REPORTS OF
THE MACHINE LEARNING AND INFERENCE LABORATORY

MLI

TOWARDS WI-FI CONTACT PREDICTION: METHOD AND INITIAL RESULTS

JANUSZ WOJTUSIAK
VARALAKSHMI VAKKALAGADDA
YING WANG
SRI SURYA KRISHNA RAMA TARAKA NAREN DURBHA
FARROKH ALEMI
AMIRA ROESS

MLI 21-1
JANUARY 2021

Towards Wi-Fi Contact Prediction: Methods and Initial Results

Janusz Wojtusiak, jwojtusi@gmu.edu
Varalakshmi Vakkalagadda, vvakkala@gmu.edu
Ying Wang, ywang86@gmu.edu
Sri Surya Krishna Rama Taraka Naren Durbha, sdurbha@gmu.edu
Farrokh Alemi, falemi@gmu.edu
Amira Roess, aroess@gmu.edu

Abstract

The presented study aims at making Wi-Fi data usable in practice for aiding contact tracing. The focus is on an approach that requires minimum preparation to be deployed over existing infrastructure. The work is intended to create a decision support tool that provides a ranked list of people potentially in contact with an infected individual. It provides a contact score that takes into consideration the likelihood of exposure, time of exposure and specific location of exposure. Increasingly complex approaches to predicting location from enterprise-level Wi-Fi data logs are studied to understand the need for data and its effects on the accuracy of results. The methods are tested using 100 scenarios completed by study participants within one building, as well as the participants' simulated contacts. The most advanced method that predicts the movement of individuals based on Wi-Fi access point locations and building floorplan achieves the best results (AUC 0.82-0.89 in predicting contacts). The results indicate that Wi-Fi data can aid the case interview process by providing a list of potential contacts. The limitations of Wi-Fi data and modeling also indicate that the approach should not be used as the only method of identifying potential exposures and only aid the traditional process of eliciting contact.

1. Introduction

Contact tracing can reduce transmission if exposed individuals are identified as soon as possible following contact with an infected person. The faster and more thorough contact tracing is, the smaller the number of people infected. Rapid contact tracing is an integral part of any strategy to contain epidemics [15]. Unfortunately, contact tracing in the COVID-19 context has been hampered by a reliance on traditional, labor and time intensive methods. Coupled with the sheer volume of cases and the high transmissibility of the virus, public health agencies are simply unable to respond to COVID-19 effectively. Instead of relying on public health agencies, large organizations will need to take over these efforts and complete in-house contact tracing. Because of the above reasons, COVID-19 has highlighted the need to develop and implement more rapid and innovative contact tracing protocols that integrate new technologies. Private and public employers alike are still looking for rapid and cost-effective solutions to contact tracing. This research provides a strategy that large organizations can follow to aid effective contact tracing by helping identify individuals that need to be contacted for follow up by the personnel conducting contact tracing. The solution is intended to aid and supplement the traditional process of eliciting contacts while conducting case interviews.

1.1. Contact Tracing vs. Contact Prediction

Contact tracing is a well-defined term in epidemiology and public health [7][11][18]. It consists of several steps: from contact identification to notification to monitoring and follow up [15]. In that process, contact identification is typically done by the infected patient during interview process. The presented work focuses on contact identification (a.k.a. contact elicitation) that aims at identifying those who were in contact with the infected individual. The term ***contact prediction*** stresses the fact that contacts are identified (predicted) through computational methods. Results are intended to be used as part of contact tracing protocols.

1.2. Wi-Fi vs. Bluetooth Protocols

The focus of contact tracing technology has been on close contact identification using Bluetooth technology, or more specifically, Bluetooth Low Energy (BLE). The main advantage of network-based Wi-Fi contact tracing is that it requires no downloads of applications and no requirements for users to turn on Bluetooth on their mobile devices. Bluetooth monitoring is a protocol set up by Apple and Google on iOS and Android phones to alert exposed individuals without reporting the identity of infected individuals. Bluetooth monitoring typically requires download of an application as well as continuous use of Bluetooth, neither of which are likely in the entire population or even a large proportion of population. In the recent version of Apple's iOS system, a new feature called exposure notification is available. It requires opt-in by phone users. It works only if an infected person has the option enabled in their phone and report their infection within the app. Additionally, the BLE does not continuously monitor contacts, but only at specific time intervals, typically every 5 minutes.

The presented approach completely relies on the use of Wi-Fi network connection logs that are routinely collected by network infrastructure. It requires no need to make any modifications to the network which may be costly and time consuming. It only requires extraction of the connection logs from network and securely transmitting them for analysis. The information included in the log file includes timestamps of users' association and disassociation with an access point. Received signal strength indicator (RSSI) information is not required in the present approach.

1.3. Contributions of the Presented Study

The premises behind the presented network-based Wi-Fi approach are that:

1. There is no need for users to install or enable anything on their phones. Simply use the existing Wi-Fi networks the way they usually do in an enterprise environment. Specifically, this work refers to enterprise Wi-Fi as network that provide individual user-based authentication.
2. The approach focuses on predicting the actual locations of individuals and reconstructing their movements. This is in contrast with approaches focused on detecting contacts relying on being connected to the same access point.
3. The approach helps organizational personnel conducting contact tracing to identify potential contacts of an infected person in addition to traditional interviews. It is not intended to replace the interviews. The focus is on developing methods that are implementable in practice with minimum effort.
4. The approach ranks potential contacts by the calculated score, which takes into consideration the likelihood of contact and duration of the contact.

The study presents five approaches to modeling Wi-Fi data. Using collected real movement data of 12 study participants, it shows that the approaches that rely on location prediction are superior to naïve ones. Further, large number of experiments have been conducted to test the approaches on simulated contacts between individuals.

2. Background

Technological support for contact tracing is not new and has been the focus of significant work within the past years, and more intensely since the COVID-19 pandemic started. Reviews of available contact tracing technologies have been presented in the recently published [15][3][6].

Digital contact tracing has existed, at least since 2007 [3][6][15], but became more prominent after the COVID-19 pandemic [4][20]. Apple and Android phones have proposed a Bluetooth protocol that is in use at this time. The Bluetooth protocol has a number of advantages over personal contact tracing. It is automated and faster than personal contact tracing. It maintains the identity of an infected person and only alerts exposed individuals. It can work on phones that have Bluetooth connectivity but requires individuals to turn this feature on. Several countries have gone through large scale implementation of digital contact tracing using Bluetooth protocol. The code for these efforts is available publicly and thus one can adopt and use it rapidly. There has been widespread criticism of Bluetooth protocol primarily because (a) it relies on a feature of the phone which is typically turned off to conserve battery and privacy, (b) it adds a new layer of surveillance, raising concerns with government intrusion in private lives [6] ,(c) to conserve phone battery it uses low a temporal resolution of about 5 minutes (two individuals need to be in proximity at the exact moment the data is collected within that interval) and (d) the vulnerabilities over Bluetooth protocol itself increase the risk of privacy violations and security concerns. It is not surprising that the experience with digital tracing has shown poor adoption by the general population. For example, only 10% to 20% of the population of Singapore participated in Singapore's implementation of digital contact tracing [15].

There are two main approaches to Wi-Fi-based contact tracing and more generally Wi-Fi-based location prediction: (1) based on measurement of RSSI and triangulation of devices to estimate proximity of a device; and (2) based on the analysis of access logs. The approach (1) can provide accurate location data for individuals, but requires specialized apps installed on phones or significant modifications to network infrastructure. This makes it not feasible in the presented setting. The accuracy is also affected by the furniture layout in the building and the movement of people surrounded. The approach (2), taken in the presented study, requires no modifications to infrastructure or use of specialized apps. Instead, the focus is on advanced data modeling to predict the likely locations of individuals.

