Privacy-Protecting COVID-19 Exposure Notification Via Cluster Events Without Proximity Detection

Paul Syverson
U.S. Naval Research Laboratory
paul.syverson@nrl.navy.mil

We provide a rough sketch of a simple system design for exposure notification of COVID-19 infections based on copresence at cluster events—locations and times where a threshold number of tested-positive (TP) individuals were present. Unlike other designs, such as DP3T or the Apple-Google system, this design does not track or notify based on detecting direct proximity to TP individuals.

The design makes uses existing or in-development tests for COVID-19 that are relatively cheap and return results in less than an hour, and that have high specificity but may have lower sensitivity. It also makes use of readily available location tracking for mobile phones and similar devices. It reports events at which TP individuals were present but does not link events with individuals or with other events in an individual's history. Participating individuals are notified of all detected cluster events. They can compare these locally to their own location history. Detected cluster events can be publicized through public channels. Thus, individuals not participating in the reporting system can still be notified of exposure.

Unlike common influenzas and other familiar diseases, COVID-19 appears to have a high degree of clustering in its dispersion and to be spread more in events where infected individuals spend time in close proximity to groups of others. There are additional dispersion clustering factors, such as whether gatherings are indoors and adequacy of ventilation. Clustering also occurs around some people who are individually linked to high number of infections. But generally, gatherings play a significant role in infection rates, whether or not they are exacerbated by other factors [5]. This includes both super-spreader large gatherings for social or cultural events as well as more moderate sized gatherings [1].

Most tests for COVID-19 initially rolled out required specialized equipment to evaluate collected samples, required days or more to return results, were expensive, or all of the above. But analyses indicate that faster, cheaper point-of-care tests can be more effective at identification of infected individuals than more sensitive tests that are slower [6, 4]. And HHS in partnership with the the DoD is now providing rapid point-of-care tests to communities across the U.S. [3]. Similarly, under the auspices of the WHO, a global partnership has plans to soon make 120 million rapid tests available in low- and middle-income countries [8]. Analyses of effectiveness are focused on identification and notification of infectious individuals. But coupled with dispersion patterns, another advantage of cheap, point-of-care tests emerges.

"[G]iven the huge numbers associated with these clusters, targeting them would be very effective in getting our transmission numbers down" [7]. Also, identifying a cluster does not require that everyone who was already infected at a particular event (location and time) has tested positive or even has been tested at all. As long as a sufficient *number* of TP individuals are associated with a given event, it is not important to identify which individuals were present, even pseudonymously, in order to indicate that the event constitutes a cluster. Since cluster events are indicators of significant risk of infection for all copresent at the event, informing individuals of those clustering events at which they were present is sufficient to notify them of potential exposure. And, individuals can make the determination of whether they were present at a clustering event entirely locally by comparing notification of clustering events with the location history in their phones.

Another potential advantage of a cluster-event based approach is that, even without limiting TP reporting to authorized individuals, it provides automatic counters to the possibility that "people may falsely report they have been infected to cause mischief or to keep people home in order to shut down school or even to disrupt an election" [2]. Someone falsely reporting a positive test cannot easily create such a result because they are unlikely to be the one to transition a location and time across the threshold of counting as a cluster event. And they cannot report at all for a location and time unless they were present at that event. Our cluster-event design thus automatically prevents, e.g., a student who is worried about an upcoming exam and anonymously reports a positive test from thereby causing his school to shut down or his whole chemistry class to be forced into quarantine. A fan of one sports team cannot trivially force a rival team into quarantine, etc.

References

[1] Muge Cevik, Julia Marcus, Caroline Buckee, and Tara Smith. SARS-CoV-2 transmission dynamics should inform policy. https://ssrn.com/abstract=3692807, September 14 02020.

[2] Lorrie Faith Cranor. Digital contact tracing may protect privacy, but it is unlikely to stop the pandemic. *Communications of the ACM*, 63(11):22–24, November 02020.

[3] COVID-19 rapid point-of-care test distribution. https://www.hhs.gov/coronavirus/testing/rapid-test-distribution/index.html, October 7 02020.

[4] Lee Kennedy-Shaffer, Michael Baym, and William Hanage. Perfect as the enemy of the good: Using low-sensitivity tests to mitigate SARS-CoV-2 outbreaks. https://dash.harvard.edu/handle/1/37363184, 02020.

[5] Sadamori Kojaku, Laurent Hébert-Dufresne, Enys Mones, Sune Lehmann, and Yong-Yeol Ahn. The effectiveness of backward contact tracing in networks. https://arxiv.org/abs/2005.02362, May 5 02020.

