2021 6.033 Design Project:

A Better Exposure Tracing System

See also DP FAQ and DP Errata

Due Dates and Deliverables

There are five deliverables for this design project:

1) **DP Prep (DPP):** In order to help you prepare with your team design effort, this assignment will require some guided analysis of the DP specification below. This assignment will be written by each student individually and is due March 12, 2021 at 5pm, EST.
2) **DP Preliminary Report (DPPR)**: This preliminary report will lay out your key design decisions, including both a functional system design and a sketch of any data structures, storage management, and/or network protocols required to achieve your design. It will not include any significant evaluation. It will be approximately 2,500 words and is due March 30, 2021 at 5:00pm, EDT.
3) **DP Presentation**: This presentation will address the feedback received on the DPPR, and any corrections or updates to the design project specification. It will also outline evaluation criteria and use cases you will use later for evaluating your design. The presentation will occur during the week of April 21-27, 2021, EDT.
4) **DP Report (DPR)**: This will be your full report. It will include your final design, all diagrams appropriate for that, your evaluation of your design and a review of how effectively your design addresses the specified use cases. It will be approximately 6,000 words and is due May 11, 2021 at 11:59pm, EDT.
5) **Peer Review**: In Tutorial your team will have done an early "review," providing informal feedback to another team on their design. For this peer review, you will individually review a few specific sections of that (same) other team's final report and will address some specific questions about that report. It will be approximately 250 words and is due May 14, 2020 at 5:00pm, EDT.

Your assignment for each of the five parts above will be distributed in separate "assignment" documents.

The prep, preliminary report, final report, and peer review should be submitted via the submission site on the 6.033 website. As with real-life system designs, the 6.033 design project is under-specified, and it is your job to complete the specification in a sensible way given the stated requirements of the project. As with designs in practice, the specifications often need some adjustment as the design is fleshed out. Moreover, requirements will likely be added or modified as time goes on. We recommend that you start early so that you can evolve your design over time. A good design is likely to take more than just a few days to develop. A good design will avoid unnecessary complexity and be as modular as possible, enabling it to evolve with changing requirements.

Large systems are never built by a single person. Accordingly, you will be working in teams of three for this project. Part of the project is learning how to work productively on a long-term team effort. **All three people on a team must be in the same tutorial.**

Although this is a team project, some of the deliverables have individual components. See the individual assignment links for more information.

Late submission grading policy: If you submit any deliverable late, we will penalize you one letter grade per 48 hours, starting from the time of the deadline. For example, if you submit the report anywhere from 1 minute to 48 hours late and your report would have otherwise received a grade of A, you will receive a B; if you submitted it 48 hours and 5 minutes late, you will receive a C.

You must complete the three team design project components, parts 2, 3, and 4 above to pass 6.033. For the other two (individual) components of the design project, the contribution to your overall grade will be whatever grade you receive on that component. Thus, if you choose not to do one or the other of them, you will receive an F for that component only as a contribution to your overall grade.

A Better Exposure Tracing System

Version Date: March 1. 2021

I. Introduction

What has historically been called “contact tracing”, and now is often called “exposure tracing”, has been practiced in many forms since the time when doctors and scientists came to understand that human-to-human disease transmission occurred. In various forms this has taken place for hundreds of years and evolved over time. From early times under different conditions there were several related objectives. Key objectives include discovery and treatment of ill people, reduction of transmission, and vaccination where possible. These are a combination of societal and individual approaches and thus involve both the individual citizens and healthcare support. In addition, from the early days it was also recognized that there was always more to learn both during any such situation and for the future.

In addition to a deepening understanding of human-to-human infection, the evolution in technology has had a significant impact on the efficacy, scale, and breadth of exposure tracing, with significant tradeoffs in some of the objectives. When the telephone became widely available, the collection of contact information could be expanded enormously, although the tradeoff was that if the healthcare worker did not do this in person, then any additional in-person contact for treatment or vaccination would need to be handled separately. It allowed for a separation of the information collection from healthcare delivery. Much later, when cell phones became both widely available and geolocation of them become widespread, tracking people’s movements in order to provide population-based predictions of potential spread of diseases and the logistics of providing health care support to large communities became increasingly feasible. Most recently with the rise in availability of smartphones and development of Bluetooth Low Energy (BLE) beaconing, there have been yet further transformations in utilizing technology to further support societies during periods of serious infections. This project focuses specifically on app-based exposure tracing, enabled by smartphones supporting these recent developments in BLE . The objective is to support the needs of both the individual who may have been exposed as well as the healthcare system in providing the infrastructure and support needed by its citizens and whole societies in any such healthcare crisis. In addition, it would be ideal to also support the work and contributions of the scientists doing more long-range studies and thinking ahead to future situations.