A comprehensive study of how Wi-Fi networks can be used for contact tracing has been presented by Trivedi [23]. This is the closest work to the approach presented here. The authors presented detailed results of simulating contacts in two large university environments. Graph algorithms are used for efficient modeling of contacts. The method used by the authors is equivalent to naïve approach N1 described later and does not model movement trajectories or actual locations when multiple rooms are covered by one access point or one room is covered by multiple access points.

Modeling and understanding the movement of humans in urban context has a wide range of applications [3] from planning the development activities like construction to controlling the spread of diseases. The RADIUS log file (Wi-Fi data) is converted to the Space Leaps (change in location that is the moving state of an artifact) to understand the movement of the artifacts in university life. Quest uses computationally secure protocols such as cQuest and iQuest data outsourcing methods to develop a system for contact tracing for the Organizational level using privacy-preserving presence algorithms[22]. These methods prevent adversaries from gaining knowledge of individual location history. The Wi-Fi data used in the study is generated using SNMP (Simple Network Management Protocol) traps.

Bluetooth Low Energy (BLE) received signal strength [2] used in the mobile handsets have a wide range of real-world settings. BLE is used to detect the close proximity of people from one another. The received signal strength of Bluetooth can vary by the orientation of the handsets and absorption/reflection of radio signals in buildings, and received signal strength may not reflect the distance between the transmitter and receiver. This implies that precise contact prediction within a specific distance using BLE is challenging. The other main drawback of this approach that the position of phone (that is in handbags, left or right side of human body) causes different attenuation of the signal. Therefore it is not possible to clearly distinguish how far people are located by RSSI values. Better accuracy can be attained by adopting social media protocols with contact tracing apps.

3. Methods

Prediction of likely contacts that can be used for contact tracing requires preparatory steps followed by the application of algorithms that estimate likely locations and contacts. First, computer-interpretable representations of floorplans and Wi-Fi network infrastructure need to be created in forms processible by contact prediction methods. While some investment in modeling environment is needed, the effort is not impossible for many organizations as long as a proper Wi-Fi network is in place and designated personnel have access to floorplans.

The following sections show the process of extracting information for one large building, and the estimated time needed to do so. The Wi-Fi connection data need to be extracted from network infrastructure on a regular basis (i.e., daily) allowing contact tracers access up to date information in the event of a need for contact tracing due to a public health emergency.

3.1. Floorplan and Infrastructure Modeling

There are a number of preparatory steps for the project. First floorplans need to be analyzed and converted to computer-processible form. More specifically, in the presented work we used graph representation of floorplans in which rooms are represented by nodes in the graph and doors/passages are represented by edges. Furthermore, the second type of nodes is added to the graph to represent Wi-Fi access points. These nodes are connected to the nodes representing rooms with edges that additionally carry signal strength information. The process is illustrated in Figure 1 in which floorplan and Wi-Fi AP locations are mapped to a graph.

In the presented study, we manually extracted the floorplan for one building. The extraction process was done by one person and verified by another. While relatively labor intensive, the work can be done within 6-7 hours for a large 500-room building with five floors. For organizations such as universities, it is not unusual that each large building has a facility manager that can be responsible for completing the work. With proper coordination between people, each responsible for their own building, the process can be completed in a relatively short time. Moreover, the extraction does not require any specific qualifications beyond the ability to read floorplans and the use of spreadsheet applications.

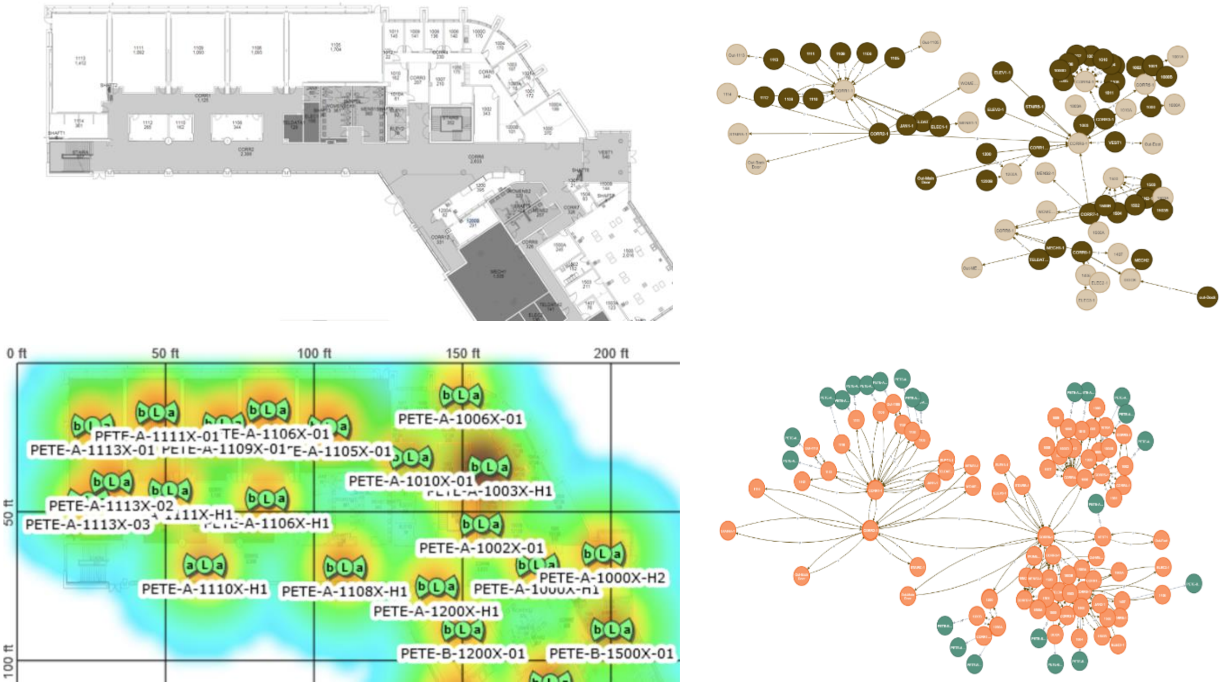


Figure 1: (top left) Floorplan of one floor of a building; (top right) graph representing the floorplan; (bottom left) locations and estimated strength of Wi-Fi access points; (bottom right) graph representing floorplan with added access points where orange color nodes represents rooms and corridors, green color represents access points.

The extracted data are in the form of a flat table that can be typed in Excel or a similar application. It consists of four columns: *location1*, *location2*, *passage-type*, and *locked*. Essentially, to model floorplan all we need is to model passages between locations. In this case, locations can be rooms, corridors or open areas as shown on the floorplan. In the undirected graph used, the order of location1 and location2 is irrelevant. *Passage-type* indicates door, double door, or hallway. Locked indicates if the doors are typically locked. Modeling which doors are locked allows for modeling access to different areas, and can be further complemented by information about access by specific individuals, if available. Connections between floors are modeled through staircases and elevators that link corridors of different floors. The graph can then be verified automatically and by another person (i.e., by searching for locations disjoint from others). There are published works on automated graph extraction from floorplans. For example, Schmitt et al.[22], presented an approach to routing graph generation from 2-dimensional floorplans. Such methods can be applied, but were not investigated in this study.

In the presented work, the mapped building is located at George Mason University’s Fairfax campus. It is represented by a graph with 529 nodes with 392 nodes representing rooms, 53 nodes representing passages and hallways, 3 staircases, 2 elevators, and 667 edges connecting the nodes. 98 Wi-Fi access points in the building are linked to locations within their coverage using 539 edges, which additionally carry information about estimated RSSI in the center of the building.

