Department of Informatics

Master Thesis

A meta-analysis on the effectiveness of digital contact tracing solutions to date

Bernhard Hjelen
Supervisor: Øyvind Ytrehus

November, 2021

Abstract

Digital contact tracing solutions were developed hastily in an attempt to combat the Covid-19 pandemic. These solutions are primarily based on proximity detection using either Bluetooth or GPS, and with an autonomous and anonymous exposure notification handling, trying to alleviate workload from manual contact tracers. The detection probability between two individuals partaking in digital contact tracing is the square of the fraction of the population that are actively using the app. Hence there is an underlying dependency on increasing the nationwide uptake for these solutions to give meaningful results. Problems arise where issues such as economical status, and minority inequalities prohibit users from using these solutions. We will discuss these issues, and look at some reasons as to why nationwide uptake is so important.

The digital contact tracing solutions are separated into centralized and decentralized solutions, these solutions have been under heavy debate regarding how privacy-preserving they are. Now that decentralized solutions have become the norm, we will look at why they are preferred over centralized, and look in-depth on how some of these solutions operate.

Given that a biological pandemic has multiple factors that are hard to properly address numerically, strategies such as generating simulation data are commonly used. In this paper we will represent an agent based model to generate new data. We will use this data, in unison with real life statistics, and comparatively with other simulated data, attempt to determine the efficacy and usefulness of the digital contact tracing solutions so far.

Acknowledgements

First and foremost, I would like to give my deepest appreciation to my supervisor Øyvind Ytrehus for his guidance during our meetings throughout the year, I would not have been able to do this without his help. I would also like to thank my friends and family for their support throughout this past year, especially Halvard Barstad.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

At the time of writing this, the COVID-19 pandemic has affected most of the global population in one form or another. With the rapid spread of the virus, nations quickly lost control and scrambled to find solutions. One of the countermeasures to the pandemic was, and still is, to create digital solutions that would detect a possible contagion between one individual and another. Given the lack of rules and procedures for developing such digital solutions, it is urgent to investigate the issues pertinent to the effectiveness of such digital solutions, as well as their impact on privacy. I believe that a framework for digital contact tracing should be established as a means of preparing for future pandemics.

In this paper we have multiple questions we will discuss:

- What % of the population would need to actively use a digital solution for it to have a reasonable affect?
- What amount of benefit is required to justify the privacy intrusion it surmises?
- How has the digital contact tracing situation evolved since first introduced?

To discuss these questions we will look at what digital solutions different nations used in the rapid response to the COVID-19, and also what solutions they use today over a year later. We will look at how these solutions work, and also take a look at the results they give. I will also create a program in an attempt to generate new data based on what we know now a year later.

1.2 Goal

The goal of this thesis is to:

- Give the reader a better understanding of how digital contact tracing works.
- Explain why we want digital solutions.

- Observe if digital tracing solutions have served its purpose so far, and if not, reasons to why.
- Take part of the discussion regarding good privacy-preserving digital contact tracing solutions.
- Look at the usage of mobile applications as a tool to increase the likelihood of detecting a contagion.

1.3 Overview

We have broken down the thesis into four parts. **Contact Tracing Background** comes first, with a brief overview over digital and manual contact tracing. Where we will also introduce some of the ideas regarding digital contact tracing, that we will build upon in the later parts of the thesis. **Protocols** serve as the technical background information chapter. Where we will go in depth on some of the common protocols that have been adopted to develop the solutions countries have used, or are currently using. **Model-Program** introduces aspects around network theory and social networking. And then we will go over a simple model I have made to generate data in order to determine the usefulness of digital contact tracing. **Discussion** is the last part of the thesis where we will bring everything we have observed together to discuss. We will try to answer the questions we have introduced by looking at data from other studies, real world statistics, and our own generated results. We will then summarize and conclude the thesis with some ideas for future work.

Chapter 2

Contact Tracing Background

In this chapter we will provide some background knowledge regarding contact tracing. We start by representing a glossary of epidemiological terms, and then look into the differences between manual and digital contact tracing. Finally we take a brief look into the history of digital contact tracing to lay the foundation for going into the next chapter where we will take a look at different protocols.

2.1 Epidemiological Background

These terms are defined from the official Norwegian health care NIPH2.1:

Term	Explanation
Index case	Person with proven covid-19 that triggers the contact tracing
Close contact	Person that could be exposed for infection after contact with the index case.
Contact tracer	Person that partakes in the manual contact tracing work.
Contact tracing	The process regarding finding, informing and eventually follow up or test close contacts.

Table 2.1: Terms from NIPH [1]

Additionally we represent a glossary table made by a research group with the purpose of bridging the knowledge gap between epidemiology and public health care. It contains terms which will be used in later sections of this thesis2.2:

Term	Explanation
Asymptomatic	A disease stage in which the infected individual does not and will not exhibit symptoms.
Basic reproduction number	The basic reproduction number (R0) is defined as the average number of secondary cases caused by a single infectious individual in a totally susceptible population
Control	Control relates to the strategies implemented to reduce the magnitude, spread, and progression of a disease in a population.
Disease	A term used in epidemiology and modelling to describe a physiological failure.
Eradication	Eradication refers to the elimination of a disease which can no longer reappear.
Exposed	The term 'exposed' is used when an individual has encountered a disease causative pathogen. This is necessary for infection or transmission to take place. However, it is not necessarily the case that infection or transmission occurs.
Illness	Illness is a subjective representation of a disease.
Immunity	Immunity refers to an individual's resistance to infection or re-infection by a causative pathogen.
Incidence	Incidence refers to the number of new cases of a disease over a period of time.
Incubation period	The incubation period represents the time period between the occurrence of infection (or transmission) and the onset of disease symptoms.
Infected	The term 'infected' refers to an individual who has contracted a disease causative agent and infection (or transmission) has occurred.
Infectious	Individuals who are infected and can transmit a pathogen (the cause of an infection) to other individuals.
Latent period	The latent period is defined as the period of time between the occurrence of infection and the onset of infectiousness (when the infected individual becomes infectious).
Pre-symptomatic	A disease stage in which the individual exhibits no symptoms, but is infectious and can transmit the disease.
Prevalence	Prevalence is defined as the number of cases of a disease at a specific time point.
Prevention	The term 'prevention' refers to the lack of disease occurrence despite exposure to, or transmission of a causative disease agent.

Table 2.2: Glossary made by a research group to bridge the gap between epidemics and public health [2]

2.2 Manual Contact Tracing

Contact tracing is the process of finding, informing and eventually backing up or testing the close contacts of the index case[3]. The process of contact tracing will vary from one disease to another, due to the difference in latent period, incubation time, infection method and such. For covid the Norwegian government has defined their manual contact tracing(mct) process like this:

- The lab confirms a positive test result of covid19.
- A doctor in the infected persons local area gets contacted.
- The doctor contacts the infected.
- The doctor considers the different contacts the infected has made in an attempt to identify other potentially infected.
- The doctor reaches out to the potentially infected to quarantine and potentially get them tested.
- The doctor registers all incidents of infection from the close contacts and stores this.