In the description below, we will discuss three modules of such a system. This is an enhancement over what is being done today, where there are generally only two modules: smartphones and the central server(s). We will discuss the role that routers might play in such a system, as an enhancement to current approaches. As we will say again below, you are not required to use this newly introduced type of module, but it is there as an option for you. In addition, we will discuss those aspects of the system that you are being asked design, as well as some key components that exist and that you may assume are available to you as part of your design. One of the elements of this system that has only been developed in the last year is the role that Bluetooth Low Energy (BLE) on the smartphones can play in such a system, so we will discuss that in some depth.

The overarching challenge for you is to design a system to support a centralized organization in identifying infection exposure as well as support of those people who have been either exposed or infected. This will require an overall design of the organization and management of data collection, storage, and analysis. With that in mind you will have two key parts to your system. The first is to

develop the overall design of the system, so that you can design what is need on each type of module (phone, router, and the central server) and how they are coordinated to provide overall data collection, management, analysis, and support of the players. This part of your design focus on which data storage, analysis and possibly archiving will happen where and how that will be organized. The second aspect your design is the communication protocols between the modules. To the extent that data or computed results will need to be moved among modules, you will need protocols for that movement. Both timing and extent of protocol utilization will be important. Your design will be driven by a combination of your choices about where information will be stored, where it might be moved, what analysis is required for your design and where that analysis will occur. Both storage and communications may be limited by available capacity, leading to a possible design tradeoff.

There is a final aspect of this project you are being asked to consider: privacy. It is important to consider in your design that the information at the core of this system is about people's locations and contacts as well has, for at least those with positive test results, about their health. In that context, as things stand now, every infection test generates a record and every positive test (maybe all tests) are known to the government, at least here in the United States. Therefore you should be cognizant of this during your design process and it should inform your design choices.

We will begin with some background, both in terms of what exposure tracing is, at a very high level (there is some debate about this, so we take a particular position on that here) and a bit about the underlying BLE technology that is enabling the current approaches. In addition, in this section we will summarize three commonly-used methods for this type of exposure tracing. Then in Section III we will discuss the system you will be designing. In that section you will find specifics about the setting in which your system is to operate, a university, the particular functionality expected by the people and organizations involved, and some concrete numbers that will help you in understanding the constraints your system will be operating under. That will be followed by a set of use cases that will both guide your design and provide a basis for evaluation of the effectiveness of your design. Finally, the document closes with some thoughts on evaluation of your design. It should be noted that more specific evaluation will be discussed later in the term, and only be a significant part of your final report.

II. Background: current systems

For a system of the sort we are considering, we begin by reviewing the needs of the players or participants of interest here. They are the users (owners of the smartphones who have contact with each other and may be infected and/or infect each other) and the healthcare organization involved in supporting them and a whole community. Across these there are a number of key functions to consider and balance against each other. We begin with a particular assumption about transmission which we accept is not universal, but provides a starting point; this is about transmission. For this study, we are assuming airborne transmission, recognizing that it is not universal. We, therefore, make an underlying assumption that transmissions are most likely to occur at a distance between people of 3 meters or less, for 20 minutes or more. (For any such situation, it will be necessary to have a metric such as this, but will vary from one situation to another.) In determining the desired balance, some systems choose not to support some of these (note that there is more detail below on the terminology used here):

- Determine contact: to do this, the system must be able to figure out which people were within a close enough distance for long enough to be considered a contact event.
- Infection notification: There are two aspects of this. The first is to notify the person who has been found to be infected. The second is to notify the exposure tracing system of an infection. We highlight this here because in two of the examples below this notification to the system is

enforced by the central organization and service, while in the third it is voluntary by the infected person.

- Exposure determination: this aspect of the system determines whether a particular person was exposed during a contact event to someone who was infectious at that time.
- Exposure notification: notify those who were found to have had a contact event with an infectious person. Note that historically this has often been done with direct phone calls, which is both slow and does not scale.
- Provide information for wide-scale medical support deployment and other public health needs.
- Support the privacy of the individuals involved.