Based on the distance to an access point, theoretical RSSI can be calculated given the transmit power of access point:

$$RSSI = P_{tx} - L_{FS} - L_m$$

$$L_{FS} = -27.55dB + 20\log(f) + 20\log(d)$$

Where f is the frequency of Wi-Fi in unit of MHz, and d is the distance in unit of meter between the observing location and the access point. P_{tx} on average is 18dBm for most Wi-Fi access point. A majority of Wi-Fi access point in the building is at frequency of 5,000 MHz, some of them are 2,400 Mhz. L_m is miscellaneous losses including fading margin, wall penetration loss, body loss, and other losses. Based on the formula and the empirical measurement, the relationship between $RSSI$ and d at the same floor is simplified as shown in Figure 2. With across floor, an additional 25dB is added.

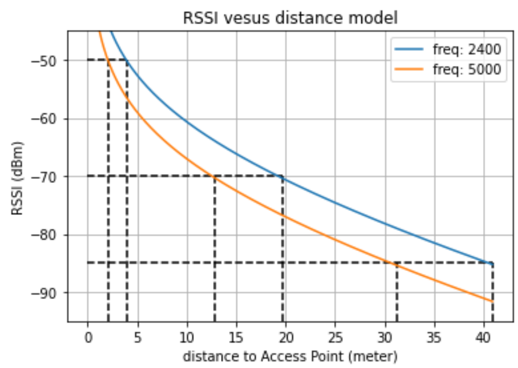


Figure 2: Approximate relationship between distance to assess point and signal strength.

In the presented work, floorplans with Wi-Fi locations were readily available in the Information Technology Office, and transcribed into a spreadsheet. However, one difficulty that was addressed in this work is that floorplans are “flat”, but Wi-Fi signal is 3-dimensional. Specifically, adjacent floor coverage was modeled with signal strength reduced by 25dB. This approach allows for turning 2-dimensional floorplans into a 3-dimensional Wi-Fi coverage model.

In addition to the theoretical coverage map, used in this study, one can construct empirical coverage based on the actual measurements. An example of real signal strength measurement is presented in Figure 3. The measurements have been taken in specific locations indicated in the figure using NetSpot software [26] while results are superior to theoretical coverage maps, this is a time-consuming process that is considered not feasible in the presented work that aims at the fast deployment of tracking systems.

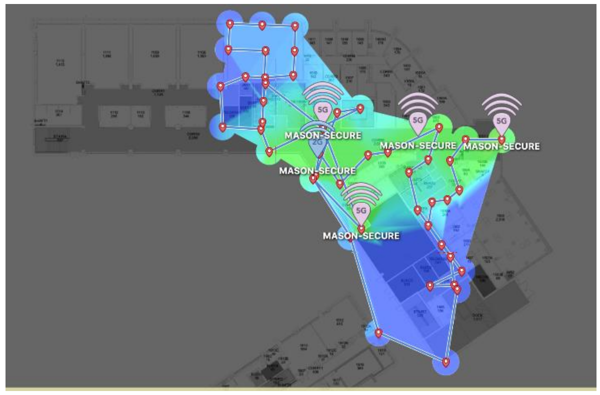


Figure 3. Empirical measurement of signal strength on one floor of a building.

3.2. Wi-Fi Data

Data used for technology-based contact tracing are collected by enterprise-level Wi-Fi networks. Each time a device is connected to a network, a record in a network log is created. Since the connection requires users to login using their credentials, it is possible to identify individuals connected at the same time. There is wide use of Wi-Fi networks in environments such as university campuses. Our data indicates that within a 31-day period in October/November 2020, the Wi-Fi network at George Mason University has been accessed by about 21,000 unique users. Interviews of selected students indicate that all of them use the campus Wi-Fi network. While this information is anecdotal and does not provide evidence of the Wi-Fi use, it is consistent with the high number of Wi-Fi users. A proper way of estimating Wi-Fi usage would be to count people entering the building in a given period and compare that data to the Wi-Fi usage.

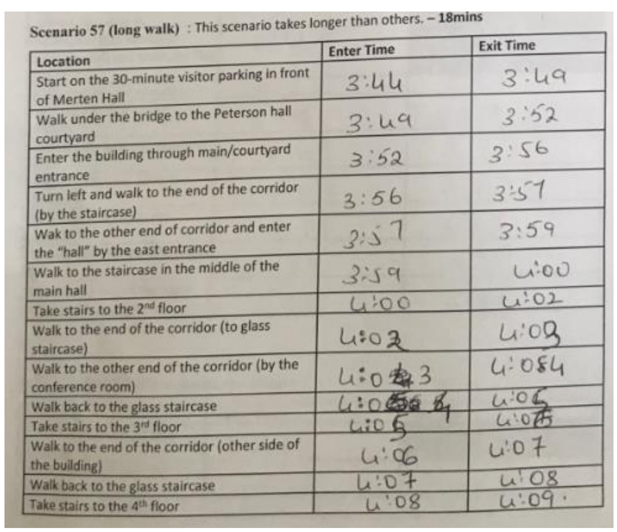
The Wi-Fi log data are relatively large. For example, the one month of data collected around October 2020 in a large building consists of approximately 50,000 rows. When in full operation (outside of pandemics limitations), one can estimate up to 200,000 rows in the log file per building per month.

Wi-Fi log data can be automatically transmitted to the analytic server. Within the infrastructure used for the presented study, network logs can be automatically emailed or securely transmitted through SFTP/SCP protocol on a scheduled basis, i.e., daily at a specific time. Then, the logs are automatically processed to include another day of data into the decision support system. If needed, logs can also be extracted manually by networking personnel if urgently needed to track a specific case outside of the daily schedule.

3.3. Scenario Data Collection

Development and evaluation of movement tracking methods require data in which the person's actual movements are known and compared to recorded Wi-Fi access information. In the presented work, 101 scenarios have been developed to account for different types of movements within a building. The scenarios were distributed to 12 study participants who recorded exact times they were in specific locations. In preparation for the study, the participants received temporary access to typically locked parts of the building (their keycards were programmed for access, see below). Most scenarios have been completed by 2-3 participants. Figure 4 shows an example scenario completed by a participant with handwritten times in specific locations. The times and location were then transcribed and converted into a data file. There are multiple typos and illegible values across the collected data, creating noise in the data used for modeling and evaluation.

Scenarios were randomly assigned to participants; thus, each scenario was intended to be performed by one, two, or three participants. However, three of the scenarios were not performed by any participants because of issues with access to certain parts of the building. This is because their keycards were not programmed to include all locations before the study began. In some cases, participants indicated that they waited for a certain time in front of a door they could not unlock. These were marked in data as variants of the original scenarios. In total, 158 scenarios were performed. When merged with the Wi-Fi log data (see section 3.4), 120 out of the 158 (76%) completed scenarios had corresponding Wi-Fi data. The distribution of assigned and performed scenarios by participants is shown in Figure 5. Further investigation of the reasons for missing scenarios indicates problems with phones not connecting during the entire study period for one participant, and during one day for another participant. For other participants, some of the missing scenarios are the ones that were performed outside the building (to test external Wi-Fi coverage). While the missing scenario data are invalid for the purpose of the study, they were used to calculate recall (sensitivity) of the Wi-Fi tracking, indicating that some people may be missing in the access data.



Scenario 57 (long walk) : This scenario takes longer than others. – **18mins**

Location	Enter Time	Exit Time
Start on the 30-minute visitor parking in front of Merten Hall	3:44	3:49
Walk under the bridge to the Peterson hall courtyard	3:49	3:52
Enter the building through main/courtyard entrance	3:52	3:56
Turn left and walk to the end of the corridor (by the staircase)	3:56	3:57
Wak to the other end of corridor and enter the "hall" by the east entrance	3:57	3:59
Walk to the staircase in the middle of the main hall	3:59	4:00
Take stairs to the 2nd floor	4:00	4:02
Walk to the end of the corridor (to glass staircase)	4:02	4:03
Walk to the other end of the corridor (by the conference room)	4:03	4:054
Walk back to the glass staircase	4:04	4:05
Take stairs to the 3rd floor	4:05	4:06
Walk to the end of the corridor (other side of the building)	4:06	4:07
Walk back to the glass staircase	4:07	4:08
Take stairs to the 4th floor	4:08	4:09

Figure 4: Handwritten scenario completion record from a study participant.