This strategy is more commonly known as the TITQ strategy, "Test, Isolate, Trace and Quarantine (TITQ)"[3]. Early identification of contacts is vital in order to break chains of infection and getting a successful strategy. In 2020, Norway reported 50 130 confirmed cases of which 31 155 were infected in Norway, and 4360 abroad[4]. In the last quarter of 2020, with greater testing capacity, Norway was able to test anyone with symptoms or suspected exposure. In these cases 30% had their country of infection unknown, while the cases in Norway had their source of transmission or exposure missing in 20%[4]. This indicates an issue with manual contact tracing. This same issue was also found in a study about the efficacy of manual contact tracing for coronavirus in the UK, using a similar definition of close contacts as in Norway, they state; "We would expect 10–15% of cases to generate at least one unidentified secondary case which would need detecting by other means."[5]

There is no doubt that the manual contact tracing model is an efficient tool in limiting the spread of diseases. However it is both labor and time consuming. Depending on factors such as the capacity of the contact tracing departments, the number of contacts per case and the cases knowing their contacts[6]. The number of cases that needs tracing is highly dependant on how you define a contact: if you define it too vague, then you have many untraced cases, but if you define it too strict, then you have a larger number of cases to be traced. Defining probability of exposure in the terms of length and duration of a contact is troublesome, and is something we will see return later in this paper when talking about digital solutions and model implementations.

Due to the manual contact tracing being a highly human based process, it is natural to have human fault hampering the efficiency of the process as a whole. These faults may be[4];

- Availability or willingness to pick up the phone when approached by the contact tracing teams
- lacking or incorrect information for the contacts
- gaps in memory or just a general unwillingness to collaborate

This could likely lead to delayed detection of a case, or the case going undetected. As a supplementary tool digital solutions was proposed, and have the advantage of rapid notification, and accurate contact information, in the case of detection. However digital solutions have their restrictions as well.

2.3 Digital Contact Tracing

2.3.1 What is Digital Contact Tracing

Digital Contact Tracing (DCT) is a description of the different tools made to aid with exposure notification. These solutions are usually software-applications made for phones or tablets, but alternative solutions with QR-codes have been used. There has also been a surge in other alternative solutions such as e-bracelets, or keycards which is more commonly known as "wearables". The common goal with these digital contact tracing tools is to correctly identify possible cases of exposure between its users. To identify a potential exposure it is vital to achieve some estimate of the distance between individuals. The lower the distance, the higher the risk of infection. Most of the contact tracing solutions rely on location data achieved by either Bluetooth, or some combination of GPS and Bluetooth[7]. We will look more into the specifics of Bluetooth later on. The required distance and duration used to define a potential exposure may vary from country to country depending on their policies and their implementations. This distance estimation neglects if the distance is measured indoors or outdoors, when it is known to be a huge factor in infection probability[8]. This problem could be its own thesis, and will not be of further focus in this paper. Here we showcase how different policies would affect the number of users traced2.1: [9]

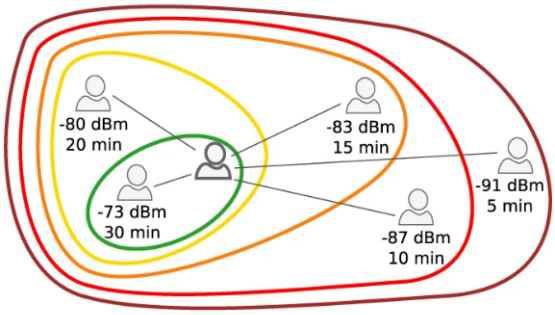
The DCT solutions efficacy is highly dependent on the number of users. There are many factors that results in the number of new users (adoption rate) of such solutions. Marketing, trust, accessibility, purpose and quality to name a few. Lacking cross-compatibility between the solutions is another issue that led to low efficacy. Given the need of a quick solution, each country made their own, which in turn led to competitiveness instead of cooperation at first. Cross compatibility has later been introduced.

2.3.2 History

The field of digital contact tracing is not particularly new. The concept goes back to 2007[10] but saw a slow development due to the low necessity. However, after the global outbreak of SARS CoV-2 (Covid-19) governments throughout the world saw a sudden emergence in need for this technology. Among the first solutions made for nationwide

	Signal strength threshold T_p (dBm)	Duration threshold T_d (min)	Fraction of CNS contacts
Policy 1	-73	30	2.2%
Policy 2	-80	20	7.3%
Policy 3	-83	15	13.4%
Policy 4	-87	10	25.9%
Policy 5	-91	5	56.7%

(a)



(b)

(**a**): The signal strength threshold T_p and the duration threshold T_d defining the policies are reported. Contacts with a duration larger than T_d and signal strength larger than T_p are considered at risk. The last column gives the fraction of the total number of interactions of the CNS data set that they correspond to. A larger value of the magnitude of the signal strength tends to correspond to a larger distance, such that in the second column the thresholds go from the least to the most restrictive policy. The policies are sketched in (**b**).

Figure 2.1: Showcase of how different digital contact tracing policies affect what users would be picked up

adoption was the application TraceTogether for Singapore, which was released the 20th of March 2020[11]. Given that the outbreak originated in China I would speculate that the Asian countries had less time to develop a solution, than the rest of the world. Aarogya Setu is the Indian app launched April 2nd[12], and the chinese launched their first solution (close-contact-detector) in February [13]. Another reason that the Asian countries were quicker than the rest of the world could be the cultural differences. It is easier in many of the Asian countries to maintain mandatory usage of the app, use centralized protocols and geolocate their residents in other ways than what would be acceptable in western countries[14]. As we will see more throughout this paper the ethics and practices around privacy policies is surrounded by this gray area. Together with competetiveness, and the lack of unified solutions, leads to inefficiency and inconvenience for its users.

As an example to this, in China, they integrated their solution "Health Code" with their national paying methods AliPay and WeChat[15]. In short, citizens receive a colored QR-code, "green", "yellow" and "red" indicating their risk of exposure. Stating that the app was voluntary to use, but at the same time demanding a "green status" to be able to access public points of interest, such as; public transit, schools, airports, grocery stores, restaurants and hotels, there is a obvious contradiction in what is said versus what is practically achievable for the population. And about the competetiveness, each city has their own "Health Code". So a green status in city A, might not give you access to public points of interests in city B, requiring you to get the app for city B as well.

Many of the aforementioned characteristics of the Asian approach were seen incompatible with the European legal and ethical view of individual privacy. Therefore, the day before Aarogya Setu was launched, April 1st, the Pan-European Privacy-Preserving

Contact Tracing (PEPP-PT) was announced [16], a non-profit organization that would deliver solutions that were both "centralized" and "decentralized". One of the major decentralized solutions that were run under the PEPP-PT umbrella was the Decentralized Privacy-Preserving Contact Tracing (DP3T), which is the protocol the majority of European DCT apps were using before many of them swapped over to GAEN. TraceTogether, the application used in Singapore mentioned earlier, open sourced their application to become the BlueTrace protocol. Utilizing a similar centralized reporting system as in PEPP-PT. There exists other viable protocols such as but not limited to; "PACT", "OpenCovidTrace", "ViraTrace" and "Whisper". But they have not been adopted by any nation, and will therefore not be looked further into in this paper. It is also worth mentioning that there are DCT solutions with custom or unknown protocols that have been adopted by nations, and we will not be covering every single existing solution.