[6] Daniel B. Larremore, Bryan Wilder, Evan Lester, Soraya Shehata, James M. Burke, James A. Hay, Milind Tambe, Michael J. Mina, and Roy Parker. Test sensitivity is secondary to frequency and turnaround time for COVID-19 surveillance. *medRxiv*, 02020. https://www.medrxiv.org/content/early/2020/09/08/2020.06.22.20136309.full.pdf.

[7] Zeynep Tufekci. This overlooked variable is the key to the pandemic: It's not R. *The Atlantic*, September 30 02020. https://www.theatlantic.com/health/archive/2020/09/k-overlooked-variable-driving-pandemic/616548/.

[8] Global partnership to make available 120 million affordable, quality COVID-19 rapid tests for low- and middle-income countries. https://www.who.int/news/item/28-09-2020-global-partnership-to-make-available-120-million-affordable-quality-covid-19-rapid-tests-for-low--and-middle-income-countries, September 28 02020.

Augmenting GAEN with opt-in case linking

Google and Apple's Exposure Notification system (GAEN) can be augmented to allow users to provide additional information to their health authority. This information would allow contact tracers to integrate contacts notified via GAEN into their existing case management systems. Doing so would enable more accurate risk assessment, reduce duplicated work, and enable notified individuals to hear tailored advice. The primary modifications needed are allowing health authorities to store a set of Temporary Exposure Keys (TEKs) with a case file, and allowing notified individuals to send TEKs that that they matched with to their public health authority.

The modified system would work the same as the current GAEN implementation if neither the positive individual or contact opts in to share additional information. If the positive individual opts in, then their TEKs are stored with the rest of their case data in the case management system. If the contact opts in, then their name and phone and exposure notification data (including the TEKs matched with) are sent securely to the contact's public health authority. The time and location of exposure could also be sent to the health authority and access could be restricted unless both parties have opted in. In scenarios where both people have opted in, contact tracers would be able to see who was involved in a potential exposure, the (estimated) proximity and duration of exposure, and optionally the time and location. They would be able to compare this information with notes collected from the case investigation and provide more tailored risk assessments. If capacity is low, they could de-prioritize calling people have already received a notification. Cluster analysis and backward contact tracing would also be be enabled.

Currently GAEN works almost independently from manual contact tracing. Combining information from both systems could improve risk assessment, quarantine compliance, and cluster analysis, without significant loss of privacy or increased workload.

Adoption Metrics for Proximity Technologies

By: Scott L. David, Director of Information Risk Research Initiative (IRRI),
University of Washington Applied Physics Laboratory
Seattle, Washington

The Need:

We need metrics to help promote and steer adoption of Exponential Socio-Technical Systems (ESTS)
ESTS are systems that are:
(1) powered by higher-order effects of Moore's law,
(2) which rely on "network effects" for value, and
(3) depend on reliable/predictable performance/behaviors of humans and institutions as BOTH users AND as critical system components and data providers.

Our Approach:

IRRI creates programs and materials that support distributed testing ecosystems AND which smooth the pathway to adoption/commercialization of "technically feasible" technologies.

Source of the Problems is the Source of the Solutions:

Where the success of the technology is dependent on sufficient adoption and deployment across a large segment of a given population, it is critical that testing go beyond mere "technical feasibility" and consider additional factors affecting stakeholder uptake.

IRRI programs test "technically feasible" technologies for adoptability through "BOLTS" reasonableness.

BOLTS tests are based on metrics drawn from **B**usiness, **O**perations, **L**egal, **T**echnical and **S**ocial domains.

IRRI supplies the process and artifacts for thorough and complete mapping of the threat, vulnerability, risk, and opportunity landscape into which any new technologies are introduced, including those associated with proximity detection in pandemics and other public health settings.

Benefits of IRRI approach:

BOLTS tests are familiar to stakeholders because they reflect patterns of practice, language, and interaction rules drawn from real world settings. IRRI processes enable stakeholder discovery of new solutions to interaction risks (e.g., privacy, security, liability) that none of them can achieve unilaterally.

Other pathways to testing ecosystems and commercialization:

Many other current forms of technology transfer facilitation and institutionalization are not sufficiently broad and/or do not sufficiently address multiple stakeholders' needs to fully gauge potential for adoption of new technologies. IRRI's distributed testing approach can help to bring together existing pathways to enhance technology testing and commercialization, particularly in the case of ESTS.