The average length of completed scenarios was 32.7+/-18.8 minutes and included 11.55+/-3.42 distinct locations. On average, participants stayed in a location for 86.79 sec (144.5 sec when excluding corridors and hallways). The summary of the scenario data is available in Table 1.

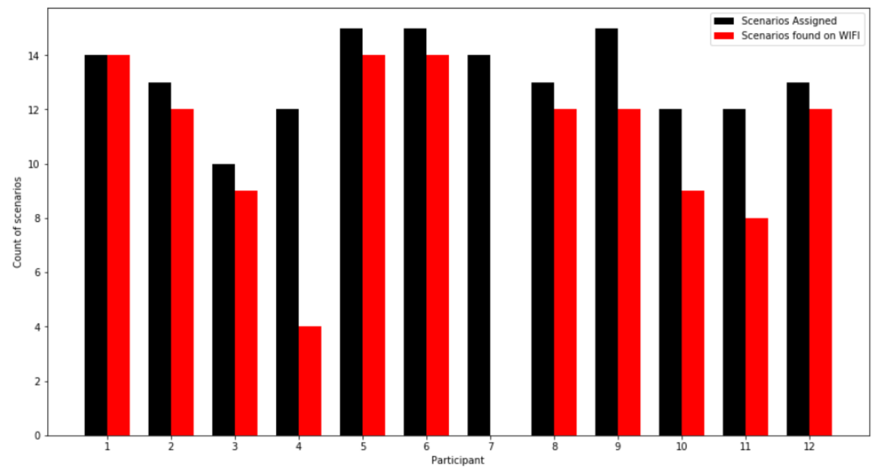


Figure 5: Completion counts of scenarios for 12 study participants.

Table 1: Summary of the collected scenario data.

	All scenario locations				Rooms only			
	Mean	Std	min	max	Mean	Std	min	max
# locations	19.85	7.21	7	48	2.71	2.3	0	10
# distinct locations	11.55	3.42	5	26	2.07	1.31	0	5
Scenario duration in (sec)	1962.28	1121.2	420	6360	653.33	539	0	2630
Time at location in (sec)	86.79	101.50	0.0	1800	144.5	124.8	0	935
# floors in scenario	2.45	1.4	0	5	-	-	-	-
# Access Points	3.19	1.58	1	9	-	-	-	-
# distinct Access Points	2.89	1.21	1	6	-	-	-	-
Time connected to AP in (sec) each scenario	1250.96	659.65	302	3336	610.63	538.85	0	2630
# locations connected to the RSSI: -50	2.29	1.92	0	10	0.61	1.1	0	6
# locations connected to the RSSI: -70	5.13	3.69	0	20	1.38	1.95	0	12
# locations connected to the RSSI: -85	8.85	4.97	0	22	2.6	2.63	0	13

The average number of access points connected to by participants in each scenario is not necessarily consistent with number of locations connected with in RSSI -50dBm, -75dBm, -80dBm because at every point of time the participant could either stay in a room or walk along corridors or hallways (this implies the participant is crossing multiple locations at a point of time). Furthermore, the scenarios were designed to emphasize movement of individuals. Thus, in real data one can expect people to stay at one location for much longer. Consequently, the real data are expected to be somewhat easier to analyze.

3.4. Linking Scenario and Wi-Fi data

To test the created methods, the scenario and Wi-Fi data were merged to allow for comparing real and predicted locations. The two datasets are merged by a participant ID and aligned by time. Since there is no one-to-one correspondence between scenario steps and the Wi-Fi dataset, additional rows of the data are created, as illustrated in Table 2.

Table 2: Illustration of the data merging process. The scenario data (left) are merged with Wi-Fi log data (center) and a resulting merged analytic file is created (right).

Scenario				Wi-Fi				Merged			
Location	Enter	Exit		Access Point	Connect	Disconnect		Location	Start	End	Access Point
Parking	15:58:00	16:00:00						Parking	15:58:00	16:00:00	
Out-East	16:00:00	16:01:00						Out-East	16:00:00	16:01:00	
VEST1	16:01:00	16:01:00						VEST1	16:01:00	16:01:00	
CORR6-1	16:01:00	16:01:00						CORR6-1	16:01:00	16:01:00	
ELEV1-1/ELEV2-1	16:02:00	16:04:00						ELEV1-1/ELEV2-1	16:02:00	16:04:00	
ELEV1-4/ELEV2-4	16:02:00	16:05:00		PETE-A-1108X-H1	16:03:16	16:08:27		ELEV1-4/ELEV2-4	16:02:00	16:03:16	
CORR4-4	16:05:00	16:06:00						ELEV1-4/ELEV2-4	16:03:16	16:05:00	PETE-A-1108X-H1
CORR2-4	16:05:00	16:06:00						CORR4-4	16:05:00	16:06:00	PETE-A-1108X-H1
4800	16:07:00	16:18:00		PETE-A-4800X-01	16:08:27	16:18:38		CORR2-4	16:05:00	16:06:00	PETE-A-1108X-H1
CORR2-4	16:18:00	16:18:00						4800	16:07:00	16:08:27	PETE-A-1108X-H1
CORR4-4	16:18:00	16:18:00						4800	16:08:27	16:18:00	PETE-A-4800X-01
ELEV1-4/ELEV2-4	16:18:00	16:19:00		PETE-A-4000X-01	16:18:38	16:20:00		CORR2-4	16:18:00	16:18:00	PETE-A-4800X-01
ELEV1-1/ELEV2-1	16:18:00	16:19:00						CORR4-4	16:18:00	16:18:00	PETE-A-4800X-01
CORR6-1	16:19:00	16:19:00						ELEV1-4/ELEV2-4	16:18:00	16:19:00	PETE-A-4800X-01
Out-Main Door	16:19:00	16:20:00						ELEV1-1/ELEV2-1	16:18:00	16:18:38	PETE-A-4800X-01
Parking	16:19:00	16:20:00						ELEV1-1/ELEV2-1	16:18:38	16:19:00	PETE-A-4000X-01
								CORR6-1	16:19:00	16:19:00	PETE-A-4000X-01
								Out-Main Door	16:19:00	16:20:00	PETE-A-4000X-01
								Parking	16:19:00	16:20:00	PETE-A-4000X-01

3.5. Data Modeling

Location modeling creates a list of likely locations in which a person was during a given period of time. In the presented work, the list of locations is constructed based on Wi-Fi access data as well as the floorplan graph described above. This can be further expanded by additional information such as movement patterns from historical data, known office locations, class registration and meeting rosters, door smartcard access logs, and dedicated sensors such as RFID readers.

Once lists of likely locations are created for all people who were in a building at a given time, these lists can be intersected to calculate chances and time of contact between individuals. Consequently, if a person tests positive for COVID-19 or any other highly contagious disease, a list of people with a possible contact to the infected individual can be created.

Naïve Approach 1: Common Wi-Fi Access Point (N1)

The simplest approach that can be used to list individuals that were potentially in contact is to check Wi-Fi connection logs for people connected to the same access point at the same time. The advantage of this method is that there is no need for any information about the location of access points, floorplans, or network structure. The approach simply calculates an intersection between connection logs of individuals. It is fast and can be readily deployed without any additional information about floorplans or AP locations.

The disadvantage of this method is a *possible low recall* (sensitivity) in identifying potential contacts. As previously shown in Figure 1, most locations within a building are covered by more than one access point, sometimes on different floors. The theoretical data used in this work shows an average of 3.61±2.01 rooms (this includes corridors, restrooms, staircases, and elevators) per access point at RSSI -50, 14.87±6.53 rooms per access point at RSSI -70, 73.98±26.31 rooms per access point at RSSI -85. Consequently, each room/location is covered by an average an 0.72±0.71 APs at RSSI -50, 2.95±1.55 APs at RSSI -70, 14.69±4.56 APs at RSSI -85.