The countries that did attempt to create their own custom protocols saw difficulties during development. Calculating distance between users was, and still is, majorly relying on Bluetooth signal strength. However, when using phones from different operative systems they might not get a signal at all. One of the custom protocols that was created is the first Norwegian contact tracing solution "smittestopp" released 16. April 2020. During the testing of detection between phone pairings Android-Android, Apple-Apple and Android-Apple, they saw a decline in the probability of detecting Apple phones. After troubleshooting this, they discovered that the phones running iOS would not announce or scan for signals when the screen was off, or if the application was in the background [17] [18] [19]. Also given that these issues were underlying in the operative system under Apple's control, and that Apple was in development of their own solution they did not want to cooperate with other custom solutions such as the Norwegian.

Chapter 3

Protocols

In this chapter we want to dive further into understanding the difference between centralized and decentralized protocols. These protocols act as a foundation for the development and implementation of the digital contact solutions. Different protocols might fit better for the ideals that the different health authorities have for their nations. There has been a wave of new protocols aimed to solve the issues revolving around what a good protocol needs to do. Some of these protocols have later seen nationwide adoption, others have been adopted for secondary solutions in countries/states. We will primarily look at the largest protocols that has seen nationwide adoption in multiple countries. The developers of the earlier protocols have stated issues around the development of the protocols, which is something we will cover at a larger extent later on in the discussion part of this paper. These development issues acts as the underlying reason as to why the Google—Apple Exposure Notification (GAEN) protocol/framework/api is so important for the current state of digital contact tracing globally. Issues regarding privacy and security will also be postponed untill later, as the scope of this chapter is to get an understanding of the fundamentals around these protocols.

3.1 Bluetooth

After soon 18 months of digital contact tracing development it is safe to say that the majority of implementations or solutions use bluetooth as their way of communicating between its users. The goal of this section is to get a broader understanding of how Bluetooth operates. Protocols differ in the ways they utilize this technology, but that will be covered in their respective sections.

Bluetooth devices has been around since the millennial change, and has seen different versions and settings since then. Covering all of bluetooth is not entirely our intention, and a bit out of scope so we will focus more on what is related to digital contact tracing. Bluetooth Low Energy, or BLE for short, is the bluetooth specification most contact tracing solutions use. Mainly due to the significantly reduced power consumption and cost, while still maintaining a similar range to other specifications e.g. "Bluetooth Basic" and "Bluetooth EDR(enhanced data rate)".

The BLE protocol specification utilizes beacons, which is a composition of data. The

composition is in this order: Preamble - Access Address - Payload - CRC. The Preamble and Access addresses are fixed values used to help with receiving beacons, assisting with timing estimation and synchronization. The Payload holds the majority of the data, firstly some structural information such as defining if the beacon is non-connectable or connectable, or has a public or random address. Then lastly the payload itself, which differs from protocol to protocol, contains the data. The data section can be empty, in which the beacon would be used to form a handshake of some sort. If it is not empty the data section often details information used in contact tracing matters. Which could be temporary identifiers, signal strengths, time information and so on. As mentioned these beacons may be non-connectable or connectable, and accordingly the devices doing the communication takes on roles. For non-connectable, or one way communication, the roles are *Broadcaster* and *Observer*. The Broadcaster will broadcast beacons, and the Observer will scan for them. For connectable, or two way communication, the roles are *Central* and *Peripheral*. In which the Peripheral broadcasts, and acts as the "slave". And the Central scans, and acts as the "master". The communication between the phones can generally be described as this.

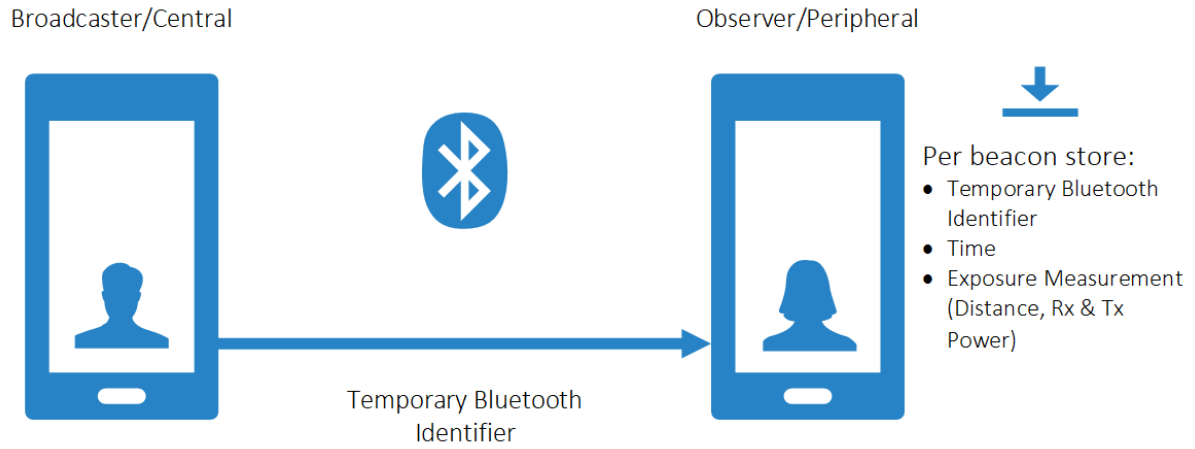


Figure 3.1: Bluetooth communication between two phones

The central is the only one that can do difficult operations, but as we will see in the protocols section later this can be circumvented to obtain mutual information between the devices.

To achieve a low energy cost protocol, devices taking on a broadcasting role will stay in sleep when not actively broadcasting. The time delay between each sent packet is referred to as the advertising interval. A smaller interval, or faster frequency of packets, gives a higher power usage but additionally reduces the time a device needs to wait for a broadcasted beacon. In a similar fashion the listener in the communication, or the scanner, would scan for packets in what is called a scan window. BLE operates on the 2.400–2.4835 GHz ISM band, a band shared with other services such as classic bluetooth and WiFi. This band is separated into 40 different radio channels, where each channel

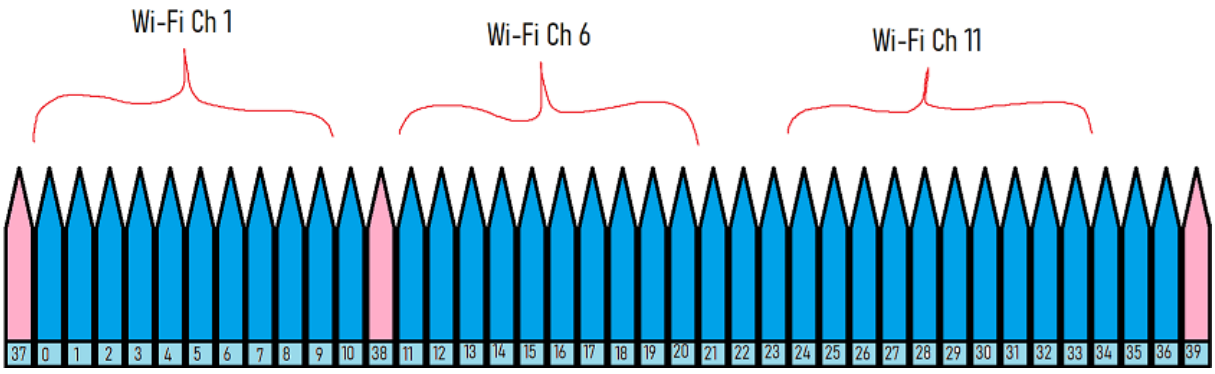


Figure 3.2: BLE advertising channels on the ISM band

is spaced by 2MHz. To avoid collisions the advertisers split up their workload onto three channels. These channels are specifically chosen, to avoid the most common WiFi channels. Illustrated here 3.2 the advertisement channels in pink for channels 37, 38 and 39. While the other blue channels are regular data channels. And the different channels span over 0-9, 11-20, and 23-33.