Modeling the impact of automatic exposure notification for vulnerable communities

Krister Shalm
Applied Physics Division, NIST and Physics Department, University of Colorado, Boulder, CO

While most efforts in automatic exposure notification focus on large population groups, we study the effectiveness of such systems in smaller vulnerable communities. For example, it has been estimated that over one-third of COVID-19 related deaths have occurred among nursing home residents in the US. Understanding what criteria must be met to successfully deploy interventions such as automatic exposure notification in smaller, targeted, communities could lead to a major health impacts in this and future pandemics. Smaller communities also tend to be more homogeneous, and it might be more feasible to achieve high adoption and compliance rates with such groups. In this talk we will report some of the recent results from our modeling efforts that compare the effectiveness of automatic exposure notification to simpler quarantine and testing strategies and show how the performance of devices used for tracking encounters in small communities is important.

Understanding and Rewiring Epidemic Networks: A Data-driven Approach Towards Enabling Quarantine in-Motion

Radu Marculescu
Department of Electrical & Computer Engineering
The University of Texas at Austin
radum@utexas.edu

In 2020, the world grinded to a halt to fight against an invisible enemy: **COVID-19**. The virus traveled swiftly through today's highly connected world. Though we find ourselves still actively trying to mitigate the current pandemic, one wonders how can we learn from our mistakes and grow resilient to other infectious diseases in the process?

Imagine yourself being able to visualize the risk of exposure to a virus while on your commute to work or while running errands around the town. Imagine that you and your friends are able to know first-hand when any possible infectious disease is spreading throughout your local community, even *before* local authorities begin raising alarms. Imagine knowing exactly *when* and *where* to avoid going to spend your evenings or weekends. If a mobile application gave you a safer route to lower your exposure to an infectious disease, would you take it?

This kind of application is precisely what the aim of this project is. More precisely, based on GPS mobility data, we try to identify the safe paths during an epidemic and minimize an individual's exposure to a virus. Similar to a navigation app that detours users away from traffic accidents, this application can direct users to avoid exposure to a virus.

To this end, this project targets human mobility at micro-, meso-, and macro-scale via a data-driven and machine learning-based approach. In other words, we are concerned with people's mobility in relatively small areas such as buildings or shared pedestrian ways, such that the trajectory and speed of a person changes due to crowd dynamics, traffic congestion, social interactions, or various other reasons. At mesoscale, we are concerned with people's mobility across larger areas, say an entire neighborhood or section of a city. Finally at macro-scale, we are looking at social sensing and phenomena induced by mobility over large areas, e.g. across a city or even an entire county.

As we are continue our research in this project, we hope to discover how modern mobility can grow resilient to infectious disease via trajectories and contact networks rewiring. When fully deployed, this can be a gamer changer as it would enable what we call "Quarantine in Motion", meaning a way of letting people go by their business while navigating the neighborhoods (and the city) in the safest possible way.

Interoperable Privacy Preserving Digital Contact Tracing

Yaling Yang (Virginia Tech, yyang8@vt.edu)

During a pandemic, contact tracing is an essential tool to drive down the infection rate within a population. To accelerate the laborious manual contact tracing process, smartphone-based digital contact tracing (DCT) tools have been proposed. These systems estimate proximity either through geolocation (e.g. Global Navigation Satellite System (GNSS)), or wireless technologies like Bluetooth.

Since the onset of COVID-19, various DCT systems have been developed and deployed around the world. A majority of them identify contacts using Bluetooth rather than using GNSS due to privacy concerns. However, if the privacy issue can be solved, geolocation-based DCT systems will offer two major advantages over Bluetooth-based systems.

First, geolocation-based DCT systems are compatible with manual contact tracing, whereas Bluetooth-based systems are not. As of today, most of the contact tracing work is carried out manually. Contact tracers work with patients to identify their close contacts and the places they visited a few days before the onset of symptoms [7]. Apart from obtaining identities of close contacts from patients, manual contact tracing also identifies locations patients visited. This verifies contacts whom the patient does not personally know, such as a cashier in a supermarket or a server in a restaurant [2, 4]. Unfortunately, Bluetooth-based and other proximity-sensing-based DCT systems do not collect such absolute location data. Hence, they can neither benefit from nor aid manual contact tracing.

In comparison, geolocation-based DCT systems and manual contact tracing can exchange location information for mutual benefits. For example, a geolocation-based DCT can provide exposure risk levels in public areas (e.g. restaurants and grocery stores) to help manual contact tracing to identify high-risk service workers. A contact tracer can share a patient's past visited locations to a geolocation-based DCT system even if the patient did not use the DCT app before the onset of symptoms. By getting additional patient histories from manual contact tracing, geolocation-based DCT systems can warn more users and may achieve a reduction in disease spreading at a lower app adoption rate. This makes geolocation-based DCT systems useful under a gradual deployment timeline, and in a demographic where a large portion of the population does not have the app. The latter scenario reflects reality. In the US for example, only 53% of US seniors older than 65 own smartphones [14]. Yet, this age bracket is the most vulnerable to the disease.

