Vecsim into Pathsim: Optimizing Cellphone Contact Tracing using Realistic Pathing

Computer Science and Information Systems Master’s Thesis Proposal

Michael Mark

CSIS Master’s Thesis

Saginaw Valley State University

December 15, 2020

**Abstract**

Vecsim was written about and introduced by A. Mukherjee et al. in their paper *Vecsim: Carrier-based, Privacy-Preserving Cellphone Contact Tracing* [1]. It is a novel contact tracing idea that focuses mainly around using cellphone network data (also known as MRO data) to formulate a method showing user proximity based on that data. However, the method proposed did have many assumptions, including Line of Sight issues, and being outdoors only. Pathsim is a new concept that takes users realistic walking paths, simulated over tens of thousands of attempts, to further enhance and predict user proximity based on their predicted pathing. Compared to Vecsim, Pathsim may be a little bit more memory intensive, but will provide a more accurate estimate of user proximity in a real-world setting.

**Background**

Contact tracing takes place in many forms. A few notable mentions are:

Contact tracing with cell phone signal, contact tracing with QR code, manual contact tracing, Bluetooth methods, social networking, Wi-Fi, and wearable devices [1]. All these methods require the user to be connected or localized in some way. Vecsim was developed to remove the localization requirement (input or data from the user’s end), and to make it carrier side only. This way, we can further enhance our technology to find and trace any contact we want without installing unnecessary software on the user side, and instead have the server be able to locate the user.

Contact tracing is developing at a fast rate and is helpful in many real-world situations, including currently during the COVID-19 pandemic, where we could potentially find infected individuals and other potentially infected individuals to save people’s lives.

At the current time near the end of 2020, many contact tracing programs are reaching their limits [2] and a national effort still needs to be made for the spread of the pandemic to be reduced through the efforts. This paper is an effort to help relieve localization programs and to help in the effort against the spread of COVID-19 via contact tracing.

**Vecsim**

Vecsim is a novel contact method that shows the distance between two users based on a formula. This formula Vecsim uses in its initial estimate is:

The meaning for these symbols are:

: the distance between the two users

: users 1 and 2’s *Tadv* values, which represents the length of the signal from the base station to the user. This value is proportioned between 0-10 with a granularity proportion of 78 meters (represented by )

: user 1 and 2’s Angle of Arrivals, which is the real value between [0, 2]. This is the angle of the signal from the cellphone observed at the base station.

: 78 meters, the granularity proportion of the MRO (maintenance, repair, operations) file given by the cellphone base station

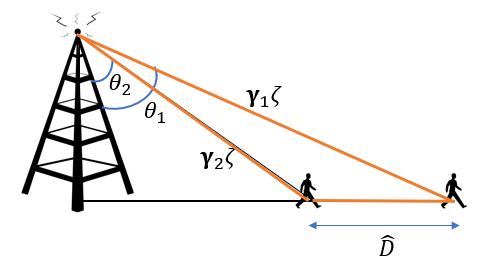


Figure 1: Vecsim initial estimate pictorial

The issue with this initial estimate is that the above formula assumes that the users are in Line of Sight with the tower (see Figure 1).

However, Vecsim further utilizes this initial estimate by throwing it into a probability density function to estimate its maximum likelihood based on the distances. The maximum likeliness formula also utilizes the signal strength, or RSRP value from the MRO data to determine whether the initial estimate is correct or whether it actually reflected off a building.

Vecsim will also create a database of Discontinuity Pairs, called the Discontinuity Pairs Database (DPD), which will keep track of signals that may be abnormal, such as a signal strength being very different after a user comes out of an alleyway. These pairs are all simulated using random walks that are generated in a random manner.

**Vecsim Pros and Cons**

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| Good starting point for non-localized contact tracing | Can only detect outside of buildings |
| Uses maximum likelihood distribution to estimate distances if initial distance estimate is incorrect | Has issues with LOS signals, but can deal with them using maximum likelihood, probability density functions and DPD learning |
| **Very good in non-urban areas with less LOS issues** | May not be able to adapt quickly to indoor signals and changing of environment |
| Preserves user privacy | DPD may take up to 500,000 paths before 90% accurate |

Figure 2: Table of Vecsim Pros and Cons

In Figure 2, I have listed a few of the pros and cons of the Vecsim method. It is noted that the method is very good in areas with little to no Line-of-Sight issues. However, in the real world, there will be many fluctuations, and signals that will reflect off buildings, or signals that become odd while inside buildings even (signal boosters). To overcome these issues, we must take Vecsim to the next level and propose realistic walking paths and try to imitate real people walking.