In terms of contact tracing, there are some bluetooth attributes of particular interest. RSSI, tx power, rx power and interference. All of these attributes are related to estimating distance between two devices, something that is of utmost importance for contact tracing purposes. TX power and RX power is namely the transmission and receiving power, RSSI (Received signal strength indication) goes hand in hand with dBm serving as an indication on good or poor connection. Worse connection leads to lesser probability of receiving packets. Some implementations can use this RSSI to determine the floor of when a connection should be established or not.

The radio waves between two devices react to the environment. In general, the further the distance between two devices are, the worse the signal gets as it is spread over a larger area. However radio waves can also be absorbed or reflect off different objects, either decreasing or increasing the signal respectively. Items containing conductive metals such as copper or aluminum have a greater risk of interfering with the signal. A distance estimation between two phones on the bus could differ from the bus stop even though the distance actually was the same. Human bodies also affect the signal strength so a distance estimation could be greatly affected by which pocket your phone lies in, or if it is in your hand. Different phones have different hardware emitters, and these vary greatly in efficiency. A research made by the developers of TraceTogether, the Singaporean solution we will look at later, put different phones in a anechoic chamber (a isolated chamber with no reflections) and found a huge variance in efficiency amongst the phone-models.

3.2 Decentralized Protocols

Decentralized protocols has been seen to be the favored type of protocol when developing contact tracing applications when you have privacy in focus. The aim of having as much

calculations done, and information stored, locally on the devices instead of on the national backends, is the major difference between a decentral and central protocol. As mentioned previously there are many protocols that have emerged recently, and probably more to come in the future. Some have seen greater success than others in terms of nationwide adoption, and we will therefore in this section, mainly dive into the two protocols that is currently being used. That largely being the DP-3T protocol, and to the lesser extent, the TCN protocol. I acknowledge that the EN protocol from Google—Apple (GAEN) also is a decentralized protocol, but we will look at GAEN in its own section later.

3.2.1 DP-3T

The goal of this section is to get an understanding of how the DP-3T protocol is designed. Developed by a subgroup of the PEPP-PT team that wanted to implement a decentralized alternative. The protocols are open source and can be found at their GitHub repository: [20] Everything below in this section is described from [20] unless stated otherwise. It is separated into three different variants; Low-cost, unlinkable and hybrid, and we will explain all three variants in separate sections below. Possible design flaws associated with either privacy or security will for the most part be overlooked for now, and brought back up later on in the paper under the "Discussion".

In all three of the variants, the devices generate *ephemeral identifiers* which change frequently. These identifiers are then shared through beacons in a typical Bluetooth communication. As mentioned previously the Bluetooth communication in particular is the LE (Low energy) mode, where the device in question broadcasts their beacons, and other devices observe to receive these beacons. The beacon is then stored on the device, together with a time indication and a measurement of signal strength. Also in common through all three variants is the need of a backend server. This backend acts solely as a communication platform and does not perform any calculations. It is considered to be a untrusted part of the architecture in terms of protecting user privacy. When given consent the user can upload a representation of the ephemeral identifier to the backend to be stored. Other devices can periodically query the backend for these representations, which in turn then is used on the device to reconstruct the ephemeral identifier of positive infected users locally. The device can then compare the recorded beacons with the new list of infected to look for potential matches.

Ephemeral IDs

Common in many of the decentralized protocols is the use of temporary identifiers. In DP-3T this is called an ephemeral identifier, or EphID for short. Its purpose of identifying the device is the same in all three variants, but the generation differs slightly.

Low-cost decentralized contact tracing

This variant of the protocol is named low-cost due to its small bandwidth requirements while still retaining a good privacy property.

In this protocol the generation of the EphIDs happens daily through the "secret day seed" denoted as "SK", you then create SK_t for day t, by computing $SK_t = H(SK_t - 1)$ where H is a cryptographic hash function such as SHA-256. SK_0 is calculated by a standard secret key algorithm such as the Ed25519. The device will store the past X days worth of secret day seeds, where X usually is a 14 day period, determined by the health authorities. The duration of which a device broadcasts the same EphID is referred to as an **epoch**. The length of an epoch is defined in minutes, denoted as L. Each day, the device computes

$$EphID_1 \,||\, ... \,||\, EphID_n = PRG\,(\,PRF\,(\,SK_t,\ "broadcast\ key"\,)\,)$$

where **n** is the amount of EphIDs to be created for the day through (24*60)/L. **PRF** is a pseudo-random function,"**broadcast key**" is a fixed public string, and **PRG** is a pseudorandom generator producing n*16 bytes. The bytes are then split into chunks of 16-bytes each to obtain the n EphIDs for the day. The devices selects a random order to broadcast the EphIDs throughout the day.

As mentioned previously when other devices receive a beacon under the low-cost protocol they store the raw EphID, the exposure measurement, and the day of which the beacon was received.

When a user has been confirmed as a positive infected user they instruct their phone to upload the seed SK_t and the day t corresponding to the first contagious day. After uploading the user would then generate a completely new seed SK_0 to avoid tracking. Every other user can then download the SK_t seed, and generate $SK_t + 1$, $SK_t + 2$ etc. for the contagious window. Once they have the seed they can compute all EphIDs for the days t, $t + 1$, $t + 2$ ... and check for potential matches with the EphIDs they have in the list of recently seen EphIDs. For each match the user would additionally upload the day and exposure measurement to compute a risk analysis.

Unlinkable decentralized Contact tracing

The second variant is the unlinkable decentralized Contact tracing protocol. It is similar to the low-cost but offers greater privacy properties in return for a larger bandwidth cost.

The EphID generation is a bit different. For this protocol the EphID for epoch i is defined:

$$EphID_i = LEFTMOST128\,(\ H(\ seed_i\ )\ )$$

Where w is a cryptographic hash function, the seed is a new random 32-byte value for each epoch, and LEFTMOST128 just refers to the first 128 bits from the hash output.

When other devices receive a beacon using the unlinkable protocol they store a hashed string

$$H\,(\,EphID \,||\, i\,)$$

together with the exposure measurement and the current day. Note that the EphID is now stored as a hash instead of raw as it were before, additionally including the epoch. This is primarily due to security reasons.