Second, geolocation-based DCT systems can inform the health authorities about the spread of the disease in a particular location. By aggregation of patients' trajectory history, the health authorities may learn the disease's spreading pattern within a community, and make informed decisions such as allocating more medical resources to hot-spot areas , announcing lockdown for a specific region instead of the whole state, or warn about a recent high-risk activity to a local community. Unfortunately, Bluetooth-based DCT systems cannot provide such information to the health authorities.

Challenges: Despite the above obvious benefits of geolocation-based DCT systems, they face two major design challenges:

(1) How can a DCT system preserve all users' privacy while recording their past trajectories?
(2) How to minimize the effects of GNSS localization error in contact tracing, especially in complex environments like indoor or urban canyon environments?

The first challenge is critical since privacy concerns have greatly hindered the wide adoption of DCT apps [8, 13]. We consider two perspectives: (a) trajectory privacy that protects a user's visited places and any personal information inferred from the trajectory histories; (b) exposure privacy that hides the user's exposure risk provided by the DCT app. To protect users' trajectory privacy, we propose to use k-anonymity to mix users' real trajectories with synthesized trajectories, such that users' real trajectories are hidden from eavesdroppers and other parties in the DCT system [3, 10, 12]. Next, we protect users' exposure risk through an efficient multi-party computation (MPC) design. MPC is a secure computation technique, where a function is computed by multiple parties jointly over their input, while keeping these inputs and the output private. However, designing an MPC-based DCT protocol is nontrivial since MPC for general functions is currently not scalable due to its huge computational complexity [5]. To enable our MPC scheme to compute trajectory intersection efficiently even when a large number of synthesized trajectories exist, we customized the MPC scheme with a partial-homomorphic Paillier cryptosystem [9], where the computation of location intersection and exposure risk is decomposed into a small set of homomorphic operations.

For the second challenge, we propose to integrate the Bluetooth technology with our geolocation-based DCT system, as Bluetooth has a better accuracy of proximity detection than GNSS in indoor environments. However, existing Bluetooth-based methods [1, 6, 11, 15] cannot fit into our geolocation-based design. A naive integration of Bluetooth and geolocation-based DCT system may cause a privacy risk which may be exploited by curious users to deanonymize patients' identities using exposure locations. To prevent this problem, we designed our Bluetooth protocol to be compatible with our geolocation-based DCT system, which maintains k-anonymity based user privacy.

Overall, in this paper, we propose KHOVID: a ***H***ybrid DCT system that enhances geolocation-based contact tracing with Bluetooth while preserving user privacy using ***K***-anonymity. In KHOVID, degradation of user privacy is avoided by building a customized MPC mechanism using the Paillier cryptosystem. In addition, to increase the accuracy of contact detection in complex environments, we combine the geolocation-based method with Bluetooth technology by designing a new Bluetooth DCT protocol. KHOVID offers accuracy and availability for multiple physical environments, interoperability with manual contact tracing, and the preservation of the privacy of all users in the system.

REFERENCES

[1] Apple and Google. 2020. Exposure Notifications: Using technology to help public health authorities fight COVID-19. Retrieved Dec 11, 2020 from https://www.google.com/covid19/exposurenotifications/

[2] BBC. 2020. Coronavirus: How Does Contact Tracing Work? *BBC News* (2020). https://www.bbc.com/news/explainers-52442754

[3] V. Bindschaedler and R. Shokri. 2016. Synthesizing Plausible Privacy-preserving Location Traces. In *Proceedings of 2016 IEEE Symposium on Security and Privacy (SP)*. IEEE.

[4] USA CDC. 2020. Key Information to Collect During a Case Interview. (2020). https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/keyinfo.html

[5] Y. Dou, K. Zeng, H. Li, Y. Yang, B. Gao, K. Ren, and S. Li. 2016. p^2-SAS: Privacy-preserving Centralized Dynamic Spectrum Access System. *IEEE Journal on Selected Areas in Communications* (2016).