Naïve Approach 2: In Building at the Same Time (N2)

Another naïve approach is to simply list people who were in the same building at the same time. The only knowledge about Wi-Fi access points needed is the building in which they are located. The method has *very high recall* (essentially 100% of people who connect to Wi-Fi, which is a very high percentage of all people in the building). The disadvantage of this approach is that it is impractical for use for contact tracing – potentially hundreds of individuals are in a building simultaneously.

To address deficiencies of the above naïve methods, the approach taken in this study is first to predict a set of likely locations in which individuals are (with associated probabilities), and then use those predicted locations to check who was likely to be at the same place at the same time. More specifically, the study aims at maximizing the precision (or specificity) of models to reduce the number of potential contacts returned by the method while maintaining recall as high as possible. The approaches described below use increasingly sophisticated methods that rely on more information.

Location Prediction Approach 1: Wi-Fi Coverage (P1)

This approach relies on a Wi-Fi coverage map converted to a list of locations as described earlier. From the Wi-Fi log data, it is impossible to determine in which of the locations covered by an AP the person is. Thus, all locations covered by an AP are considered. When no additional information is available, one can consider the equal probability of a person being in any of the covered locations:

$$\mathrm{p}(\mathrm{loc_i}|\ \mathrm{A\ P}) = \frac{1}{\mathrm{N_{AP}}}.$$

where N_{AP} is the number of locations covered by the access point AP. When occupancy information is available, the probability of a person being in a location is proportional to the occupancy of a given room, with the assumption that there are more people in larger rooms. Here, occ_j is an occupancy of location *j*.

$$\mathrm{p}(loc_i|\ A\ P) = \frac{occ_i}{\sum_{j=1}^{N_{AP}} occ_j}$$

If no actual occupancy data are available, they can be estimated from area of the room as defined by the International Building Code [15].

Location Prediction Approach 2: Wi-Fi Coverage + Previous Location (P2)

The probability of being in a given location depends on the previous location, and more specifically, on the coverage of the previous access point to which a person was connected. This is illustrated in Figure 6.

Suppose a person is at location A and is connected to AP 1108X. If the person moves to location B, it is unlikely that AP connection will change because the new location is still in range. However, if the person moves from A to C, a switch to a new AP (1110X) will likely occur. Consequently, if a person is observed being collected to AP 1108X and then 1110X, it is unlikely that the person is at location B (covered by both APs), and is more likely at location C.

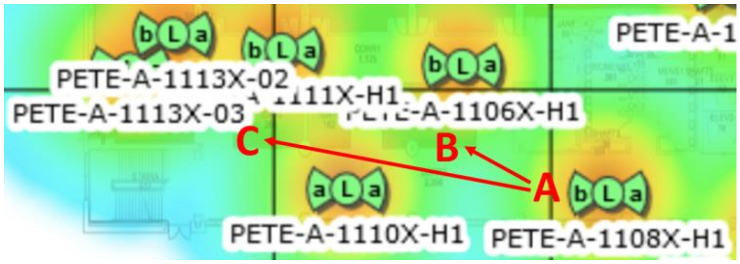


Figure 6: Illustration of switch between access points.

The probability of being at a given location given information about the current, AP_j, and previous AP_k access point in the Wi-Fi log is calculated as follows:

$$p(loc_i | AP_j, AP_k) = \begin{cases} \frac{occ_i}{\sum_{j=1}^{N_{AP_j \backslash AP_k}} occ_j} \; when\; loc_i \in cov_{AP_j} \backslash cov_{AP_k} \\ \frac{1}{\epsilon} \; when\; loc_i \in cov_{AP_j} \cap cov_{AP_k} \end{cases}$$

Here, cov_{AP} is the coverage of access point AP, $N_{AP_j \backslash AP_k}$ is the number of locations covered by $AP_j$and not by AP_k. The parameter ϵ is used to allow for small probability that the switch of AP occurred even if within the range of the previous one. In the presented work ϵ was set to 100, but that value needs to be optimized. The if occupancies are not known, occ_i can be set to one for all locations.

While the above approach is reasonable, it has limitations. One can construct a movement pattern in which a person moves from location A to C and then to B in Figure 6. The last move B to C does not trigger AP switch. Because of the first move A to B with Wi-Fi switch, the location B has low (or zero) probability which is incorrect. This type of situation can be addresses by limiting the use of the formula above to certain period of time, after which returning to standard probability calculation. Further work is needed to parametrize this method based on real movement data.

Location Prediction Approach 3: Wi-Fi Coverage + Movement (P3)

This approach takes into consideration how individuals may move between locations in a building. It is based on the building graph G, previously described in Section 3.1 and exemplified in Figure 1. Consider how a person may move within the building. Let, *sloc* be a previous known location of the person, and *dloc*, be a new known location. There are typically many such paths, known in the graph theory as simple paths (paths in graph with no repeated locations). The building graph G, allows to construct possible paths *paths(G, sloc, dloc)* between *sloc* and *dloc*. Let $sp(sloc, dloc)$ be the shortest path between *sloc* and *dloc* and $|sp(sloc, dloc)|$ be its length. When considering distant locations in building, there are potentially thousands of combinations of possible routes between points with the number growing exponentially with the length of the path. In the presented work length of paths is limited to $(1+\gamma) \cdot |sp(sloc, dloc)|$ where γ is a parameter. For example, the shortest path between rooms 3100 and 4800 is 3100 -> CORR12-3 -> CORR11-3 -> STAIRB-3 -> STAIRB-4 -> CORR4-4 -> CORR2-4 -> 4800. When considering all paths no more than 50% longer than the shortest path, there are 68 possible paths through the building exemplified below. When considered all possible simple paths in the building, there are thousands of possible ways to get from 3100 to 4800.

```
['3100', 'CORR12-3', 'CORR8-3', '3500A', 'PANTRY1-3', 'CORR3-3', 'CORR7-3', 'STAIRA-3', 'STAIRA-4', 'CORR2-4', '4800']
['3100', 'CORR12-3', 'CORR11-3', 'CORR3-3', 'CORR7-3', 'STAIRA-3', 'STAIRA-4', 'CORR2-4', '4800']
['3100', 'CORR12-3', 'CORR11-3', 'STAIRB-3', 'STAIRB-2', 'CORR3-2', 'ELEV1-2', 'ELEV1-4', 'CORR4-4', 'CORR2-4', '4800']
...
['3100', 'CORR12-3', 'CORR11-3', 'ELEV2-3', 'ELEV2-5', 'ELEV2-4', 'CORR4-4', 'CORR5-4', 'CORR2-4', '4800']
['3100', 'CORR12-3', 'CORR11-3', 'ELEV2-3', 'ELEV2-5', 'CORR6-5', 'ELEV1-5', 'ELEV1-4', 'CORR4-4', 'CORR2-4', '4800']
['3100', 'CORR12-3', 'CORR11-3', 'ELEV2-3', 'ELEV2-5', 'CORR6-5', 'STAIRB-5', 'STAIRB-4', 'CORR4-4', 'CORR2-4', '4800']
['3100', 'CORR12-3', 'CORR11-3', '3300', 'CORR3-3', 'CORR7-3', 'STAIRA-3', 'STAIRA-4', 'CORR2-4', '4800']
['3100', 'CORR12-3', 'CORR11-3', 'CORR9-3', 'CORR3-3', 'CORR7-3', 'STAIRA-3', 'STAIRA-4', 'CORR2-4', '4800']
```

Assuming that all paths are equally probable for a person to take, one can count how many times a specific location is at those paths, thus, calculating a probability of passing through a specific location. Note that the assumption of equal probability of paths, can be further improved with the actual movement data. This results in the creation of a lists of locations with associated probabilities as exemplified in Table 3.