When confirmed positive the user gives instructions to upload to the backend, this protocol enables the option of removal of identifiers. Given that the identifier is now tied with the epoch, a user could limit the EphIDs they wish to share with the backend. Then, once uploaded to the backend, the backend utilizes a Cuckoo filter in order to store the

$$H\,(\;LEFTMOST128\;(\;H(\;seed_i\;)\;)\;||\;i\;)$$

You can read more on the cuckoo filter here, [21] the result of using this filter in a per-user regard, is that the lookup property reduces the computational cost when compared to the low-cost seed generation.

Hybrid decentralized Contact tracing

The hybrid decentralized Contact tracing protocol is a design that acts like a middle ground between the first two variants. In this variant the devices generate random seeds for a time window w. Given the epoch L the time window needs to be a multiple of L in order to create w/L EphIDs. w can vary from minutes to a full day. A w of a full day would represent a design that is very similar to the GAEN design, as GAEN took high inspiration from this protocol particularly.

The EphID generation is almost identical to the low-cost, but with a different fixed public string. For each time window the device would compute:

$$EphID_1\;||\;...\;||\;EphID_n = PRG\,(\,PRF\,(\,seed_w,\;"DP3T-HYBRID")\,)$$

where again **PRF** is a pseudo-random function, **DP3T-HYBRID** the fixed public string and **PRG** the pseudorandom generator. As in the low-cost protocol the devices picks a random order to broadcast the identifiers.

When the devices observe and receive other beacons in similarity to the low-cost they would store the EphID, the exposure measurement and the time window w in which the EphID was received, giving a higher accuracy than the other two models which only stored in terms of days.

If the user is diagnosed positive and is going to upload to the backend, they can with this model choose to remove certain time windows they do not wish to upload. In time windows where the user did not observe any EphIDs close enough to be considered as an exposure the corresponding seed is automatically removed. Other users would as described in the low-cost protocol download the seed for time window w and reconstruct the EphIDs to look for potential matches.

3.2.2 TCN

TCN is the other protocol that has seen some use. There are a handful of solutions based on the TCN protocol. If a nationwide adoption is primary, these solutions are secondary/second hand. Some US-states and solutions in France, Germany and Italy use TCN.

This decentralized protocol is centered around Temporary Contact Numbers, or TCN for short. Similar to the Ephemeral Identifier from DP-3T, the TCN serves a purpose

of being a shareable identifier to the individual user. Developed by the TCN Coalition network and described on their GitHub: [22] Originally having a rather basic model of: "Generate a random TCN, store the TCN, and broadcast it using Bluetooth." [23] referred to as "Strawman Protocol" providing good server privacy, receiver privacy and preventing passive tracking, but additionally does not prevent a user from observing another user's TCN and stealing it. The model itself also poses a scalability problem as a user has a list of every TCN they broadcast, and every user need all lists that exists. To address some of these issues the "TCN Protocol" as we know it was implemented. Primarily changing from random TCN to seed-based generation in similarity to the DP-3T solution.

The protocol goes as follows: A key-pair *report authorization key* rak and *report verification key* rvk is made, can be derived through Ed25519. Then, to create the initial *temporary contact key (TCK)* tck_1:

$$tck_0 \ <- \ H_tck(rak) \qquad , \ tck_1 < -H_tck(rvk \ || \ tck_0)$$

where H_tck is a domain-separated hash function with 256 bits of output (often SHA-256). Every tck_i other than the first initial tck_0 can be computed as

$$tck_i \ <- \ H_tck(rvk \ || \ tck_{i-1})$$

Then a temporary contact number is derived from a temporary contact key as such:

$$tcn_i \ <- \ H_tcn(le_u16(i) \ || \ tck_i)$$

where H_tcn is a domain-separated hash function with 128 bits (could be a truncated SHA-256).

Once an infected user wishes to upload their list of infected TCN they create a report including the tck necessary for other users to re-recreate the infected users $TCNs$ locally. Also included in this report is the original rvk for that period, and the report itself is signed with the rak, verifying the source integrity for the other users.

3.3 Centralized Protocols

Centralized protocols differ from decentralized protocols in that the backend server has a larger role in the architecture. The backend will usually do more of the heavy lifting in terms of computation. There is also often some sort of registration necessary in the centralized approaches. Contacting the backend is a necessity to obtain exposure status in centralized protocols as the verification operation is performed on the server-side. The centralized versus decentralized discussion is something we will return upon later in this paper. However the aftermath of that discussion has led to a large adoption of decentralized protocols. Implying that the adoption of the centralized protocols is rather small. We will take a look at the ROBERT protocol, a project from the PEPP-PT consortium, and we will also dive into the BlueTrace/OpenTrace protocol.

3.3.1 ROBERT

ROBust and privacy-pres**ER**ving Contact **T**racing protocol, ROBERT, was originally a proposal for the Pan European Privacy-Preserving Contact Tracing (PEPP-PT) initiative, developed by a collaboration between the French Inria and German Fraunhofer AISEC. It is now currently adopted in France as one of the only centralized approaches in Europe.

As mentioned before the centralized approaches rely more on backend. The server-side has a initialization phase where the server key **Ks**, and registration key pair **SKs** (private key) and **PKs**(public key), is made. The server key **Ks** and private key **SKs** is only known to the server, and the key-pair is defined over the elliptic curve NIST-P256.

After a user registers, the server creates a entry in its local database. In this entry the following information is stored for further user:

K_A^{auth}	Authentication Key for user A	Key for authenticating messages from A
K_A^{enc}	Encryption Key for user A	Key used to encrypt from server to A
ID_A	Permanent Identifier for A	Identifier for A known only to server
UN_A	User A Notified	True/False if user A is flagged "at risk"
SRE_A	Status Request Epoch for user A	Last epoch when user requested a status request
LEE_A	List of Exposed Epochs	List of epochs where A's EBID was found in an infecteds list

Table 3.1: Data base entry per user

The app generates a ephemeral key pair, and uses its private key together with the servers public key PKs to generate the "SharedSecret". The server confirms with the app's public key and the server private key SKs. From the SharedSecret the authorization and encryption keys are given value as follows:

$$K_A^{auth} = HMAC\_SHA256(SharedSecret,\ "authentication\ key")$$

$$K_A^{enc} = HMAC\_SHA256(SharedSecret,\ "encryption\ key")$$

These keys are then part of what is referred to as "HELLO" messages. These messages act as the beacons we have seen in other protocols. When communicating on an app to app basis the messages are sent through Bluetooth low energy, in similarity to other solutions. The user is given a ephemeral bluetooth identifier, "EBID" for short, an encryption of the **ID_A** from the users database entry, and the server key **K_S**. This EBID is used to create a Encrypted Country Code, which is also a part of the "HELLO" messages, but is mainly used for interoperability between back-ends from other countries. The "HELLO" messages that are continuously being broadcast and received is composed of four parts; Country Code — EBID — Time — MAC. Where the MAC is a authentication of the other three parts hashed with the authentication key of the user that broadcasted said message. A confirmed infected user would notify the server that they are in fact infected. It should be noted that the protocol assumes this communication goes through a trusted health authority on trusted servers. Based on time, the past X EBIDs from that user would be marked as infected EBIDs. When other messages are received they are put

into a local "ContactList", storing every EBID the user has seen the last 14 days. The application regularly sends a request to the server to compute their exposure status. The server then goes through the *LEE* (list of exposed epochs) of all infected users, to check if the ID of the user that requested a exposure status is found. If the ID is indeed found the user would then be flagged "at risk" following the UN_A value in the database entry. It is also worth noting that in this protocol there is no distance estimation from bluetooth data, it is simply a binary check of received messages from other infected users.