[6] Inria and Fraunhofer AISEC. 2020. ROBERT-proximity-tracing/documents. Retrieved Dec 11, 2020 from https://github.com/ROBERT-proximity-tracing/documents

[7] L. Jaffe. 2020. What's It Like to Be a Contact Tracer? *Probulica Illinois* (2020). https://www.propublica.org/article/whats-it-like-to-be-a-contact-tracer-we-spoke-with-3-to-find-out

[8] C. Lehmann. 2020. Privacy Concerns Hindering Digital Contact Tracing. *WebMD* (2020). https://www.webmd.com/lung/news/20200928/privacy-concerns-hindering-digital-contact-tracing

[9] P. Paillier. 1999. Public-key Cryptosystems Based on Composite Degree Residuosity Classes. In *Proceeding of International conference on the theory and applications of cryptographic techniques*. Springer, 223–238.

[10] L. Pappalardo and F. Simini. 2018. Data-driven Generation of Spatio-temporal Routines in Human Mobility. *Data Mining and Knowledge Discovery* (2018).

[11] ppep pt.org. 2020. PEPP-PT Pan-European Privacy-Preserving Proximity Tracking. Retrieved Dec 11, 2020 from https://github.com/pepp-pt/pepp-pt-documentation

[12] L. Sweeney. 2002. k-anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* (2002).

[13] C. Timberg and D. Harwell. 2020. Government Efforts to Track Virus Through Phone Location Data Complicated by Privacy Concerns. *The Washiongton Post* (2020). https://www.washingtonpost.com/technology/2020/03/19/privacy-coronavirus-phone-data/

[14] C. Timberg, D. Harwell, and A. Safarpour. 2020. Most Americans Are not Willing or Able to Use an App Tracking Coronavirus Infections. That's a Problem for Big Tech's Plan to Slow the Pandemic. Retrieved Dec 11, 2020 from https://www.washingtonpost.com/technology/2020/04/29/most-americans-are-not-willing-or-able-use-an-app-tracking-coronavirus-infections-thats-problem-big-techs-plan-slow-pandemic/

[15] C. Troncoso, M. Payer, J. Hubaux, M. Salathé, J. Larus, et al. 2020. Decentralized Privacy-preserving Proximity Tracing. *arXiv preprint arXiv:2005.12273* (2020).

Function Secret Sharing for PSI-CA: With Applications to Private Contact Tracing

Steve Lu, Stealth Software Technologies, Inc., steve@stealthsoftwareinc.com, CEO

We present a new privacy-preserving framework for proximity/risk aware contact tracing based on new cryptographic constructions. In this work, we describe a token-based solution to contact tracing via Distributed Point Functions (DPF) and, more generally, Function Secret Sharing (FSS). The key idea behind the solution is that FSS natively supports secure keyword/token search on raw sets of keywords without a need for processing the keyword sets via a data structure for set membership. Furthermore, the FSS functionality features a linear "homomorphic encryption" property which enables aggregating up numerical payloads associated with multiple matches without additional interaction between phones and servers. Combined with new models to represent proximity/risk as these numerically aggregated values and a secure process from phone-to-server, these features combined make an attractive tool for lightweight privacy-preserving searching on a database of tokens belonging to infected individuals.

Modelling multipath interference for BLE proximity detection and exposure scoring

Dr Ramsey Faragher,
CEO Focal Point Positioning,
Fellow in Computer Science, Queens' College, University of Cambridge

This talk will cover the development of, and some results from, a physics-based BLE signal simulator that can accurately model the impact of multipath interference on BLE signal strength measurements. This is one of the major sources of error for BLE measurements, impacting the accuracy of proximity and exposure measurements, and leading to false positives and false negatives. The talk covers findings from the simulations that can improve contact tracing apps.

The author is an expert in this field, with almost 1000 citations on the topic of BLE multipath interference and its impact on distance and positioning measurements.

The presentation will cover:

- How the simulator models multipath interference for BLE signals
- Experimental validation of the simulator's accuracy using real-world datasets
- Findings of the simulator when testing GAEN settings used by smartphones today
- Recommendations for improving BLE-based proximity detection

Experimental validation was performed by comparing the simulator output to real test data from both University of Cambridge and Trinity College Dublin. These confirm that the simulator can reproduce key signal behaviours and power variations that are measured on real tests.