Table 3: Example locations on paths between room 3100 and room 4800.

location	Count	Probability	location	Count	Probability	location	Count	Probability
3100	68	2.061	**CORR11-3**	67	2.030	**ELEV1-3**	30	0.909
CORR12-3	68	2.061	**STAIRB-3**	4	0.121	**ELEV1-1**	18	0.545
CORR8-3	1	0.030	**STAIRB-2**	2	0.061	**CORR6-1**	2	0.061
3500A	1	0.030	**CORR3-2**	4	0.121	**ELEV2-1**	18	0.545
PANTRY1-3	1	0.030	**ELEV1-2**	19	0.576	**ELEV1-5**	19	0.576
CORR3-3	4	0.121	**ELEV1-4**	30	0.909	**CORR6-5**	4	0.121
CORR7-3	4	0.121	**CORR4-4**	64	1.939	**ELEV2-5**	19	0.576
STAIRA-3	4	0.121	**ELEV2-2**	19	0.576	**STAIRB-5**	2	0.061
STAIRA-4	4	0.121	**ELEV2-4**	30	0.909	**ELEV2-3**	30	0.909
CORR2-4	68	2.061	**STAIRB-4**	4	0.121	**3300**	1	0.030
4800	68	2.061	**CORR5-4**	21	0.636	**CORR9-3**	1	0.030

When using Wi-Fi access data, person's location typically cannot be uniquely identified. Therefore, one needs to consider a set of previous likely locations $S = \{(sloc_i, p(sloc_i)), i = 1,2,..\}$ where $p(sloc_i)$ are probabilities associated with locations $sloc_i$ (see Approach 2), with $\sum p(sloc_i) \leq 1$. Similarly, $D = \{(dloc_j, p(dloc_j)), j = 1,2,..\}$. Then, $paths(G, S, D)$ is a set of all possible paths between all combinations of points from *S* and *D*. Assuming the independence of sources and destinations for each path, the probability of that path is multiplied by $p(sloc_i) * p(dloc_j)$. When counting locations of all paths, a list of all potentially visited locations is created, as exemplified below. The list constitutes potentially visited locations with associated probabilities, and is added to the overall list of predicted locations.

The same approach is taken when entering and exiting the building. When a person is observed connecting to an AP in the building, there is a limited number of ways in which the person got there. The algorithm considers all possible routes from outside the building to the detected locations. After the last AP connection is observed, all possible routes from that location to the outside are calculated.

AP Connection Time and Predicted Location Time

To model temporal relationship between recorded AP and predicted location the following observations are made: (1) a Wi-Fi device cannot connect to an AP before being in range; (2) the device can connect to AP at any time while in range; (3) the device can be reported as connected after it is no longer in range (lag in disconnect/switch to new AP); and (3) the device can disconnect from AP anytime while still being in range. These are illustrated in Figure 7 that shows two cases. In the case 1 Wi-Fi connection is established after time δ from the moment of arriving within range of an AP. The device is disconnected in the interval ε_1 before it is moved out of range. In the case 2 Wi-Fi connection is established after time δ from the moment of arriving within range of an AP. The device is disconnected in the interval ε_2 after it is moved out of range.

The same principle applies when switching between two APs. The device can be already in range of the second access point while still connected to the first one. When calculating time at location, the presented method assumes the longest possible connection time.

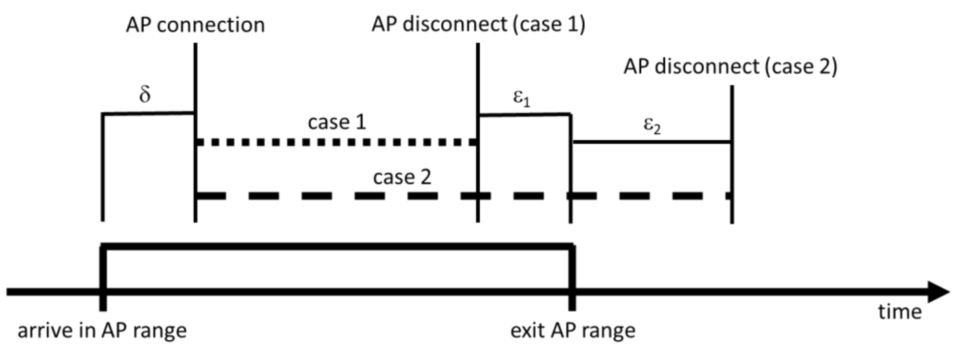


Figure 7: Relationship between location and connection time.

Location Prediction Approach 4: Wi-Fi Coverage + Patterns (P4)

Patterns of previous movements are used to help estimate location. More specifically, the additional data is used to better estimate probabilities of locations used in approaches P1-P3, thus improving overall method accuracy. Past frequently visited locations and movement patterns can be modeled using an approach previously used to analyze GPS data for people with Alzheimer's Disease [24] and is being applied is a study on social distancing during COVID-19 pandemics [25]. Furthermore, the prediction can be improved by linking the Wi-Fi data with additional digital data sources such as keycard access logs, a directory with office locations, student class registration lists, and additional sensor data, if available. The method is outside of scope of the currently presented work and is only discussed for completeness.

3.6. Contact Score Calculation

Contact scores provide direct information to public health personnel conducting contact elicitation during contact tracing process. Contact scores are assigned to individuals that were potentially in contact with an infected person. The higher the score, the more at risk an individual is. Intuitively, one can think of the contact score as an expected time of contact, but the actual formula is more complex. Let $\mathrm{p}_I(loc_i)$ be probability of probability of the infected individual I being at a location loc_i at a given time. Similarly, let $\mathrm{p}_C(loc_i)$ be probability of probability of potential contact individual C being at the location loc_i at the same time. Assuming that $\mathrm{p}_I(loc_i)$ and $\mathrm{p}_C(loc_i)$ are independent, i.e., there is no relation between movements of the two individuals, the probability of both being at the location loc_i is $\mathrm{p}_I(loc_i) \cdot \mathrm{p}_C(loc_i)$. Let $\mathrm{t}(loc_i)$ be predicted time interval in which both individuals were at loc_i. The contact score is defined as:

$$CS(I, C) = \sum_i \omega_i \cdot \mathrm{p}_I(loc_i) \cdot \mathrm{p}_C(loc_i) \cdot \mathrm{t}(loc_i)$$

where ω_i is a parameter characterizing type of location in relation to the risk of the disease transmission. Locations such as elevators or small offices have high values of ω, while large areas such as lecture halls have low ω. For simplicity, in the presented study, all values of ω were set to one. Based on the disease spread characteristics one can estimate their values for different locations.

3.7. Accuracy Measures

The presented work evaluates the accuracy of the presented methods in terms of widely used metrics adapted for the location prediction and contact prediction problems.

Location prediction metrics

Ideally, one would like to construct a model that precisely and correctly predicts the location of an individual. In the presented work, location-recall (L-R) specifies the proportion of locations at which a person was really present and that were correctly identified. Location confidence (L-C) is defined as the proportion of correctly identified locations to all possibilities returned by the model. The terms location-recall and location-precision are specifically used to avoid confusion with similar standard metrics used in binary classification as described in the later section. Let TP be number of true positives that is a set of correctly predicted locations for a person to be at; FP be number of false positives that are locations at which the person was not present, but that were identified by the model; TN be number of true negatives that is a set of locations correctly predicted for the person not to be at; and FN be number of false negatives that is a set of locations in which the person was present, but that were missed by the model. $|predictions|$ indicates the number of predicted locations in which a person can be (when methods are unable to pinpoint one specific location).

$$recall = \frac{TP}{TP + FN}$$

$$confidence = \frac{TP}{|predictions|}$$

$$confidence\ gain = \frac{confidence}{no\ informtion}$$

The last metric, confidence gain (PG), shows an improvement of confidence over a situation in which a person can be in any location in a building, which is $\frac{1}{|locations|}$ with $|locations|$being the total number of rooms in the building. In the presented work there are 451 locations in the considered building. For example, *PG = 10* indicates that the prediction is 10 times better than saying that a person is somewhere in the building.

