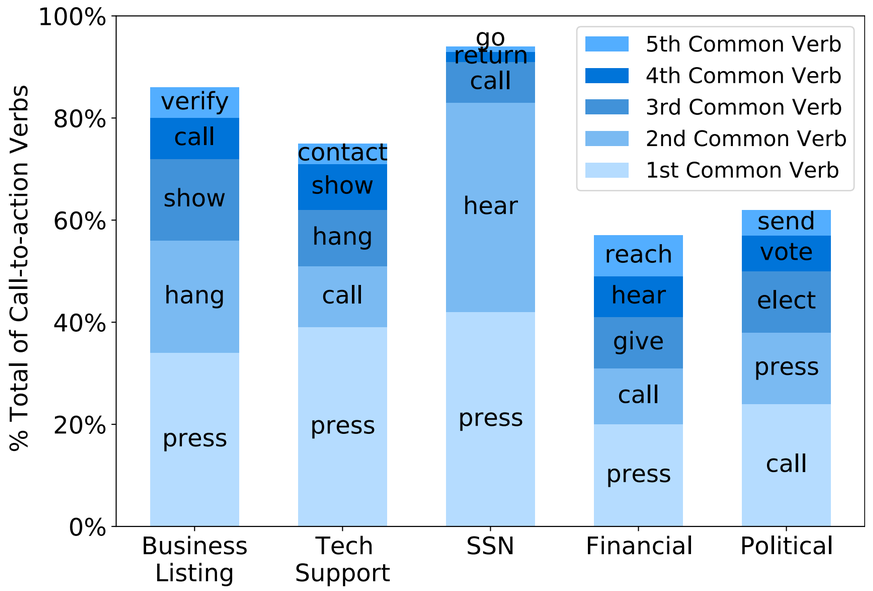
- These campaigns were labelled as both Financial and Political — highlights the robustness of SnorCall

*We did not perform a partisan analysis of political calls observed in the honeypot, i.e. attempt to classify political calls into either Democratic, Republican or Independent.

Takeaway: SnorCall enables domain experts to extract insights from large number of robocall recordings

How do Robocalls Engage with their Targets?

- We applied state-of-the-art task extraction techniques* to extract calls-to-action from robocall transcripts
- 72.79% (~18k) campaigns used at least one valid call-to-action
- Instructing users to press a digit or call a number is a popular engagement tactic



* **Lin: Unsupervised Extraction of Tasks from Textual Communication**
Parth Diwanji, Hui Guo, Munindar Singh, and Anup Kalia
In Proceedings of the 28th International Conference on Computational Linguistics, 2020

Callback Number Analysis

"Social Security Administration calling regarding a lawsuit which has been filed on your name and your social before we suspend your social number and issued your warrant of arrest against your name call Social Security immediately on ***208-203-XXXX****. I repeat* ***208-203-XXXX****. Thank you. "*

- SnorCall has a callback number extraction accuracy of 100% (including North American Numbering Plan validation)
- Callback numbers rarely match the asserted caller ID (about 95% calls have different caller ID and callback number)
- **Evidence of shared infrastructure:** we saw one robocalling campaign selling health insurance and another contacting about car warranty using the same callback number
- Ownership information about a callback number can be tracked down and used to identify the service provider (and potentially the entity/person/organization who owns that number)

But wait.. There's More (in the paper)!

- How to train and evaluate SnorCall models for any other types of robocall category?
- A comprehensive labeling codebook for various types of robocalls
- How financial robocalls impersonate the IRS, Credit Card companies, and Banks
- Strategies used by Tech Support scammers to target smart watch, smart devices, and gaming platform users
- Detailed discussion on how our work enables stakeholders to combat illegal robocalls

Key Takeaways

- SnorCall enables a domain expert to swiftly process large number of robocalls while accurately labeling the robocalls of interest.
- Well-known scams like Social Security and Tech Support have evolved over time. They continue to target the vulnerable population in our society by impersonating federal agencies and well-known public organizations.
- Robocalls take advantage of important societal events — like elections, student loan forgiveness, and worldwide pandemic — to deceive their targets.
- Our pipeline enables stakeholders to move towards a proactive robocall mitigation approach by characterizing the robocalling landscape without completely depending on reports from victims.

Project Website

https://robocall.science

The Team

Trevor Dunlap

Alexander Ross

Sathvik Prasad

Athishay Kiran Mylappan

Elijah Bouma-Sims

Brad Reaves

Sathvik Prasad snprasad@ncsu.edu

Brad Reaves bgreaves@ncsu.edu