The simulator can rapidly produce tens of thousands of realistic datasets in a few seconds that would take tens of years to experimentally gather in a range of different interior layouts and building materials (metal, wood, plastic, etc). This allows for rich sensitivity analyses to be performed to determine the best settings for contact tracing apps using BLE (i.e. sampling rates, number of measurements per exposure metrics, exposure thresholds, etc) when considering the variations in signal power at various ranges in various environments due to multipath interference. We can therefore investigate the key factors in reducing false positives and false negatives. An example analysis of BLE settings for a dynamic covid exposure will be discussed, The simulator is used to demonstrate false positive and false negative rates expected when using three different measurement rates for BLE contact tracing apps - 15 seconds (original NHSx app approach), 2.5 minutes (GAEN highest rate) and 5 minutes (GAEN lowest rate). The simulator produces similar findings on false positive and false negative rates to those that were published by the NHSx team in a BBC news article (https://www.bbc.co.uk/news/technology-53765240), again confirming the ability of simulations to help to assess the real-world behaviours of BLE based proximity detection schemes.
The talk will conclude with some recommendations on how best to use BLE in contact tracing apps, based on these studies and the findings from the simulator

NIST event: Challenges for Digital Proximity Detection in Pandemics: Privacy, Accuracy, and Impact

Area: Epidemiological modeling of the efficacy of electronic proximity detection

Technical Abstract

COSMOS Testbed – Proximity Detection and Social Distancing Estimation in COVID-19 Pandemic

Corresponding Author: *Prof. Zoran Kostic, zk2172@columbia.edu*

Authors; Mahshid Ghasemi, Zhengye Yang, Hongzhe Ye, Zihao Xiong , Mingfei Sun, Mingei Sun, Zoran Kostic, Javad Ghaderi, Gil Zussman. Electrical Engineering Dept., Columbia University, New York.

Smart-city intersections will play a crucial role in automated traffic management and improvement in pedestrian safety in cities of the future. They are also a source of information about pedestrian density and behavior, data which can be critical in assessing and controlling pandemics such as COVID-19.

The Cloud Enhanced Open Software Defined Mobile Wireless Testbed for City-Scale Deployment (COSMOS) which is being built in Harlem-New York, enables research on technologies supporting smart cities. It deploys a variety of infrastructure sensors, including street-level and bird's eye cameras, whose data is aggregated and processed by AI-enabled servers. COSMOS pilot node at the intersection of Amsterdam Avenue and 120th street has been used to experiment with video-based applications focused on pedestrian detection, tracking, density estimation, and corresponding data processing. We have used this capability to collect and process data about pedestrians during the COVID-19 pandemic, and to understand their behavior.

Two video-based approaches to monitoring pedestrians in a traffic intersection have been developed – one based on street-level cameras and the other one based on bird's eye's cameras. We designed fully automated multi-stage social distancing analyzer pipelines, which utilize: (i) customized deep-learning based object detection, (ii) high accuracy object tracking algorithms, and (iii) original methods for assessing the proximity of pedestrians and the likelihood of belonging to safe social groups. Whereas the goals of the two approaches and technical problems are the same, a number of details provide different challenges in the domains of object detection and tracking, and social distancing analysis. Bird's eye methodology provides inherent privacy preserving features.

Video data of the pilot intersection collected during many months prior and during the pandemic has been processed using the two pipelines. We have extracted the statistics in form of pedestrian proximity histograms, time-variant changes, and social distancing groups. The obtained results quantify the intuitively-expected changes in pedestrian behavior pre-pandemic and during the pandemic. The obtained detection accuracies are respectable for social distancing applications. New insights in the behavior of social groups have been gained. Further development of the pipelines can facilitate not only real time assessment of pandemic-related situations in dense metropolitan areas, but also means for alarming the pedestrians and authorities in case of overcrowding which is unacceptable from the perspective of managing of the pandemic.

Title: A Simplistic Machine Learning Approach to Contact Tracing
Presenters:
Carlos Gómez, ML-Labs, Technological University Dublin, carlos.g.tapia@myTUDublin.ie
Niamh Belton, ML-Labs, University College Dublin, niamh.belton@ucdconnect.ie
Boi Quach, ML-Labs, Dublin City University, mai.quach3@mail.dcu.ie
Jack Nicholls, ML-Labs, University College Dublin, jack.nicholls@ucdconnect.ie
Devanshu Anand, ML-Labs, Dublin City University, devanshu.anand2@mail.dcu.ie
Target Area: Machine Learning Algorithms

The SFI Centre for Machine Learning (ML-Labs) arranged teams of ML-Labs PhD students to take part in the Too Close for Too Long (TC4TL) challenge organised by NIST and PACT. The objective of the challenge was to calculate the distance between two phones given phone instrumental data. By hand crafting features and employing either a Multi-Layer Perceptron (MLP) or a Gradient Boosting Machine (GBM), our supervised classification method resulted in state-of-the-art performance.

We harnessed the data provided by NIST of almost 25,000 event files, where each event was the data collected between two devices over a period of time. An event file was accompanied by a ground truth maximum distance label. The data collected included Receiver Strength Signal Indicator (RSSI) values and data collected from other instruments such as accelerometers and gyroscopes.