At its core, there are three types of "events" of interest in this system. First, as currently defined, a *contact* event occurs when two people are within 3 meters of each other for at least 20 minutes. Second, a *positive test* event occurs when the system learns that a person has been identified as having had a positive test. Before this can happen, a test will occur, a lab will have a positive finding and then report that to the public health authority and the individual. For the exposure tracing system to begin its analysis, this fact of a positive test must be recorded in the system itself. In two of the three examples below, this is handled by the public health system reporting it into the system. In the third, that reporting is voluntary and initiated only be the patient directly. The time of such reporting into the contact tracing system will initiate the contact tracing analysis and in conjunction with the time of the original test will determine a preceding infectious period. In this system, only the onset of positive testing is important, because from that one can extrapolate to an infectious period. It is assumed that once a person has tested positive they will quarantine until they have recovered and therefore cause no further exposure. Third, an *exposure* event occurs when a person who has had a positive test event has also had a contact event with another person during the infectious period. In other words, an exposure event occurs when someone who has not tested positive is determined to have been exposed to someone who has.

In addition to these events, it is worth giving some attention to the Bluetooth Low Energy (BLE) technology underlying many of the current contact tracing efforts. The smart phones broadcast extremely simple messages, primarily consisting of an ID for themselves. Each smart phone within reach of such a broadcasting smart phone records a timestamp, signal strength, and the received ID for each such broadcast. Because BLE signals do not reach very far and because the signal strength provides an estimate of distance away, these broadcasts provide the basis for estimating whether two smart phones are within a distance considered viable for a possible infection to have been exchanged. The currently widely used metric here is that exposure to infection requires proximity of 3 meters or less for 20 minutes or more. We will provide more numbers for all this below, but the general concept is important in understanding many of the current approaches to exposure tracing, using smartphones.

There are many apps that have been developed in this space with a wide variety of features and capabilities. To begin with, those that are not adopted by entire countries are generally not very effective, because without significant participation they do not have an adequate basis; at a minimum they are not widely enough used to provide a reasonable approximation to exposure events. So, we consider here three slightly fictionalized examples used at national scale in different sized countries, large, medium and small, none of which fully meets the objectives described above.

To set the stage, in each case there is a central organization. In these examples, they are countries and their public health services. In our study below the central organization will be the university, which manages the housing and food for the resident students, all healthcare capabilities including testing, vaccination, and medical support, and the support for isolation and quarantine. The central organization

is also responsible for the operation of the central (computing and storage) service. This service is centralizing logically, but at the scale of a country, it is likely to be physically distributed across a number of physical servers although in our research we have not found any reference to the organization of those "distributed" central services. In your design below, we are assuming that the university is small enough that only a single, central server will be provided. In all cases, because the central organization is providing the testing, positive test results can be "known" in the central service, although in the Small Country App example below, this is moderated.

A. Large Country App

The Large Country App (LCA) and central service comprise a system that is being operated and utilized by the central government. Using the app on their phone to communicate with the central service, each user registers with the government including their cell-phone number and a significant amount of personal information. The central service generates a unique id (we will call them the person's LCID) for each user, which is transmitted to the app running on the phone and used by the phone in its BLE proximity broadcasting. Each phone logs all such signals it receives within the bounds mentioned above, specifically including the LCID from the transmitting phone and a timestamp. When a positive test result becomes known to the public health system, it reports that to the LCA central service, which, in turn, utilizes the citizen's phone number to identify their phone and pull the contact logs, including the LCIDs and timestamps of all those whose signals were received by this phone, as well a fair amount of personal information about the person's health, profession, travel, etc. Included on the phone is GPS tracking, and additionally, the central service has access to phone company call records, both of which allow for geolocation tracking. It should be noted that this could be used for isolation compliance, but is not at present. The issue is whether or not people who have been exposed actually comply with the isolation expected of them. These logs provide the identities of the phones that had been in proximity over the "infectious" period. The central service analyses these logs for contact events and then exposure events, notifies the infected person and notifies all exposed people. The system also collects and makes use of geolocation information, primarily through the phone numbers and call record tracking. Thus, the central government has direct information and can provide support, both in terms of health and everyday life resources to both those found to be infected and those exposed. This system allows for human intervention to provide increased support of the individuals, which is part of the "gold standard" for effective compliance with isolation and quarantining, as well as adequate support. It also includes geolocation information about the phone both prior to infection or exposure events and Aafterwards. Finally, the central service in the LCA system keeps all records for 180 days, after which it deletes them. There is ongoing debate within the country about whether the public health service should or does have access to this data and can keep it indefinitely.

B. Medium Country App

The Medium Country App (MCA) is rather similar to the LCA, except that the handling of the identifiers in the BLE broadcasting is different. The user registers with the central service through the app. To reduce the probability that an individual can be traced easily (because they had one LCID as in the LCA situation), the phone requests batches of MCIDs from the central server. The central service generates batches for each client and keeps track of them all. The client then can select randomly among the set and will use the selected MCID for only about 15 minutes, then moving to another one randomly. The client can request a new batch whenever it would like. As with the LCA, the central service is notified when a positive test occurs and polls the phone of that person for all its contact logs. It then uses that information to determine contact and then exposure events. The app can now notify those exposed, but again the centralized system that includes phone numbers allows for human intervention for support. This system does not provide geolocation information, so tracking compliance is not supported at

present. Also here, the contact logs on the phones are kept for 2 weeks and the records are kept in the central service for 180 days.