3.3.2 BlueTrace

As mentioned previously TraceTogether is one of the first digital contact solutions that was developed in the beginning of 2020, made for Singapore. "TraceTogether is the first national deployment of a bluetooth-based contact system in the world." [19] Given that the solution looked promising other governments took interest. They wanted to adopt or adapt the solution for themselves. In response to this interest the team behind TraceTogether made a protocol BlueTrace, which have been adopted by countries such as Australia and Fiji. In addition to releasing the protocol they also made some "default" implementations called "OpenTrace" for both android and ios separately, which can be found on their GitHub page OpenTrace. [24]

BlueTrace as a centralized approach starts with the registration. When a user registers they put in their phone numbers. The back-end then generates a randomised unique UserID associated with the given number. The phone number's primary function is to allow health authorities to contact the infected person for further followup if they need. Which is a issue we will talk more about later in this paper. Also in contrast with the ROBERT protocol, the back-end in BlueTrace effectively has some PII (Personally identifiable information) on its users. Which serves many issues we will discuss later on and not in this section. As we have seen with many other solutions, the use of temporary identifiers for the bluetooth app to app communication is a standard approach. The "TempID" generated in the BlueTrace protocol is a combination of the randomised UserID from registration, the creation time for the TempID and the expiry time for the ID. These three parts are encrypted with AES-256-GCM, then added on, is a random Initialisation Vector (IV) and an "Auth Tag" for integrity checks. These five parts is what comprises the TempID, and are then Base64 encoded by the health authorities. The health authorities holds the secret key to encrypt and decrypt the TempIDs. As with other temporary identifiers, BlueTrace recommends its implementations to choose a shorter lifetime, e.g. 15 minutes. This is primarily for security reasons which we will discuss later. The back-end supplies its users with batches of TempIDs at a time, to prevent users with unstable internet connection from missing valid TempIDs.

BlueTrace has the devices take on either Peripheral or Central roles. The Peripheral advertises its "services", which is a collection of data, and the central scans for these advertisements. It is only the Central role that can perform more intricate actions in the communication. The central can read off the data from the peripheral, in addition to write data back, while the peripheral only advertises the data it has. Only the device in central can read the RSSI from the communication, which is information that can be

used to estimate distance. Thus, when a connection is made between two devices (one in peripheral and one in central), the central would write the RSSI back to the peripheral in order to have symmetric information between the two devices. This symmetric information setup is a clever way of trying to preserve resources, as being in central is more consuming than peripheral. A device such as a phone can act as both central and peripheral at the same time, and BlueTrace recommends a duty cycle of 15-20% in central, and 90-100% of peripheral. "Duty cycle" is a term that defines the time a device is active in relation to the time it is inactive. When BlueTrace recommends a peripheral duty cycle of 90-100% that effectively means that a device should always listen, but should only send with an active time of 15-20%. Most importantly is that the sum of the duty cycles is greater than 1. This is to ensure there are no setting where two devices cannot detect each other. When a device has seen another device, BlueTrace implements a "blacklist" to avoid multiple connections to the same device in rapid succession. The devices seen is put in this blacklist for a scanning cycle or two, and additionally in a encounter history for a set number of days depending on the implementation.

The encounter history can be uploaded to the health authority backend, but an authorisation code from the health authority is needed in order to do so. The health authority can decrypt the TempIDs for the different encounters to obtain the UserID, and determine risk analysis by looking at information such as time of exposure and an estimate of distance. This system was designed for the information to be used alongside manual interviews (where the phone number comes in) with the users in hope to determine individuals with higher likelihood of exposure. The BlueTrace developers has acknowledged that the protocol could have been decentralized by removing the process of manual interviewing, and thus deleting the phone number storage in the registration, but consciously chose not to. The protocol recommends withdrawal of consent to be supported by the implementations.

3.4 GAEN

Many of today's solutions are made from the Google—Apple Exposure Notification *(GAEN)* framework, often referred to as "Exposure Notification System" (ENS). An API made from the joint effort of Google and Apple in the aid for interoperability between Android and iOS smartphones. Released May 20th 2020 it is very inspired by the hybrid approach seen in the DP-3T protocol, and additionally brings in aspects from the TCN protocol. Which both are protocols we have taken a closer look at in this paper, and both of which was made public some time in April. It is as mentioned previously a decentralized protocol, but uses a centralized backend, if consented, to distribute keys about infected users. As with all other contact tracing protocols discussed in this paper GAEN also utilizes the BLE specification. Apple released documentation on bluetooth specification amongst other information about the API here: [25] Google refers to the same specifications from their webpages.

3.4.1 Advantages of Operative System Access

Due to previous developers of other solutions only having access to the application layer of the phone, they met aforementioned implementation issues, mainly tied to iOS devices. They discovered that iOS devices having the contact tracing application in the background would not broadcast properly, only broadcasting formats readable for other iOS devices resulting in a lower cross device interoperability[19]. Additionally the device will not be able to sufficiently scan for other devices when in the background. Having the application in the foreground but with the screen turned of effectively meant the same as having it in the background. As a workaround users were prompted to have the application in the foreground with the screen on, and were advised to turn their phones upside down to enter "power saver mode"[19]. I think it is fair to assume that the efficiency behind this "upside down in your pocket" solution is significantly reduced when it relies on people not using their phones while outdoors.

However, as Apple and Google are organizations who develop a large share of the current phone market, they have access to more resources of the devices compared to other developers. Patching this bluetooth detection issue on the iOS devices was possible for their released solution. But is not something that was shared with the custom developers. GAEN has been able to be adopted, allowing for some smaller modifications to be done by countries for their solutions, however Google and Apple are strict with the policies around privacy before a health authority is given access. Each country is typically only allowed to deploy one application based on GAEN, and may only be released by the public health authority.

3.4.2 Adoption

The GAEN framework has been sought out by countries even though they might already have had solutions in place. The bug fix for iOS devices not properly working in the background is a major efficiency upgrade compared to custom solutions. A research done in the UK during development for their custom solution: "Specifically, the software registered about 75% of nearby Android handsets but only 4% of iPhones." [26] Where they also stated that the GAEN solution found 99% for both handsets. As of writing there are 67 different applications utilizing the GAEN framework. In the US, each state is given the decision on how they wish to combat the contact tracing challenge. Currently there are 25 contact tracing applications using GAEN in the US[27]. The limitation of one applications per country also applies for the states, meaning one state can only have one authorized GAEN application. An interesting observation is that the majority of states adopting for the GAEN contact tracing solution also are politically democratic states.