The most significant finding of the project was a distance formula for estimating the distance between two devices using Bluetooth data. This is a known formula in the domain. Given two phones, it is possible to compute a distance estimate **d'** using equation (1) where RSSI is the RSSI signal measured by the receiving phone, N depends on environmental factors ranging from 2 (good conditions) to 4 (bad conditions) and measured Power (TX) indicates the 1 meter RSSI. The optimal values of TX and N for fine and coarse grain events were found using a greedy search.

$$\mathbf{d'} = 10^{\frac{TX-RSSI}{10*N}} \tag{1}$$

We built two models, an MLP and a GBM that were trained on the predicted distance feature and additional categorical features such as the transmitter and receiver device. Both models were tuned and were evaluated using the software provided by NIST. The results are shown in the table below.

	Fine Grain 1.2 m	**Fine Grain 1.8m**	**Fine Grain 3m**	**Coarse Grain 1.8m**	**Average nDCF**
GBM (proposed)	0.6	0.52	0.58	0.37	**0.5175**
MLP (proposed)	0.61	0.48	0.55	0.44	0.52
Contact-Tracing-Project	0.68	0.54	0.59	0.41	0.555
LCD	0.6	0.58	0.63	0.55	0.59

Given the significant architectural differences of the MLP and GBM, these results demonstrate how a supervised classification technique trained on the aforementioned handcrafted features is a well-founded approach to the problem. In addition to achieving state-of-the-art performance, the training time for both models is less than thirty seconds on a CPU. The success of this simplistic approach can be mainly attributed to the 'predicted distance' variable which is based on received RSSI signals. Both models saw a drastic performance improvement when adding this handcrafted feature. This feature interacted with variables relating to the transmitter and receiver devices to further improve the accuracy of the model. This highlights how a fusion of machine learning and domain knowledge can result in superior performance.

The Feasibility of Co-location Detection through a Deep Learning Fusion of Mobile Sensors

Sheshank Shankar
PathCheck Foundation
sheshank.shankar@pathcheck.org

In this talk, we discuss the feasibility of effective proximity sensing through deep learning models trained on temporal Bluetooth Low Energy (BLE) and mobile sensor signals. This is in the context of the PathCheck team's submission to the NIST Too Close for Too Long (TC4TL) Challenge. Provided two datasets (further referred to as the TRAIN and TEST datasets) containing the mobile sensors logs from multiple mobile co-location experiments, we trained a temporal one-dimensional convolutional network. This model achieved an average normalized decision cost function (nDCF) value of 0.6425, placing 3rd in the challenge.

A major challenge we encountered across our experiments is the lack of generalization in deep learning predictors. Since the TRAIN and TEST datasets were collected through the same protocol by different organizations in different environments, accurate predictions on both datasets are a good sign of a generalizable model. Even with complex architectures (such as LSTM, GRU, CNN/RNN hybrids) and data representations, models are not able to achieve a high level of accuracy on the other dataset. To assess the generalizability of the datasets, we switch the TRAIN and TEST datasets, but it does not yield any notable improvement in generalization. Informed by these results, we try to estimate the gap between the two datasets. The inconsistencies between the two distributions can be observed by the high values of $\ell 2$ between the pairwise distances of any two feature vectors. Looking at the cross-dataset nearest neighbour pairs, we find that a significant number of the pairs had significantly different distances. Further proving this data discrepancy, we receive similar results when training on an optimal training subset that includes only the 2 nearest points in the training dataset for each point in the test dataset is included. We also find that the average distance between a nearest-neighbor pair is fairy low; however, if we isolate nearest-neighbor pairs with different classes, the average distance is around 8 times as much! This implies that the two data distributions (although collected through the same protocol) are not similar enough to capture any generalizable information. A thorough statistical analysis would be needed to confirm this argument, as highly non-linear manifolds that can fit both distributions may exist.

To better understand the potential for an accurate predictor in these two data distributions, we performed low dimensional projections of the TRAIN and TEST datasets to identify underlying target class clusters in the feature space. These plots showed heavily overlapping clusters, without a clear decision boundary between any two distance buckets (1.2, 1.5, 3, 4 meters). We suspect that this could be caused by external factors such as, physical barriers present between phones, the number and time spread of observed chirps, and multi-path signals reflected from surfaces (e.g. indoor vs outdoor). It is, however, possible that higher dimensional hyper-planes dividing the classes exist. Lastly, we briefly discuss potential solutions to these challenges, including hybrids of other co-location modalities.

