CS 455 – Computer Security Fundamentals

Dr. Chen-Yeou (Charles) Yu

Introduction

- Someone used to ask me how to describe the great picture in Computer Security
- My answered is --- interdisciplinary study

- A high level algorithm will be presented to fight against unwanted calls.
  - Introduction
  - The power of crowdsourcing
  - Edge computing can help us
  - Contextual Info. can help us
  - Calling content can help us
  - Finally, our high level algorithm

Introduction

- The growth of telecommunication fraud has caused tremendous loss to end users.
- In this study, it is purely from an end user's perspective.
- In the US, the Federal Trade Commission (FTC) have worked with telephone carriers to blacklist robocalls or scam callers by periodically announce latest blacklisted numbers as datasets on their website.
- Google, have embedded signature-based blocking mechanism in their latest Android systems --- less effective
  - They try to involve OEM phone manufacturers, carriers, and end users to maintain the blacklists
  - At least, this is their 1st step

Introduction

- With the advancements in internet and telephony technology, scammers, spammers and criminals are exploiting new types of abuses on top of it.
  - Caller number spoofing --- by taking advantage of low cost and spamming campaign to easily reach and deceive end use
    - Most of the people would reject the call from 1-800 calls, but they still pick up the local calls (the same area code)
    - More recently, open-source software has made it possible for almost anyone to spoof calls with little cost or technical knowledge.
    - One of the most prevalent ways of spoofing is through VoIP.
    - VoIP stands for “Voice over Internet Protocol” and is basically a phone service delivered via the Internet.
    - If your internet connection is of decent quality, then your phone service can be delivered through the internet rather than your phone carrier.

Introduction

- So it is like the video streaming, no need for video channels
- In this way, we don’t even need the landlines
- VoIP service provider can give you a choice to setup the phone number
  - The representing phone numbers
- And it can be **changed at any time**
  - What!??

Introduction

- Criminals may target special ethnic groups or the people in specific location.
  - I will talk about this later in the Social Engineering
  - For example, international students in the college towns are easily targeted
- Purely signature based (phone# based) blacklisting is not good enough

The power of crowdsourcing

- Since the unwanted calls would like to pretend to be local callers, we can do “simple statistics” to sort the “local heavy hitters”
- Local people can contribute their efforts to catch the local heavy hitters during a period
- Still, they will use the number spoofing technology and the local heavy hitters has to be re-ranked periodically

Edge computing can help us

- The resources are used to be in the centralized cloud
  - Now, services are pushed to the edge (Distributed Computing)
  - Cell phone base stations are good candidates to deploy edge servers / edge services
    - If someone would like to make a phone call to you, the last step would be a “search” performed by the cell phone base station (because you registered in this base station), and get the call directed to your device (from the perspective of topology)
  - Edge servers can run strong location binding services. For example, local weather, local traffic
    - They can be a good service pool to compute the “local heavy hitter list”
    - They can push the most updated “local black list” to end users
    - End users can contribute their local black lists to edge services to help them compute the “local heavy hitters”

Contextual Info. can help us

- Yes, it is. But you need to know machine learning a little bit. Or? You can use simple statistics for approximation
- For example, this is someone's "working hours" in the company

Day	7-8AM	8-9AM	9-10AM	10-11AM	11-12AM	12-1PM	1-2PM	2-3PM	3-4PM	4-5PM
Day 1	Driving	Working	Working	Working	Eating	Working	Meeting	Working	Working	Break
Day 2	Driving	Meeting	Working	Working	Working	Eating	Working	Meeting	Working	Working
Day 3	Eating	Working	Working	Working	Working	Eating	Meeting	Working	Meeting	Break
Day 4	Meeting	Driving	Eating	Working	Working	Eating	Meeting	Working	Working	Break
Predicted Day 5	Driving	Working	Working	Working	Working	Eating	Meeting	Working	Working	Break

- There is a great purpose to know someone's predicted contextual info.