3.4.3 Protocol

As mentioned previously the protocol itself is inspired by the hybrid approach from DP-3T. If we recall, the idea is to create temporary bluetooth identifiers with a lower lifetime of e.g. 10-20 minutes. These identifiers are created from a daily rotating "seed key". In GAEN these daily seeds are called TEK (Temporary Exposure Key), and the temporary identifiers are known as RPI (Rolling Proximity Identifier). As mentioned app to app communication goes over BLE. The BLE payload contains the current RPI valid for the timeframe, together with some "Associated Encrypted Metadata" which contains data such as versioning etc, but most importantly the transmit power for the payload. This metadata is encrypted by the TEK, and only decrypted

once a positive diagnosis is confirmed. The bluetooth communication is defined such that a device will broadcast their packets every 250ms[28], but will only scan for packets "minimum every 5 minutes" according to the official documentation. In practise, different working GAEN solutions typically scan for packets between 3.5 and 5 minutes. As mentioned the payload contains the RPI (called "tempID" in Bluetrace, or "ephID" in DP-3T) bluetooth identifier. This identifier is meant to be changed synchronously with the devices address and metadata. The simultaneous synchronization is mainly a security feature towards linkage attacks, but is not important for our paper. As in the DP-3T protocol, we first generate some initial seed key "TEK" from a cryptographic random number generator. This seed key is effectively a way of tracking the time of which the TEK and RPI is viable. Then a RPI key (RPIK) is generated using a HDKF such as SHA-256. From this RPIK the RPI for the current interval is derived from AES-128 of the RPIK and some data padding. The encryption of the "Associated Encrypted Metadata" mentioned previously is also done in a similar fashion through AES-128, and a HDKF such as SHA-256. Usage of SHA-256 and AES seems to be common practice in developing secure procedures for these contact tracing solutions, therefore we will not use more time on repeating similar formulas. The formulas themselves can be found here for further reading: [29] When a user has a positive diagnosis they can choose to upload the respective TEKs for the relevant time periods to the trusted backend server. The server will then distribute these TEKs to the other users such that they can re-generate the RPIs and evaluate potential matches locally on their devices. If a RPI match is found the "AEM" will be decrypted to determine risk score based on distance from the RSSI in the metadata.

3.4.4 ENX

The protocol we just saw is the ENS, or Exposure Notifications System. However in September 2020 Google and Apple released the Exposure Notification Express. In short, a public health authority provides the ENX with a configuration file that contains instructions regarding user messaging and risk parameter calculations, and the ENX generates a custom application in return. This substantially decreases the deployment time when a health authority wishes to adopt some digital contact tracing solution. This update also included an option for iOS devices to activating ENS by a simple "opt-in" feature in the phone settings rather than downloading an application. With the faster deployment and easy access, GAEN saw an substantial increase in its users [30].

Technologypower

GAEN is as said before made by Apple/Google, two large globally known companies. Any changes they make regarding the GAEN framework can affect every other implementation that relies on the GAEN-api. Given that Apple effectively controls access to efficient Bluetooth contact tracing on iOS devices, they are made indispensable in a global setting in terms of efficient contact tracing.

3.4.5 Information Privacy in GAEN

Google and Apple has released papers regarding GAEN and its surrounding development, including a paper on privacy preservation [31]. Given that GAEN solutions have become the new norm in digital contact tracing, the privacy preservation has been deemed well enough to become

used globally, and Google—Apple can claim with good conscience that the solution is privacy preserving. However, GAEN solutions running on android devices is handled through Google Play Services, which in short is an application handler on android devices that has elevated privileges. The documentation on play services is severely lacking. In a June 2020 study on GAEN data handled by Google Play Services[32], they claim that Google Play Services pings google servers every 20 minutes containing information such as your WiFi MAC address/IP address thus being able to roughly determine the location of the user. Neither Google nor Apple would confirm these accusations so it is speculation, but the fact that GAEN can not operate on android devices without Google Play Services, effectively gives Google the opportunity to collect data on their users, while they can still state that the solution "GAEN" is privacy preserving.

3.5 Data Privacy

Data privacy, information privacy or data protection are all synonyms describing what information should be shared, with whom and what procedures that should be in place for the collection and storing of said information. In recent years privacy have been more and more in focus for developing public services. Rules and policies such as the GDPR, CCPA, HIPAA and DPA 2018 detail different rule sets and procedures good privacy-preserving systems or architectures should adopt. In terms of digital contact tracing data privacy is of great concern. We have already seen by the protocols that data security is done by state-of-the-art cryptography schemes, however the data privacy differs from the different protocols and implementations. How does one implement applications that observes who is in proximity to whom, while at the same time preserving the identity of whom you are observing, and the people they are in proximity with? This is one of the questions the developers of good privacy-preserving solutions are trying to solve. We will look more into the privacy-preserving debate in the discussion chapter.

Chapter 4

Model-program

In this chapter we will represent a model program based on networking. Therefore we will firstly explain some terms as background knowledge in network theory and social networking as these will be good to know when discussing the program later.

Vertex/Node	An object in the network
Edge/Tie	Connection between two nodes
Node degree	Number of connections that node has to other nodes
In degree	Number of nodes pointing to the node
Out degree	Number of nodes pointing out of the node
Undirected graph	The order of connected vertices is unimportant
Directed graph	A connection has a direction from one node to the other
Average degree	Average number of links per node
Network Size	Number of nodes in a network
Homophily	Tendency of nodes to make ties with similar nodes versus dissimilar
Density	Number of direct ties relative to the total number of achievable ties
Triadic closure	If connection A-B and B-C exists, A-C is more likely to be formed
Distance	Minimum number of ties required to connect two nodes
Diameter	The longest achievable distance in the network
Clique	A group of nodes where every node has a direct tie to every other node
Clustering coefficient	Measure of the degree of which nodes tend to cluster together

Table 4.1: Social network attributes and definitions

4.1 Agent Based Modeling

My approach on gauging the potential efficacy of digital contact tracing is a simple agent based model(ABM) programmed by myself in java. In this section I will explain roughly what agent based modelling is, and why I chose it for my model. Agent based modelling is a modelling approach where you give individuals certain attributes, behaviour or traits in order to see how it alters the outcome. Often the model is run where there are individuals called "agents" that do some autonomous decisions based on probabilities. This model leverages computation power instead of pure mathematical methods. A model that is similar to ABM is metapopulation

modeling where you would see behaviour on larger groups instead of individuals, e.g the NIPH has a metapopulation model to estimate the basic reproduction number in Norway[33].

Given that I am writing my thesis under the Department of informatics, and not the Department of mathematics, a programming model suited me better as a platform to learn more than a mathematical approach. Standing with the choice between an individual or metapopulation based model the decision was easily ABM for me. I found that agent based modelling better suited my wishes. I had an idea as to implement a network of random people, random in the sense that people have varying values from each other, then spread some arbitrary disease throughout the network, and look at how many cases we get when we alter the different attributes. Primarily the attribute we will focus on is to give the individuals phones, which puts them into the category of available digital contact tracing users. The effect we hope to achieve is to put users in earlier isolation and see how that affects our network compared to situations where we have more or less digital contact tracing. There exists open source simulation tools such as SERIA [34](*use described in this paper)*, AnyLogic which is a tool you could use for smaller experiments: https://www.anylogic.com/, and multiple other simulation tools found: [35]. But in order to have full control over the experiments I want to do, I chose to make a simple simulation myself.

4.2 Model-program

I developed a simple program to simulate a population (Source Code: Appendix A). The goal of this program is to notice tendencies from my results after I alter the parameters going in. The tendencies I am interested in testing, and which I focused the development around is: 1) The difference in a uniformly distributed population versus a non-uniformly distributed. 2) Different adoption rates of digital contact tracing and their results.