Further Reference: S. Shankar, R. Kanaparti, A. Chopra, R. Sukumaran, P. Patwa, M. Kang, A. Singh, K. McPherson, R. Raskar. Proximity Sensing: Modeling and Understanding Noisy RSSI-BLE Signals and Other Mobile Sensor Data for Digital Contact Tracing, 2020.

Presenter: Nicholas Maynard
Affiliation: The MITRE Corporation
E-mail address: nmaynard@mitre.org
Title: Entropy Based Discretization's and Weight Optimization for Configuring the GAEN system.

Abstract:

The Google-Apple Exposure Notification (GAEN) system, currently in use aims to provide a meaningful improvement to the failures of manual contact tracing. Namely it allows a public health authority (PHA) to take advantage of Bluetooth signals emitted from cell phones to determine if an exposure occurred rather than relying on the memory of the infected individuals. In addition to this capability, the GAEN system is privacy preserving for the individuals who decide to opt into it.

While this tool has the capability to enhance manual contact tracing, it currently requires the setting of a configuration file which consists mostly of sensor thresholds/weights for the bins created by these thresholds. In this regard, there are limited tools available for setting a configuration file other than opening notepad and looking through the file itself to make the change. Additionally, there is limited insight in how to translate guidelines set by a public health authority (say less than 6ft for 15 minutes) into the most ideal corresponding configuration values. This is where we step in in an attempt to bolster the tools available for setting a configuration file, as well as assist a PHA in making informed decisions to best approximate their definition of too close for too long.

In this talk, we will first present the current data available for use as pulled from the datasets created by Mark Krangle. Next, we will go into the overall approach/tools built for configuring GAEN to a PHA's needs based on real sensor data. More specifically we will detail how a PHA may translate tracing based on time spent in various distance ranges into proper sensor values through a method called entropy-based discretization's. Additionally, we will detail an optimization approach taken to set the weights of the attenuation bins based on simulated interactions sampled from the datasets generated by Mark Krangle. Lastly, we will run through an example from start to finish using said tools suite to get an idea of how a PHA would interact with the tools developed and get an idea of what sort of false positive/false negative rates they could expect.

With the ability to make informed sensor settings a PHA should be able to take full advantage of the GAEN system in making accurate predictions of risky interactions and hopefully use the knowledge to inform the people who need it the most.

Presenter: Brian Thompson
Affiliation: The MITRE Corporation
E-mail address: bthompson@mitre.org
Title: Efficacy of Current Approaches and An Alternative Paradigm for Digital Contact Tracing

Abstract:

The Google-Apple Exposure Notification (GAEN) system, currently being used by at least 33 countries and 17 U.S. states and territories, aims to complement and overcome some of the shortfalls of manual contact tracing. However, evidence indicating the effectiveness of GAEN, in terms of reducing the number of new infections, has been minimal, even after taking low uptake rates into account. As more countries and states are making decisions about whether to invest resources into pushing out GAEN to their constituents, understanding its potential impact as well as its limitations is crucial.

A common approach to measure the performance of a digital contact tracing mechanism is by its ability to detect "close contacts", typically based on the WHO or CDC's definitions. By those definitions, most contacts do not lead to infection — estimates from the literature suggest a false alarm rate upwards of 90%, and likely significantly higher. Such a high rate of notifications to uninfected users is likely to drive down app usage and compliance. GAEN aims to make more nuanced determinations of exposure risk by leveraging a granularity of data that is not available to manual contact tracers, but its accuracy at notifying only infected users does not seem to be adequately assessed by existing analyses. Studies that examine the impact of digital contact tracing on population-level health outcomes often assume high compliance of app users with quarantine recommendations and do not consider the consequences of a high false alarm rate on compliance.

In this talk, we will first present the results of a modeling study that explores the efficacy of digital contact tracing using simulated interactions between infected and susceptible individuals carrying GAEN-enabled mobile devices. A key finding is that, although improvements in the accuracy of proximity estimation methods can increase the efficacy of digital contact tracing, even perfect estimators cannot yield a high detection rate of secondary cases without also having a high false alarm rate, which limits the potential impact of such technologies.

Finally, we propose an alternative paradigm for digital contact tracing that bootstraps the GAEN framework to detect and mitigate the effect of super-spreading events on viral spread using "lateral" rather than "forward" contact tracing. Our proposed approach has the potential to overcome fundamental limitations on the efficacy of the current paradigm, can have impact at lower adoption rates, and works even if the infected person does not have the app. Mitigation strategies that include manual contact tracing, the current approach to digital contact tracing implemented by GAEN, and our proposed paradigm for detecting super-spreading events may present the best opportunity to finally control the COVID-19 pandemic.