Contact prediction metrics

Contact prediction can be defined as a binary classification problem. Let *P* be a population of individuals that access building within a specified period, such as one day. Let *v* be a person who is infected and possibly contagious. The problem is to identify all people *s* in *P* such that *s* and *v* were in contact. Let TP be true positive, that is individuals correctly identified to have contact; TN be true negative, that is individuals correctly identified not to be in contact; FP be false positive, that is individuals incorrectly identified as having contact, and FN be false negative, that is individuals incorrectly identified not to have contact. The results are reported in terms of accuracy, recall, precision, and F1-score.

$$recall = \frac{TP}{TP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$accuracy = \frac{TP = TN}{TP + TN + FP + FN}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

3.8. Simulated Contact Data Generation

The presented work has been conducted at the peak of COVID-19 pandemics, thus not allowing the research team to ask study participants to be in contact when performing scenarios. Instead, the approach taken was to allow study participants to complete scenarios at any time, and then align the scenarios in time to simulate contacts.

To do so, two approaches to generating synthetic data have been used. In both cases, the data were generated by shifting in time the real data of scenarios completed by participants. Location-based synthetic data are generated by aligning completed scenarios in time so that a pre-defined number of participants “meet” in specific locations. The one-day synthetic data simulates a pre-defined number of individuals accessing a building within an 8-hour period.

Location-based synthetic data

The scenarios were developed so that participants visit some locations more frequently than others. While such data is not intended to simulate real usage of locations, it allows for simulating multiple individuals at the same room at the same time.

Four locations were selected for further analysis: Suite 3000, Suite 3600 and Classroom 3904 on the 3rd floor of the building, and laboratory space 4800 on the fourth floor. Suites are represented by their lobbies through which people need to pass to enter rooms within. The laboratory is an open space with five workspaces and one collaboration area.

One-day synthetic data

For a more detailed study of the performance of the methods, synthetic data for one day (8 working hours) was created. The data were created by sampling with replacement and randomly positioning in time participant-scenarios. Each randomly selected scenario has been assigned a random start time within the 8-hour period. If the scenario overlapped in time with another copy of itself, it was discarded, and another scenario was selected. This process was repeated until the desired number of synthetic individuals was created.

In the presented work, 100, 200 and 500 synthetic individuals were considered.

3.9. Implementation

The presented Wi-Fi data modeling has been implemented using Python programming language and utilizes Pandas library for general data processing, Scikit-Learn for machine learning and model evaluation, NetworkX for graph modeling and analysis.

All source code has been written for the purpose of the presented study. The source code will be made available on the project website.

4. Results

4.1. Location prediction

The location prediction methods N1, N2, P1, P2 and P3 have been applied to scenario data collected from 12 participants. Data for scenarios with no corresponding Wi-Fi logs were removed. Not surprisingly, methods that predict locations have better performance than naïve approaches. It is also clear that the performance increases when locations are limited to those in which individuals stayed longer (i.e., 5 minutes, 10 minutes). Note that the sample size for locations in which individuals stayed for more than 10 minutes is small. The methods have relatively good recall and gain over no room information more than 100-fold. By far the best results are obtained from method P3 that combines the movement of people through the building with P1 approach to predicting when people are connected to a Wi-Fi AP. When only movement is modeled, method P3-Move also shows accurate results.

Table 4: Summary of experimental results comparing approaches to location prediction on scenario data. Results are reported in terms of recall, confidence and confidence gain.

	All data, N = 1271			5+ minutes, N = 100			10+ minutes, N = 6		
Method	**Recall**	**Conf.**	**CG**	**Recall**	**Conf.**	**CG**	**Recall**	**Conf.**	**CG**
N1	3.462%	100%	529.00	11.0%	100%	529.00	16.667%	100.0%	529.00
N2	89.772%	0.189%	1	99.000%	0.189%	1	100.0%	0.189%	1
P1 RSSI ≥ -50	10.535%	40.988%	216.83	20.202%	35.582%	188.23	16.667%	44.048%	233.01
P1 RSSI ≥ -70	28.403%	8.861%	46.87	48.0%	7.389%	39.09	50.0%	10.053%	53.18
P1 RSSI ≥ -85	40.834%	2.894%	12.47	59.0%	2.617%	1.38	50.0%	3.337%	17.65
P3	89.773%	24.022%	127.08	79.167%	17.157%	90.76	85.714%	17.646%	93.34
P3-Move	65.379%	10.756%	56.90	64.286%	12.515%	66.20	66.667%	8.047%	42.57

4.2. Simulated Contact Prediction

Datasets constructed using the methods described in section 3.7 were used to test the accuracy of contact prediction. All simulated experiments have been repeated 5 times, and average results are reported.

For location-based simulation, the objective is the assessment of a close contact in a desired location. A randomly select simulated participant is selected as the target participant. Using the simulated data, anyone who has an overlap with the selected participant at the desired location is defined as positive, and anyone who doesn't have an overlap with the selected participant at the desired location is defined as negative. The result of location-based simulated data is shown in Table 5.

Table 5: Recall of models predicting known contacts simulated in specific locations.

	Suite 3000	**Suite 3600**	**Classroom 3904**	**Lab 4800**
Real Contacts	30	40	30	26
N1	0.47	0.43	0.53	0.69
N2	1	1	1	0.92
P1 50	0.47	0.43	0.59	0.69
P1 70	0.87	0.75	0.83	0.92
P1 85	1	0.96	0.89	0.92
P2	1	0.96	0.89	0.92
P3	1	0.96	0.96	0.92
P3-Move	0.2	0.21	0.67	0.32

When simulating one day of data in a building, the number of people entering the building was changed between 100 and 500. These numbers are consistent with the real numbers of people who accessed the building daily in October 2020: about 400 ± 30 during weekdays and about 150 ± 50 on weekends. To calculate average performance of the method, the simulation assumes that one person is infected, and all contacts are predicted to calculate accuracy metrics. This is then repeated for all people in the data.

Table 6: Results of contact prediction on simulated "one day" data with varying number of individuals. Accuracy, recall and precision are calculated at classification threshold 0.5.

N	100				200				500			
Real Contacts	36				61				113			
	AUC	Acc	Rec	Prec	AUC	Acc	Rec	Prec	AUC	Acc	Rec	Prec
N1	0.53	0.06	0.06	0.06	0.54	0.08	0.08	0.08	0.54	0.09	0.09	0.09
N2	0.91	0.70	0.83	0.70	0.90	0.64	0.81	0.64	0.91	0.62	0.82	0.62
P1 50	0.61	0.14	0.14	0.14	0.61	0.22	0.22	0.22	0.625	0.25	0.25	0.25
P1 70	0.83	0.59	0.61	0.59	0.80	0.48	0.50	0.48	0.78	0.51	0.55	0.51
P1 85	0.83	0.60	0.67	0.60	0.79	0.54	0.58	0.54	0.78	0.50	0.57	0.50
P2	0.83	0.60	0.67	0.60	0.79	0.54	0.58	0.54	0.78	0.50	0.57	0.50
P3	0.89	0.70	0.79	0.70	0.82	0.57	0.64	0.57	0.83	0.56	0.66	0.56
P3-Mov	0.56	0.11	0.11	0.11	0.53	0.06	0.06	0.06	0.54	0.09	0.09	0.09

Results for both location-based and one-day simulated datasets confirm the superiority of methods that predict locations, instead of naively using Wi-Fi connection data. Method P3 which applies graph routing algorithms to model likely paths of movements of individuals gave the best results.

