Week 7 Blog

Sree Ram Boyapati	a1775690@student.adelaide.edu.au
Course	Comp Sci 7092 - Mobile and Wireless Systems
Group	CovidGuard-F
Supervisor	Zach Wang

What have I done this week?

For the demo purposes, I have mostly refactored a few parts of the application and reviewed code of my teammates for showing tangible outcomes during demo especially storing of TEKs and RPIs in the SQLite database.

I personally refactored a few parts of the system like permission workflow on the client-side and logging backend at the server to google cloud v2 specification to get more data from the cloud.

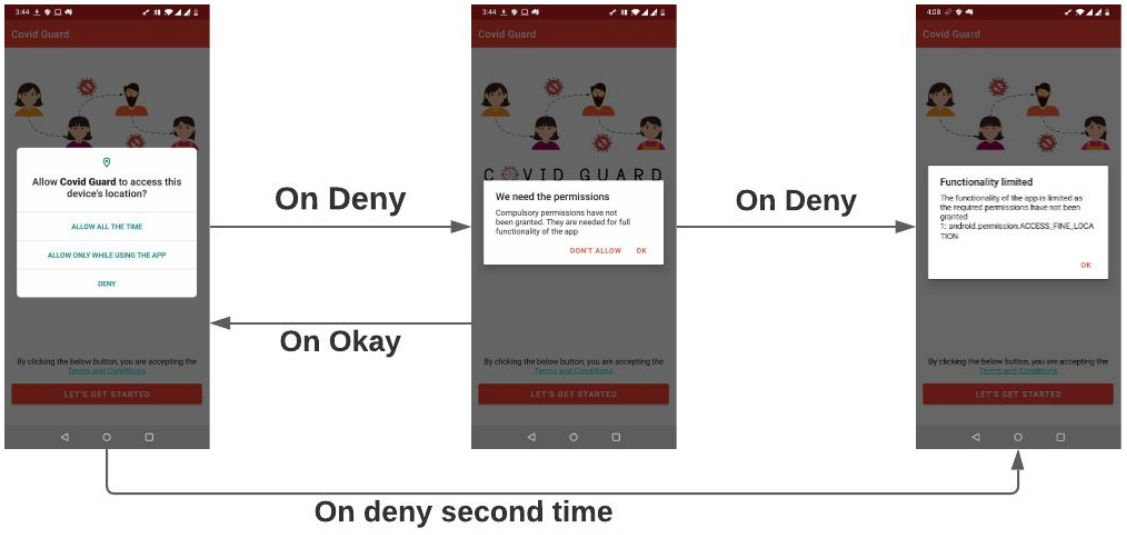

Figure 1: Permission Workflow

I have also refactored storage of registration token in encrypted shared preferences using AES encryption scheme.

I have started working on upload diagnosis keys of the server. Some of the challenges involved is -

1. Authentication using registration token.
2. Upload of large size of data using multi-part body

Amal and I are working on this from the client and server-side respectively. We are also encountering some technical debt:

1. Using data transfer objects[1], We can use Moshi JSON library for easy serialization and deserialization of requests.
2. Using room library to make the queries. (I want to work on this during the break.

Open Issues:

1. Feature - Submit Diagnosis keys to server https://github.cs.adelaide.edu.au/2020-Mobile-and-Wireless-Systems/CovidGuard-F/issues/32
2. https://github.cs.adelaide.edu.au/2020-Mobile-and-Wireless-Systems/CovidGuard-F/issues/26
3. Fix Logging specification issues - https://github.cs.adelaide.edu.au/2020-Mobile-and-Wireless-Systems/CovidGuard-F/issues/39
4. Room Refactor - https://github.cs.adelaide.edu.au/2020-Mobile-and-Wireless-Systems/CovidGuard-F/issues/40

Closed PRs:

1. https://github.cs.adelaide.edu.au/2020-Mobile-and-Wireless-Systems/CovidGuard-F/pull/30

[1] "Data transfer object - Wikipedia." https://en.wikipedia.org/wiki/Data_transfer_object. Accessed 12 Sep. 2020.

Paper Review # 1

Jun Xu, Jinliang Fan, M. H. Ammar and S. B. Moon, "Prefix-preserving IP address anonymization: measurement-based security evaluation and a new cryptography-based scheme," 10th IEEE International Conference on Network Protocols, 2002. Proceedings., Paris, France, 2002, pp. 280-289, doi: 10.1109/ICNP.2002.1181415.

Why have I chosen this paper?

One of the main challenges of privacy in a client-server interaction that the server knows the client IP address of the user which is privacy-sensitive. Routing through a mix of servers was suggested in Andrew Y. Lindell et al. 2018. [1] and asymmetric encryption-based authentication scheme for Anonymous authentication.

However, there are some caveats:

1. Compatibility with existing protocols
2. Mitigation of attacks on mix servers that route the traffic from client to server.

In Eckert C et al., 2001 [2], Headers were classified into critical and privacy-preserving fields, These can be added on the client-side -

1. Referrer - contains the information on previous searches by the user. This field may not be suitable for our needs.
2. X-Forwarded-For, Client-IP - contains the information on the client IP address.
3. Cache-Control - data from CDN may be cached for 24 hours and cache-info should not reveal the user's identity.
4. User-Agent - User-agent contains a lot of data on the origin of the request which may or may not be necessary for the server.

An attacker can see the packets that are being sent and can launch a slow-loris attack [3] as the server has no identifying information about the client to blacklist the IP. Request time-outs and using a proxy server to manage the connections for the server can slow the attack. As IP-based strategies may longer be applicable.