C. Small Country App

The Small Country App (SCA) has several interesting differences from the other two. In this case, the central service has only a very limited role. The phone generates a key and a seed and uses a hash algorithm (see the box below for an introduction to hash functions) to generate a set of SCIDs. It generated 1440 (24x60) such IDs each day and selects randomly among them for an SCID used again for about 15 minutes. If a person is found to have tested positive, they can then voluntarily contact the central server and include the seed and key pairs for the period over which they were infectious. On a regular basis, each participant's phone can check with the central service for all the key and seed pairs reflecting a reported infection. Then the local phone generates the appropriate 1440 SCIDs for each reported infection seed and key pair and searches for matches that also reflect a contact event (3 m. for 20 min.) If matches are found then an exposure event occurred and the owner of the phone is notified, but the central service is not. The logs never leave the phone and are purged after 2 weeks. Neither the contact events nor the exposure events ever leave the phone. This approach provides significantly more privacy at the cost of any involvement by the central organization. It also has some potentially significant scaling problems; consider a server for a medium-sized metropolitan area. If infection rates are high, then the number of "infection" reports will be high and without any regionalization, the phone of each person in the Boston area would be evaluating every infection reported by a possible contact. For someone who rides public transit across the area this might be important, but for many people it would not, and the system does not track geolocation.

> ***What is a hash function?***
>
> *A hash function maps an input string x, which can be any length, to an output string y, which is a fixed length (let's say 256 bytes). We will often say that y is the "hash" of x. This mapping has a few properties:*
> *1. it's very hard to reverse; given y, it's virtually impossible to determine x*
> *2. It's virtually "collision-free"; given two different strings, x1 and x2, it's *extremely* unlikely that the hash of x1 will be equal to the hash of x2. In fact, for our purposes, you can assume that such a collision will never happen.*
>
> *The way this mapping is computed is out of scope for now, but suffice to say that hash functions are *very* carefully designed; only a few functions with these properties exist.*
>
> *Sometimes hashes are used in conjunction with random (or pseudorandom) numbers to generate multiple hashes of the same string, for example a sequence of hashes. In this case given a carefully selected number r, we can compute y1, the hash of x concatenated with r. We can then repeat that and take the hash of y1 and r to generate y2. And so on. Often x is called the seed and r is called the key. Remember, no collisions! — from the same base string x. As long as one can keep track of the sequence of random numbers, this can be a useful way to continually generate some sort of identifier from a string x.*
>
> *You may not find hash functions useful at all for your design project; you should feel no pressure to use them. But since some contact tracing apps use them today, we wanted you to have a quick primer. We'll talk about hash functions more in-depth in the last part of 6.033, because they're used for many things.*

D. Some considerations with respect to these apps

In this section we discuss a number of significant considerations with respect to these apps, although we note that this list is only a sampling of such issues.

One of the issues reported by essentially all providers of exposure tracing is that the potential users find it either untrustworthy or lacking in utility. The result of these is chronic lack of participation, making the system of limited effectiveness. From the users' perspective, the primary goal is correct and timely exposure notification. This means that the system needs to be both accurate enough and widely enough used to ensure valid exposure notification. In addition, users need to be informed. Direct calling contact tracing is believed to be the most effective at providing useful and supportive information both for infection and exposure. As mentioned, it is extremely difficult to make this scale, leading to delays and lost contacts. A third challenge with respect to the users is privacy. The users often do not want information about them exposed and utilized by third parties. Even more concerning in this domain is that of surveillance. This leads us to our final concern having to do with compliance with isolation, an area where we seek creativity. At most, violations of compliance with isolation should be reported to the central authority. Other less intrusive proposals that would positively effect compliance with isolation will also be welcome. Finally, it is important to the users that the app on their phone not significantly degrade operation of other apps on the phone. This can take the form of memory, storage, computation, communication, and generally battery drain.