**Vecsim Optimization, Pathsim**

To further optimize the Vecsim method proposed by A. Mukherjee *et. al*, there are a variety of methods that we could use to do this.

1. **Refine the random walks to become more realistic walking.** In our Teams group which comprised of Beata A. Hejno, Anthony Galea, Morgan B. Simmons, Morgan B. Simmons, and me, we have been developing code to create more realistic walking paths and predicting where users may walk. This included extensive MATLAB research, coding, and a simulation done on a city with various blocks. Here is an example of coding in MATLAB that shows what is possible.



Figure 3: MATLAB plot of example shortest path

However, in conclusion, we found that users tend to want to hug buildings and cross the street perpendicularly. Figure 4 gives example of what we want to output, where the green line path is a more realistic path.

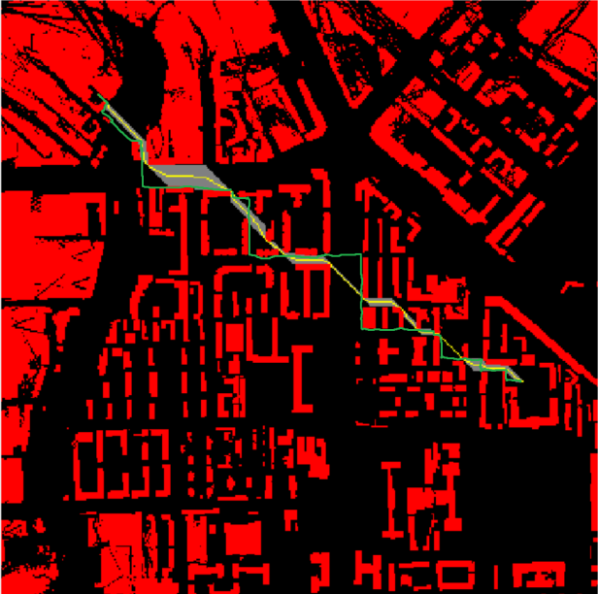


Figure 4: Shortest path vs Realistic Path

We are still working as a team to optimize the code to create realistic paths more quickly and Pathsim will in the end use our code to generate a large amount of walking paths at any given time in a matter of minutes or even seconds.

1. **If we know the routes and directions of an entity, we can use an algorithm to find their actual location with much more precision than just relying on the MRO data method utilized by Vecsim.**
2. So, to do this, I propose that we create a more realistic method named ***Pathsim***, based on pathing predictions between two different entities and accurate location retrieval.

**Proposal, Pathsim**

Thus, we propose to introduce Pathsim, a more sophisticated approach based on two entities’ movement to make a more accurate prediction on where they are based on the MRO log data as well.

Pathsim works the same as Vecsim in the fact that it is a software installed on the server side. Therefore, user privacy is still protected. It also is a bit more sophisticated in terms of its algorithm to find where the users are. It will conduct more realistic walking paths to figure out where the entities may be moving toward and use this data to find where the users are. Still, localization is not being used in this case as the user does not have to turn on Bluetooth or install any special software. However, Pathsim will be much more accurate in this case since it is able to retrieve more data regarding DPDs and figure out discontinuities in the MRO data better than Vecsim can.

Pathsim will be able to determine a user’s location with much more precision and reduce the time to learn unnecessary DPD pairs, so it will have less of a database storage and strain on the server. Pathsim will also be able to function quickly and more effectively in urban areas since it does not rely on the maximum probability function as much to determine whether the initial distance estimate is correct.

Sources:

[1] Avishek Mukherjee, Yaoguang Zhong, Zhenghao Zhang, Tingting Zhao, and Jinfeng Zhang. 2020. Vecsim: Carrier-based, Privacy-Preserving Cellphone Contact Tracing. In Proceedings of the 18th ACM Symposium on Mobility Management and Wireless Access (MobiWac '20). Association for Computing Machinery, New York, NY, USA, 47–55. DOI: https://doi.org/10.1145/3416012.3424627

[2] “Model shows potential contact tracing impact against COVID-19.” [Online]. Available: https://med.stanford.edu/news/all-news/2020/08/model-shows-potential-contact-tracing-impact-against-covid-19.html. [Accessed on December 13, 2020].