Contextual Info. can help us

- What if we have a list of predefined states with the “willing” to pick up the phone? “1”: Yes, pick up. “0”: No, I don’t want to pick up!
- For example, in a daily schedule, not just the working hours, I might have the following states = {Driving, Working, Eating, Sleeping, Break, Shower, Cooking, Recreation, Reading, Workout}
- Here is the example output for working hours (8AM~5PM)

Day	7-8AM	8-9AM	9-10AM	10-11AM	11-12AM	12-1PM	1-2PM	2-3PM	3-4PM	4-5PM
Day 1	Driving, 0	Working, 0	Working, 0	Working, 0	Eating, 1	Working, 0	Meeting, 0	Working, 0	Working, 0	Break, 1
Day 2	Driving, 0	Meeting, 0	Working, 0	Working, 0	Working, 0	Eating, 1	Working, 0	Meeting, 0	Working, 0	Working, 0
Day 3	Eating, 1	Working, 0	Working, 0	Working, 0	Working, 0	Eating, 1	Meeting, 0	Working, 0	Meeting, 0	Break, 1
Day 4	Meeting, 0	Driving, 0	Eating, 1	Working, 0	Working, 0	Eating, 1	Meeting, 0	Working, 0	Working, 0	Break, 1
Predicted Day 5	Driving, 0	Working, 0	Working, 0	Working, 0	Working, 0	Eating, 1	Meeting, 0	Working, 0	Working, 0	Break, 1

Contextual Info. can help us

- Another thing, what if we are just getting tired to differentiate the spam/scam/robo calls?
  - We can simply put all of them into “unwanted”.
  - This is the human centered design which starts from knowing your daily schedule first. (It would be good to have a “predicted” schedule!)
  - If I can know your following schedule, in advance, it would be wonderful!
    - Hourly
    - Half-hourly
    - …
  - For example, if someone calls you during the weekend, I can give this phone call with higher “unwanted” possibility
  - Or, when I’m sleeping in the midnight. The phone calls can be totally seen as “unwanted”, except my **close friends.**

Contextual Info. can help us

- If we can collect someone's schedule, we can do some predictions
  - What are you doing in the next hour, next half-hour or even next minute?
  - We can use Neural Network or Recurrent Neural Network (RNN) for time series prediction
  - Ideally, if a machine learning model is well-trained, it can output a good prediction of a person's daily schedule
  - For example, if someone calls me at 9 PM, it is supposed to be for my shower time.
    - If this phone number is in **my contact list** in my phone, just let it ring
    - If this phone number is **unknown**, I can see my phone with a **pop-up a warning notification**, saying this one could be "**unwanted**"
    - It serves as a suggestion, not to pick up this one!

Calling content can help us

- Keyword identification
  - It's a good manner to listen to other’s speaking ^_^
  - Because it can give you lots of “features”
  - If we can keep a dictionary of keywords
    - All of the U.S. governmental agencies or public offices, they do not call you.
      - Instead, they are very traditional, then mail you the letters, NOT emails.
    - If you can identify they are saying they are from some agencies… (You know what I mean)
    - If you can identify their speaking with very strong foreign accents…
    - If you can identify their speaking are about insurance (property, vehicle, life, health,…)
  - Then, you know those calls are possibly unwanted!

Calling content can help us

- Maybe we can try to listen to their speaking for the previous 10 secs, if we can develop such kind a of App
  - Then, we can quickly make the decision, hang up the phone number and blacklist it, or not?

Finally, our high level algorithm

Data: incomingCallerID, CurrentContactList

Result: BlackCallerIDList

```
While incomingCallerID do:
        Read incomingCallerID;
        if incomingCallerID NOT IN CurrentContactList
                if previous 10 secs matches some of our keywords
                        BlackCallerIDList+= incomingCallerID
                end if
                if timestamp(incomingCallerID) is mapped to “0”
                        BlackCallerIDList+= incomingCallerID
                end if
        else
                Let the phone ring
        end if

end
```

Finally, our high level algorithm

- However, this algorithm is not perfect
  - It needs some work-arounds
  - The part of the timestamp(incomingCallerID) depends on someone's living style
  - For example, if Dr. Yu get fired by Truman and went to a startup company being a CEO, his lifestyle would be TOTALLY different from his daily schedule recently.
  - We will need to re-train our machine learning model and do re-prediction
    - There is an kind of performance index, the prediction error rate
    - If the rate is getting higher, we know it is the time to re-train the model.
  - You can definitely use pure statistics but the effect wouldn't be good.

Appendix

- Everything to Know About Phone Number Spoofing
  - https://www.kaspersky.com/resource-center/preemptive-safety/phone-number-spoofing