Closely related to trustworthiness is the issue of privacy. Let us begin by considering privacy to be the ability of the individual to control who can know what information about themselves and that information can be used. One of the challenges one faces in trying to build a system for public health is that there is a potential tension. On one hand, there is the desire or intention to give the individual as much control over the privacy of the information about themselves as possible. On the other hand, there is the need for the society as a whole and those managing its infrastructure to have enough information be able to support, mitigate, and reduce societal illness. There are widely differing opinions about how to make the tradeoff in this space, and the three examples above provide a good demonstration of that. Adding the routers we have proposed here that are capable of logging the devices using them only makes the issue even more complex. This dilemma is inherent in every use of contact and exposure tracing, not just in our current situation.

The challenge with respect to privacy in this domain is whether and how to also support public health and epidemiology needs. If, for example, a country or large community would like to have some means of improving the probability that people in isolation actually stay in isolation, that is possible an approach like that of LCA, because the system is also includes call record tracking into the central system; neither MCA nor SCA has that capacity. A question one can ask in this domain is whether it is possible to improve isolation compliance while continuing to support some, perhaps more limited model of privacy. We will return to this idea below.

A second societal challenge is reflected in the lack of universal access to smartphones. There are many parts of the world where a combination of poverty and illiteracy lead to low ownership of smart phones. In a larger sense (outside the range of our 6.033 project), it is important to consider the overall population more extensively than we are able to do at present in 6.033.

Finally, these three examples raise some issues with respect to identifying exposure events. The first is turnover in IDs. The second is how accurate it is. We will consider them separately.

With respect to turnover, in LCA, there is no turnover in the ID used in its BLE broadcasts, but for the other two there is turnover. Each of them switches to another ID about once every 15 minutes, so if person A approached person B within 5 minutes of switching to a new ID, B would see 5 minutes of one ID and 10 minutes of another, with no way to link them. So that is the first part of the ID rollover problem. But more than that SCA rolls over its ID generator at midnight, so the set of IDs that might be used for a contact spanning midnight might be derived from different sources. Therefore, in considering how to evaluate whether A and B have been in contact with each other, the analysis must consider these problems.

In addition, with respect to accuracy and flexibility in the definition of "contact", LCA and MCA simply determine the signal strength of a received ID to determine whether the sender was less than or greater than 3 m. away and then determine whether the contact (less than 3 m.) was for 20 min. or more. SCA is more sophisticated. It postulates an equivalence of a closer distance for a shorter period of time to a contact event of 3 m. for 20 min. It also postulates a longer distance a longer period time should also be considered a contact event. These extensions to the definition are fixed and still based on the model of 3 m. for 20 min. equivalent, but there is growing evidence in the research that the underlying assumption may not be accurate. As an aside, with respect to the determination of "contact" that LCA and MCA determine a contact event based on data from the person with the positive test, whereas SCA does the evaluation on the non-tested person. But the computation could be done either way for either case. This is a choice you can make in your design.

III. Your improved contact tracing system

You will have noticed that each of the above examples broadly provides a partial solution to effective contact tracing, privacy of the individual, support for the public health system, and efficient and scalable utilization of resources. None is ideal. As outlined more below, you will be designing a system to operate in a university of similar size to MIT that improves on that combination of objectives. We realize that we are identifying a set of goals for this system, some of which may be contradictory, so part of your challenge will be to find a defensible compromise among them. First, we can identify in more detail the goals for the users followed by those the central organization and the underlying system to provide the intended functionality.

A. Goals

The goals for support of the users can be grouped as follows:

- Functionality: accuracy, timeliness
- Ease of use
- Low-impact on phone (storage, computation, communication)
- Privacy

For the users, probably the primary functionality is to notify them in a timely way *if* they have been exposed and in particular, to notify them *when* they were exposed, because both appearance of symptoms and the period when they may be infectious to others will be determined by when the exposure occurred. In addition, since they will be required to remain isolated, it will be important for them to be provided with adequate information and resources during their isolation period. Ease of use and timeliness are always important to users as well. On a different front, users are likely to have experienced or be aware of apps that drain the capacity of their phones. This may take the form of utilizing too much storage, requiring too much work thus making the phone run slowly, or draining the battery. In particular, using the Wifi networking of the phones in this project will be the bigger power

drain than the BLE signaling. Finally, as discussed above, users are concerned about privacy from two perspectives. The first is surveillance, or monitoring by the central organization. The second is what they consider inappropriate use of the information about them that identifies and targets them individually for purposes other than the functions directly related to exposure tracing.