This paper explores privacy-preserving IP anonymisation to give server flexibility and client privacy by anonymization of network traces for analysis of traffic offline.

This is important because, in case of attacks, network traces should be privacy-preserving to be shared with analysis teams.

Proposed Work

Important Terms:

Address Space Tree	Entire IPv4 address space can be stored in a binary of height n=32
TCP Dump Distinct IP Addresses	A network trace of a server may have a subset of a complete binary tree of the address space.
Anonymization Function	Takes an address space tree as input, and produces an anonymized address space tree with the one-to-one mapping by rearranging the IP addresses.
TCPDPRIV[2]	A tool that removes private information from TCP Dump.

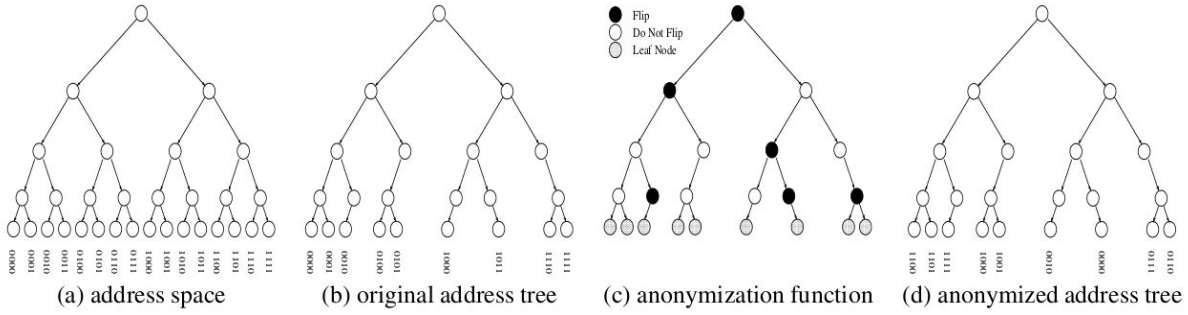

Figure 2: Prefix-Preserving Anonymization process

Novelty:

TCPDPriv cannot produce a distinct mapping if the order of the IPs in the log changes. This seriously limits the parallel processing capabilities of the TCPDPriv. Large files cannot be broken down into smaller pieces and processed independently.

The paper tries to address these concerns and challenges. They have changed the Anonymization function to be trace-independent.

[2] "man page." http://fly.isti.cnr.it/software/tcpdpriv/tcpdpriv.0.txt. Accessed 12 Sep. 2020.

$$f_i(a_1 a_2 \cdots a_i) := L(R(P(a_1 a_2 \cdots a_i), \kappa)) \qquad (2)$$

L - Least Significant Bit
R - Psuedo Random Function like AES
P - Padding to match the block size.
K - Cryptographic Key like RPIKey
For a chosen key and pseudorandom function family, An raw IP address is always anonymized to an anonymous address regardless of the relative order of trace. Using this property, the parallelization of TCPDPriv is possible.

Conclusion of the Research

1. The author underscored the need for secure distribution of key used in PRF Family to all servers of the service.
2. Attacks on the aforementioned prefix-preserving anonymization function are divided into two categories -
    a. Cryptographic Attacks
    b. Semantic Attacks
3. Security against cryptographic attacks is provided by properties of the pseudo-random function family where each function is indistinguishable from others.
4. Vulnerability to Semantic attacks depends on the entropy of the trace and all prefix-preserving anonymization techniques functions are vulnerable to cryptanalysis techniques like frequency analysis.
5. The effect of an attack is measured using -
    a. Number of the unknown sub-trees root of the address space denoted by C
    b. Number of unknown bits in the total address space that are revealed - U
    c. Total number of addresses where Highest Significant bits have been revealed - F_i
6. Frequency analysis reveals the IP addresses that are distinguishable from others. However, it is only possible if their frequency comparatively is much higher. Large internet companies which run trackers and crawlers like google are susceptible to it not individual users as they are indistinguishable by frequency analysis alone
7. Active attacks where an attacker can encode some bridging information to identify anonymized IP are very hard to counter.

The outcome of the Project

1. For our demo purposes, We can use TCPDpriv for anonymous network traces. However, at the scale of GAEN, the proposed protocol can run as a side-car. TCPDpriv maintains the same strength for a smaller setup.
2. By rotating the keys and using Google KMS [3], We can change the anonymisation scheme every N days for generating network traces.
3. The authors have given measurement of attacks and durability of the scheme using three formulas which can be adopted for experimentation of a prefix-preserving anonymisation scheme. The gist of the problem in layman terms comes down to are two binary trees equivalent? It is a favourite google interview question - https://leetcode.com/problems/flip-equivalent-binary-trees/
4. For N=32, Equivalency of binary trees should be computationally infeasible. however, authors assume that the attacker is computationally constrained. In the case of contact tracing, that assumption may not hold true.
5. The datasets were of Tier-1 ISP, In the case of GAEN protocol, The most significant bits may be the same for all the citizens of a country as the application is localized. Least significant bits are useful for personal safety.

[3] "Cloud Key Management | Google Cloud." https://cloud.google.com/security-key-management. Accessed 12 Sep. 2020.

References

1. Lindell, Yehuda. "Anonymous authentication." Journal of Privacy and Confidentiality 2.2 (2011).
2. Eckert, C. and A. Pircher. “Internet Anonymity: Problems and Solutions.” SEC (2001).
3. SlowLoris Attack - https://www.cloudflare.com/learning/ddos/ddos-attack-tools/slowloris/