5. Conclusion

The presented work confirmed that Wi-Fi data could be used to support contact tracing. The investigated approaches to location prediction and contact score calculation can be used to complement data collected during public health case interviews, as part of the contact tracing during a communicable disease outbreak or in the time of a pandemic such as the one the world is currently experiencing with COVID-19.

The study has several limitations that need to be investigated in the next phase of the project. The designed scenarios were created to study the relationship between movement and Wi-Fi data. The scenarios may not represent how real people move. For example, there are only a few longer "stops" in the data where participants were located for a longer period of time. This is in contrast to people's movements in real life, especially in the environments such as university campuses in which many people remain for 2-3 hours or longer. This will be addressed by the collection of real movement data for a large number of participants moving freely around the campus. Furthermore, the data will be collected in multiple buildings to test the generalizability of the approach beyond the one building used in the presented study. More specifically, identifiable data for consented individuals will be collected and cross-checked with their reported real locations. In addition, deidentified data of all people accessing the Wi-Fi network will be assessed to find population-level patterns of movement and tune method parameters.

The developed models are controlled by a large number of hyperparameters. The values used in the developed models are based on the investigators' assumptions but may not be optimal. The work will be further investigated by modeling these hyperparameters with data collected from people freely moving on campus by applying machine learning algorithms to find optimal settings.

The presented methods are envisioned as built into a decision support system in which a person conducting a case interview with an infected individual can have access to data about predicted locations, can interactively confirm or disconfirms known locations, and add more places in which the infected person has been. The system will then produce a ranked list of potential contacts, each with an associated score, predicted location and predicted time of contact. Such information can then be used to conduct a more informed and thorough follow up with individuals as part of contact tracing protocol.

Tracking people raises privacy concerns. This is not different in the presented approach to contact prediction. While the Wi-Fi data are routinely collected for network maintenance, their use for contact prediction and consequently contact tracing may be questioned. Presenting arguments for and against using Wi-Fi data are, however, outside scope of the presented work that focuses on technical aspects of contact prediction.

Finally, the algorithms presented in this study are implemented for transparency of source code and ease of future modification as part of the study. In later phases of this project, selected algorithms will be reimplemented specifically to improve their computational efficiency.

Acknowledgements

The authors thank Josh Shilling, Rachele Peterson and Alison Yun for their help with managing the project and Dr. Sanja Avramovic for her review and comments that helped improve the work. The authors also thank Rashid Hashem at Mason ITS for his help in extracting Wi-Fi data. The project was funded by the National Cancer Institute, Praduman Jain (PI), subcontract to George Mason University.

References

[1] "Virus tracing apps 'ready in weeks' in EU and Australia". BBC News. 2020-04-17. Retrieved 2020-04-22.

[2] Alex Hern. "Digital contact tracing will fail unless privacy is respected, experts warn". The Guardian. ISSN 0261-3077. Retrieved 2020-04-20

[3] Anglemyer, A., Moore, T. H., Parker, L., Chambers, T., Grady, A., Chiu, K., ... & Bero, L. (2020). Digital contact tracing technologies in epidemics: a rapid review. Cochrane Database of Systematic Reviews, (8).

[4] Bahri, Shamshul (2007-01-01). "Enhancing quality of data through automated SARS contact tracing method using RFID technology". International Journal of Networking and Virtual Organisations. 4 (2): 145–162.

[5] Barbaschow, Asha. Australia looks to 'go harder' with use of COVID-19 contact tracing app. ZDNet.

[6] Braithwaite, I., Callender, T., Bullock, M., & Aldridge, R. W. (2020). Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. The Lancet Digital Health.

[7] Cho, H., Ippolito, D., & Yu, Y. W. (2020). Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv preprint arXiv:2003.11511.

[8] Choudhury, Saheli Roy (2020-03-25). "Singapore says it will make its contact tracing tech freely available to developers". CNBC. Retrieved 2020-04-22.

[9] Das gefährliche Chaos um die Corona-App. http://www.tagesspiegel.de (in German). Retrieved 2020-04-20

[10] Douglas, J., Farrell, S., "Coronavirus Contact Tracing: Evaluating the Potential Of Using Bluetooth Received Signal Strength For Proximity Detection," ACM SIGCOMM Computer Communication Review, 2020.

[11]Eames, K. T., & Keeling, M. J. (2003). Contact tracing and disease control. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1533), 2565-2571.
[12]Farrahi, Katayoun; Emonet, Rémi; Cebrian, Manuel (2014-05-01). "Epidemic Contact Tracing via Communication Traces". PLOS ONE. 9 (5): e95133.
[13]Ferretti, Luca; Wymant, Chris; Kendall, Michelle; Zhao, Lele; Nurtay, Anel; Abeler-Dörner, Lucie; Parker, Michael; Bonsall, David; Fraser, Christophe (2020-03-31). "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing". Science
[14]Gupta, P., Mehrotra, S., Panwar, N., Sharma, S., Venkatasubramanian, N., Wang, G., "QUEST: Practical and Oblivious Mitigation Strategies for COVID-19 using WIFI Datasets,"
[15]https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/
[16]Huang, Yasheng; Sun, Meicen; Sui, Yuze (2020-04-15). "How Digital Contact Tracing Slowed Covid-19 in East Asia". Harvard Business Review. ISSN 0017-8012
[17]International Building Code, 2018.
[18]Keeling, M. J., Hollingsworth, T. D., & Read, J. M. (2020). Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J Epidemiol Community Health, 74(10), 861-866.
[19]ltuwaiyan, Thamer; Hadian, Mohammad; Liang, Xiaohui (May 2018). "EPIC: Efficient Privacy-Preserving Contact Tracing for Infection Detection". 2018 IEEE International Conference on Communications (ICC). Kansas City, MO: IEEE: 1–6.
[20]Mbunge, E. (2020). Integrating emerging technologies into COVID-19 contact tracing: Opportunities, challenges and pitfalls. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14(6), 1631-1636.
[21]Meneses, F., Moreira, A., "Large scale Movement analysis from WiFi based location data," 2012 International Conference on Indoor Positioning and Indoor Navigation, 2012.
[22]S. Schmitt, L. Zech, K. Wolter, T. Willemsen, H. Sternberg and M. Kyas, "Fast routing graph extraction from floor plans," 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Sapporo, 2017, pp. 1-8, doi: 10.1109/IPIN.2017.8115868.
[23]Trivedi, A., Zakaria, C., Balan, R., & Shenoy, P. (2020). WiFiTrace: Network-based Contact Tracing for Infectious DiseasesUsing Passive WiFi Sensing. arXiv preprint arXiv:2005.12045.
[24]Wojtusiak, J. and Mogharab Nia, R., "Location Prediction Using GPS Trackers: Can Machine Learning Help Locate the Missing People with Dementia?," Internet of Things, Elsevier, 2019 (in press).
[25]Wojtusiak, J., Bagchi, P., Durbha, S., Mobahi, H., Mogharab Nia, R. and Roess, A., "COVID-19 Symptom Monitoring and Social Distancing in a University Population," Journal of Health Informatics Research, in-press.
[26]http://www.netspotapp.com
[27]Zeitung, Süddeutsche. "Corona-App: Streit um Pepp-PT entbrannt". Süddeutsche.de (in German). Retrieved 2020-04-20.

A publication of the *Machine Learning and Inference Laboratory*
College of Health and Human Services
George Mason University
Fairfax, VA 22030-4444 U.S.A.
https://www.mli.gmu.edu

Editor: J. Wojtusiak

The *Machine Learning and Inference (MLI) Laboratory Reports* are an official publication of the Machine Learning and Inference Laboratory, which has been published continuously since 1971. Until 1987 while part of Ryszard Michalski's group at the University of Illinois, they were called ISG (Intelligent Systems Group) Reports, or were part of the Department of Computer Science Reports).