The central organization running the exposure tracing service also has a set of goals. It is operating under the IEEE Code of Ethics, and takes the first point about behavior and conduct as primary:

> *to hold paramount the safety, health and welfare of the public, to comply with ethical design and sustainable development practice, to protect the privacy of other, and to disclose promptly factors that might endanger the public or the environment.*[1]

First, its goal is to support exposure notification to reduce societal infection. Closely related to this, as discussed above is to support and encourage isolation and quarantining of those individuals exposed or infected, while respecting the privacy of the users. As part of the goal of public health and safety, the central organization has a commitment to improving compliance with isolation, yet they understand that total surveillance is beyond where they will go. To aid in achieving this goal, the Wifi routers have the capability of logging which phone is using them. You are not required to use this capability, but if you do, you will need to include it in your considerations about privacy as well as utility in improving isolation compliance. A goal here is a compromise position in which as long as the user is compliant, perhaps the central organization is not involved and does not collect information. The challenge to the designers is to propose an approach which might put increasing pressure or exposure on people who violate their isolation increasingly flagrantly.

To achieve the goal of public health, the overall system has two record-keeping goals: somewhere in the system, either on the phones or in the central server, the tracing information needs to be kept for 2 weeks and the exposure information for 180 days. (The tracing information is kept on the phones in the examples above, but that need not be the case, in your design.) In addition, overall within the other goals, the system should move the burden of performance, power and resource utilization to the central service. Finally, there is another goal for which proposals are solicited: the provision of data adequately summarized or anonymized that could be used for researchers and long-range epidemiologists. This will be especially important in learning from each such situation to improve outcomes in future situations.

B. The underlying capabilities

We will explain the underlying capabilities to two parts. The first is the scheme including the types of devices and their communications capabilities. The second is the sizes, quantities and capacities.

1. The modules of the underlying system

The devices available to you are a central server, a set of Wifi routers (hotspots), and a much larger set of smartphones. The physical location (with size numbers discussed below) will be a university, with enough routers to support all dorm rooms, classrooms, hallways, and public spaces. The routers will be connected to the central server by a wired network.[2] Each person will have a smart phone which will support both Wifi and BLE. The smartphones will use a full TCP stack (discussed in Lectures 7 through 10 of the course) to communicate with the central server, running over WiFi to the routers, and the wired

[1] See https://www.ieee.org/about/corporate/governance/p7-8.html for the full text.

[2] As 6.033 discusses, any network will also have internal routers. For purposes of this project, we are not considering those, because they do not have an impact on functionality of the system you are designing.

network between the routers and the central server. They will communicate directly with each other for the purposes of determining exposure tracing only, using BLE. Notice that three different types of link layer protocols (BLE, Wifi, and wired networking) are used in different parts of the system. Lastly, we note that at the bottom of the TCP stack on the wired network there will be a Media Access Control layer, based on MAC addresses. We mention this because in general MAC addresses do not change (unlike addresses such as IP addresses). So, any logging done by the WiFi routers will be based on the MAC addresses of the smart phones.[3] As you saw above, none of the example approaches use this sort of logging. Figure 1 provides a simplified layout containing the types of elements and their possible interactions, including which communications technology is used where.

