I Know Where You are and What You are Sharing

Exploiting P2P Communications to Invade Users’ Privacy

Stevens Le Blond
Chao Zhang
Arnaud Legout
Keith Ross
Walid Dabbous

Presenter: Arakadakis K.
We know where you are!
We know where you are!

RELAX! We won’t tell anybody!
Motivation

- Can less-trustworthy entities track our whereabouts using real-time communication services (Skype, VoIP)?

- Can somebody link a user behind an IP address to his internet activities (e.g. file-sharing)?
BUT

HOW?
How?

- Real-time communication services are done P2P.
- There are services that give us the location of an IP.
How?

1. I contact Bob, whose Skype id is Bob_123
How?

1. I contact Bob, whose Skype id is Bob_123

2. I sniff the datagrams and extract Bob’s IP.
How?

1. I contact Bob, whose Skype id is Bob_123
2. I sniff the datagrams and extract Bob’s IP.
3. I get Bob’s location using an IP geolocation service.
How?

1. I contact Bob, whose Skype id is Bob_123
2. I sniff the datagrams and extract Bob’s IP.
3. I get Bob’s location using an IP geolocation service.
4. I can crawl P2P file-sharing systems and see if Bob’s IP uploads or downloads any file.
Some problems to solve

- How can I find Bob’s Skype id?
- How can I know which packets come from Bob?
- How can I call Bob without him knowing?
- NAT?
So, they created a system that does all the above.
Finding a person’s ID
Finding a person’s ID

- When a somebody creates a Skype account, needs to provide an e-mail address, a Skype ID and other information.

- We can search for their Skype ID using their birth name or e-mail.
Finding a person’s ID

What a mess!

1. Use side information from their account.
2. Call them and analyze their IP (location, ISP)
Finding a person’s IP Address
Finding a person’s IP Address

● Call Skype_id, sniff received datagrams and get sender’s IP.

● If the Skype user is offline, Skype provides the most recently used IP (within past 72 hours).
How does skype call establishment work?

1. User online and not behind NAT

   Caller send packets directly to the callee.
How does skype call establishment work?

2. User online and behind NAT

Callee is instructed to initialize communication to the caller (via his supernode) and callee’s public IP will be in the received datagram.
How does skype call establishment work?

3. User offline (but was online at least once the past 72 hours)

Caller send packets directly to the callee.
How does skype call establishment work?

- Caller’s and calle’s initial communication protocol follows a specific pattern for each case shown before.
- Based on these patterns, we can infer which packets originate from the target.
Inconspicuous calling
Inconspicuous calling

- During call establishment, both TCP and UDP packets are sent between caller and callee.
- If the TCP connections are not established, the callee will not be notified about the call.
- Thus, drop all incoming and outcoming TCP SYN packets.
Mobility of Skype users
Mobility of Skype users

- For geolocation purposes, MaxMind was used.
- We check the target’s IP periodically.
Let’s scale up!
Let’s scale up!

- 13M Skype IDs -> 100,000 random IDs -> 10,000 active the past 72 hours
- Call these 10,000 IDs on an hourly basis.
- Before making next call, wait long enough, so that the complete packet pattern elapses. (~15s) -> 240 IDs / hour can be probed
- After some interval optimizing, they can call 340 IDs / hour x 30 machines.
Results
Results
What about internet usage of Skype users?
File-sharing usage of Skype users

- They focused on BitTorrent
- Three components: *Skype Tracker*, *Infocash Crawler*, *BitTorrent Crawler*, *Verifier*
File-sharing usage of Skype users

Skype Tracker

Calls 100.000 active Skype users, once per day, and sniffs their IP addresses.
File-sharing usage of Skype users

Infocash crawler

Collects file identifiers from the PublicBitTorrent tracker, including the seeds and leechers (IPs).
File-sharing usage of Skype users

BitTorrent Crawler

- Collects the IP addresses participating in the 50,000 most popular torrents from the Mainline DHT, every hour.
- Mainline DHT is a decentralized tracker.
- When a peer wants to download a new file, it contacts Mainline DHT to obtain a list of peers distributing that file.
- DHT nodes do not implement blacklisting strategies.
How to deal with the NAT problem
Verifier tool

- Tells if Skype and BitTorrent run on the same machine.
- Given an IP, participating in both Skype and BitTorrent using the IP-ID field in the IP datagram.
- At the same time, verifier calls the Skype ID, and initiates a handshake with the BitTorrent client.
- If the distance between the IP-IDs is small, it is very likely that Skype user to be the bitTorrent downloader.
Results

- 765 users used both Skype and bitTorrent
- 400 users are indeed using bitTorrent
Summary

- It is possible for an attacker to determine the IP of a Skype user (if active).
- The attacker can locate a Skype user.
- The attacker can link a Skype user with certain online activity.
Thank you!

Q&A