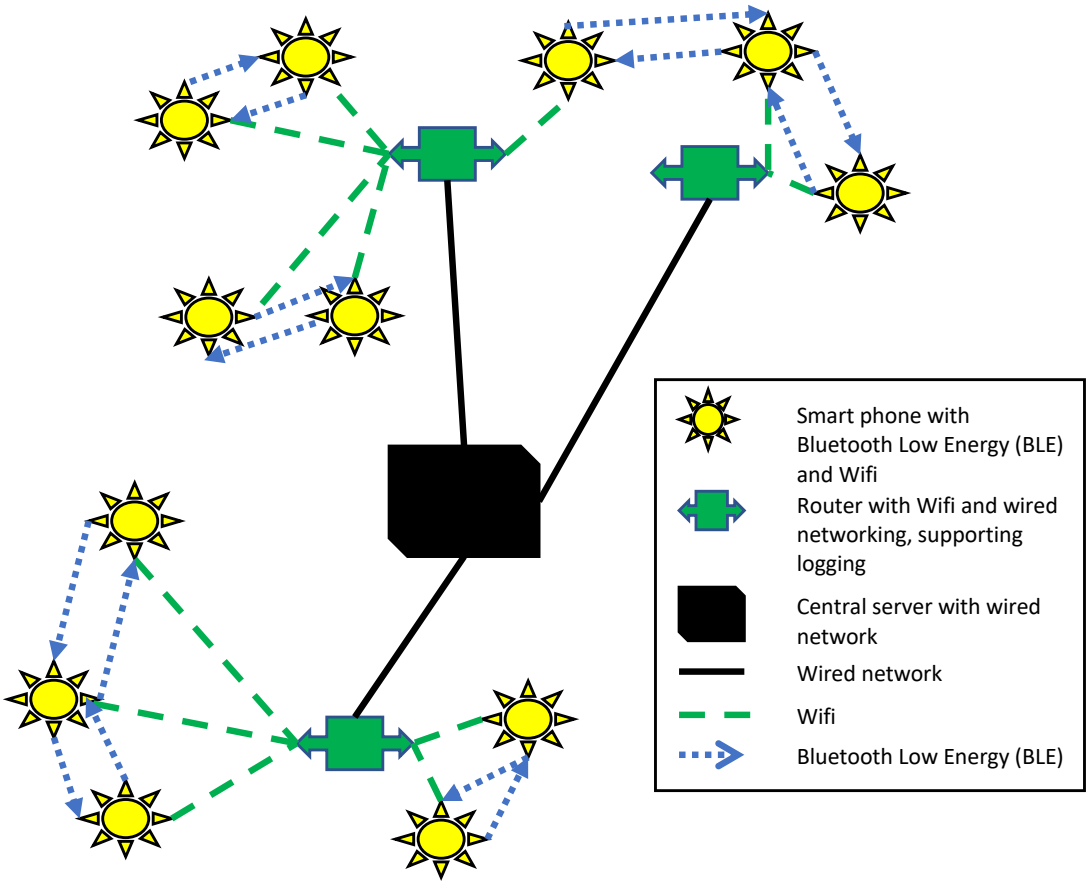

Figure 1: A simplified system layout: A central server is on the wired network. There is a collection of Wifi routers, both on the wired network and supporting Wifi devices. These are attached to walls, etc. and do not move. Then, there is a collection of smart phones, moving with their owners. Each smartphone supports both Bluetooth Low Energy (BLE), over which it broadcasts to nearby smart phones, and Wifi over which it can communicate with nearby routers.

2. Capacities and capabilities

Let's being with the capacities. In this slightly fictionalized situation, the central server is a 64 core machine, with 64GB of memory, 12TB of storage and 2 10Gb Ethernet ports. (The university believes that this should be significantly over the capacity you need, but you will need to verify that.) The routers are standard Wifi routers with a 2.5Gb Ethernet port. They are placed in all campus buildings such that

[3] In your design, if you choose to use this MAC address logging information, you will need to explain when and how it will be used, as well as where this logging information will be stored.

every person is always within range of at least one, and there are enough of them to handle any crowd capacity that would gather in any space. This includes all the living spaces as well. Since this is also slightly fictionalized, each one also contains 8GB of storage that can be used for logging. Finally, the smart phones, although smart may be rather low-end. Thus, your system must allow for 64GB devices, but since the users also want to use their phones for music and entertainment, your system will be limited to 1GB of storage on each phone.

In addition, as per the standards being developed now, each phone emits a BLE signal every 250msec. Although BLE signals can reach up to 50 or 100 meters (depending on the specific form) in freespace, the signals are significant reduced when going through barriers such as walls. For simplicity in this design, you should assume that no signals reach through walls, floors, ceilings, etc. Assume that these broadcasts are only received within a single open space such as a room.

Furthermore, we estimate that to include all the information needed to log one of these signals, each record will be 70 bytes long. This will include the ID, timestamp and signal strength, as well as some other ancillary information not needed in this particular analysis. If your design requires additional information to be stored on the smart phone, that must also be included in the 1GB limit on storage utilization. As with the three example apps above, the contact logging information must be stored for 2 weeks and at least the information about positive tests and exposures must be kept for 180 days. If your design includes saving more information to support longer-term research you must allow for that storage as well. It is your decision about where data will be stored.

C. The university

The context is a university. Here are some important facts. As you consider the sample use cases below, the relevance of these will become more apparent:

- When someone tests positive they will quarantine for at least 8 days or until they test negative, whichever is longer. Along the same lines, isolation as the result of an exposure event will also be for 8 days. Testing will be available to them. They can test at any time in the 8-day period, but if any test is positive they will need to quarantine. They must test on the seventh day and have a negative result to stop isolating on the eighth day.
- The size of the population is 20,000, of which 10,000 live off campus, 7,000 staff and 3,000 graduate students taking classes. (See below for more about classes.)
- The remaining 10,000 are undergraduates, living on campus. The university has some small living units (separate living groups, but not dorms) that house 30 students each. 3,000 students live in this type of housing. The remaining 7,000 students live in 20 dorms of 350 students each. These are partitioned into floors of 50 students each. The living group assignments are all known to the university.
- Each person living off-campus (staff and graduate students) is exposed not only the campus environment but also to the metropolitan area of the city. Because of this and the differences in control over the environments, the infection rates and daily case rates outside the campus are double that for the university. All of their testing will be done by the university under the same regime as those living on campus. These terms are discussed below in the section on use cases, Section IV.
- When one member of either one of the 100 living groups or one of the 140 dorm floors has tested positive, the whole group or floor will isolate for 8 days and be tested until there is confirmation of no infection.

- Of the staff, 30% are teaching one class or more each. Of the graduate students 40% are TAs in either undergraduate or graduate classes, separately from the courses they are taking. You can assume that the teaching load is spread relatively evenly across all classes.
- Twenty percent of the students (both undergraduate and graduate students) are taking at least one course that involves at least some in-person components and the remainder are all virtual. Each undergraduate is taking four courses and each graduate student two courses. There is no overlap among undergraduate and graduate student class. (That is, only graduate students take graduate classes and only undergraduates take undergraduate classes.) Which staff and students are in which in-person classes is known to the university. When someone in an in-person class tests positive, everyone else (students and teaching staff) in the class will go be isolated for 8 days and the class will continue only online. The normal testing routine described above will also be followed. For students living on campus, the university will provide separate isolation space for the full 8 days, in order to avoid use of shared spaces such as bathrooms within that student's living group.
- Test results are not all completed on the same day as the test is taken. As an average, 75% of the tests taken in a day will be analyzed and reported in that day. The remaining 25% will take up to 24 hours total to report back.
- There are three types of exposure events:
    - Measured exposure to an infectious person. This is based on exposure tracing.
    - Someone in the living group has tested positive, so everyone in that group is deemed exposed.
    - Someone in an in-person class tested positive, so everyone in that class who attended class during the infectious period is deemed exposed.
    - Note that an infection is assumed to have been possible for up to 24 hours prior to the test that came back positive, but tests can take up to 24 hours to be reported.

In addition to your design for managing the contact and exposure information, etc. the central server already provides:

- For each person at the university, the system will keep track of permanent information including their name, phone number, living arrangements (which dorm and floor or living group) and each class with which they are associated.
- For each living unit, a definition of its perimeter in terms of routers as well as whether it is in isolation or not.
- For each class that is at least partially in-person, everyone involved in it. In addition, the system keeps track of which in-person classes are completely virtual as a result of an exposure event, including start and end dates for the class. The end date is 8 days from the latest known infection in the class.

Further information is needed, but it is your decision as to where it is located and how it is managed:

- The "state" of each person: nothing, isolation, quarantine.
- If a person is moved to a different isolation location, that information, also including whether they transitioned from isolation to quarantine or directly quarantined as the result of a positive test.
- For each person in isolation the date at which it began, any testing during the isolation period, and final testing result at the end of the period.

IV. Use cases

To help you in your design, you must consider at least the following use cases. Our use cases at present are all relate to impact of infection rates and daily new cases. The infection mean is that each infected person infects that number of other people. The number of daily new cases is reported per 100,000 people. If you believe there are other use cases that provide an even more compelling argument for your design, you are encouraged to include them as well.

1. **Very low numbers**: Among the students living on campus, the infection rate is .83 and the number of daily new cases is 8.9. The reporting of all positive tests is spread evenly over time.
2. **Very high numbers**: Among the students living on campus, the infection rate is 2.5 and the number of daily new cases 80. Again, the reporting of all positive tests is spread evenly over time.
3. **Compressed very high numbers**: Among the students living on campus, the infection rate is 2.2 and the number of daily new cases is 80. For the set of test results available on any day, they will all be reported in a batch at 5pm that day.

In this context, it is important to remember that for those living off campus (staff and grad students) they are exposed to the larger community infection and daily case rates for the majority of their daily lives. In considering your design, it will be important to understand how well your design will supports these different scenarios while achieving the goals laid out earlier. This should include both all the aspects of storage, data analysis and communication, as well as response times. It will also be important to understand any tradeoffs you will be making with respect to privacy of the information about the individuals.

V. Thinking about evaluation

We will think about evaluation in the later stages of the design process, but it is helpful to start thinking about it early in the process as well. Here is a set of questions you might ask yourself:

1. How long should various kinds of data be archived where?
2. How much storage capacity will you need for that much data?
3. What are the factors in determining how much storage is needed?
4. How much data needs to be transferred under what conditions?
5. How quickly will the system respond? How quickly will a member of the community be alerted to an exposure event? Why would this matter?
6. Are there situations in which the system will be overloaded? In what ways and under what conditions?
7. There are things that may change with time. One example is that, with time, people may be able to be vaccinated, though even when vaccinated they may still be able to be carriers. A second is that the definition of “contact” may change with more understanding and science to amend the current definition. Does your design allow for changes in requirements?