My program is divided into four classes. "Main" will run simulations by calling a simulation "Sim", the "sim" will need a set of nodes "Node" which is created by the nodegenerator "NodeGenerator". Most of the logic lies in Node and Sim.

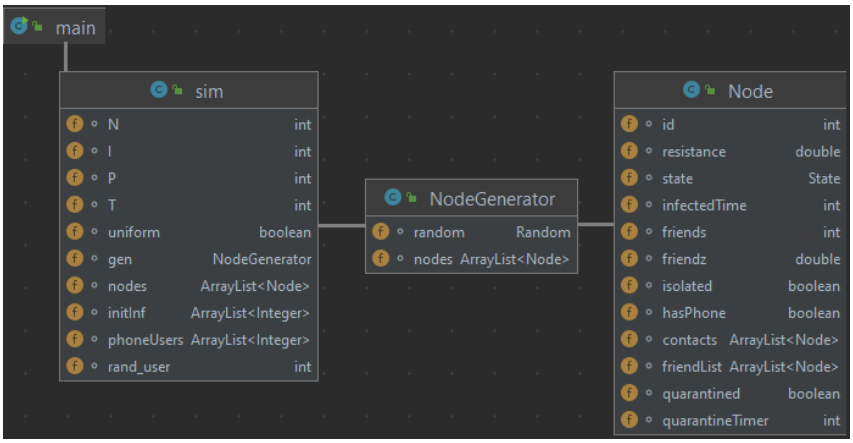


Figure 4.1: Program Structure

4.2.1 Initialization

Much of the logic in agent based modeling is related to the agent, which in our case is the *node*. Here is the parameters of interest related to the node object 4.2:

id	int, identifying the node
resistance	Double, a value representing genes
state	enum, what state the node is in S-I-R
infectedTime	int, time since the node was infected
friends	int, number of friends the node should have
quarantimeTimer	int, time left in quarantine
contact list	ArrayList<Node>, list of nodes the node has met
friend list	ArrayList<Node>, list of friends the node has
isolated	boolean, if the node is isolated or not
hasPhone	boolean, if the node has a phone for DCT or not.

Table 4.2: Node Attributes

Firstly I initialize a simulation with the parameters N,I,P,T,*uniform*. Where N is the network size, I is the initial number of infected nodes, P is the number of users having a phone for digital contact tracing, T is the number of days the simulation should last, and *uniform* is a boolean value stating if the distribution of friends per node is uniformally or normally distributed. A node is initially in the "Susceptible" state, except for the I nodes that are made infected. My epidemic model follows a S-I-R model, where a node goes from being susceptible to possibly infected, and then recovered. A recovered node remains permanently immune. After

the network is created, the infected people, and people having phones P are chosen randomly. A node selects a set number of "friends" in the network, where the number of friends follows *uniform* if it is a uniform or left-truncated normal distribution. The reason for the truncation here is that normal distributions with a low probability can falter from the mean value, therefore any value falling to the negatives will be rounded up to 0.

The initial infected people I is a static value set to 4. Instead of having just one "patient zero", I chose to have multiple to reduce situations observed during development where the simulation dies out. The parameter P refers to the number of people having a phone. App uptake can then be calculated as P/N, in modern contact tracing the realistic upper bound uptake of modern western societies is 60%, however we will be testing for all values 0-100% with a 10% interval. The attribute itself of having a phone in the simulation is defined to earlier isolation. If a infected user has a phone they will isolate and quarantine their close contacts one day earlier than the non-phone users. This is a reflection of what I think should be the ideal in the future. There will be a critique section later before we look at the simulation results, but I will repeat here that an assumption is made that DCT (digital contact tracing) has a 100% detection rate in my simulation.

4.2.2 Per day iteration

Following the simulation parameter T, we have a set number of days of which the simulation will run. For each day T, each node n will go by their day. The node will choose a random number between 0 and the size of their friendlist minus one, representing the number of contacts they will make that day. The "friends" term here is a merging of all the different contacts a person usually makes throughout the day, either being family, work, school etc. Additionally to meeting this random number of "friends", a node will also meet a totally random node at a 75% chance.

To elaborate on this "random encounter": The value of 75% was set during testing as it gave me the behaviour I was looking for. Secondarily given when *uniform* is false we follow a left truncated normal distribution where the number of friends lie in the range of 0 to 10. There will be values larger than 10, but the average over all nodes is calculated to be approximately 5, which is the same as when the distribution is uniform. If the number of friends for a node is one, then the node will make a exclusive clique with another node with a friendlist size of one. Resulting in a edge case where these nodes effectively could only infect each other if one of the nodes were amongst the initially infected set I. However, by having this random encounter the clique can get infected externally.

When meeting other nodes, either it be through iteration over the people they will meet, or random encounters, a edge is made. Edges represent connections between nodes, and have an id starting with the day they met. Implementation-wise this is to ease the process of deleting edges when expired. As seen in the protocol section, most currently working solutions store their temporary bluetooth identifiers for 14 days, following the health regulation norm of quarantining people 14 days after exposure of potentially infected. Inspired by that an edge between two nodes will be deleted after 14 days as well. After edge creation, if one of the nodes in the edge is infected, a computation is done to determine if the other node will be infected.

This computation is simple and artificial. It is not meant to represent real world disease calcu-

lations. Every node has a resistance value weighted towards 0 in the range of 0 to 1, and the calculation is simply

$$resistance - Math.random() > 0.25$$

where Math.random() returns a random value between 0 and 1. By this formula the higher resistance a node has, the less likely they are to be infected. The threshold 0.25 was chosen after iterating and testing during development as it gave me the desired behaviour. After being infected the node is set as "infected" for 14 days, but will after 6 days without a phone, and 5 days with a phone, test themselves to self-isolate and quarantine their contacts. Generally in this simulation, I repeat that the focus is to observe the tendencies of when we increase P relative to N, and if these results vary from a uniform or non-uniform distribution of friends. Therefore the way of transmission, chance of contagion, and other systems one could implement are not as important. The most important factor here is that the systems that *are* implemented are constant, and behave in similar manners for all scenarios.

Isolation

When the node has been infected for 6 days it will isolate itself. I chose 6 to average the time it takes a node to get infected, develop symptoms, and then test themselves and isolate. This combined time of incubation, latent period and behavioral pattern of actually going to test yourself varies from different sources. I mainly took inspiration from the CDC, WHO and NIPH to determine this value. The node will after "testing successfully and isolating" notify 80% of the people in its contact list. This is to simulate faulty manual contact tracing, as the index case cannot know everyone they have possibly infected. If the node that is isolating has a phone, they will isolate one day earlier. This is to represent an ideal digital contact solution that does in fact give you notifications quicker than a manual one. In my approach digital contact tracing has a 100% detection rate, implying when a isolating node has a phone, every node in its contact list that also has a phone will quarantine. This is to represent accurate digital contact solutions. Where by "accurate" I refer to a close to 100% detection of contacts that should be detected. In both cases a individual has a 40% chance not to follow quarantine and thus act a "normal day" of interactions. This follows the human failure to adhere to quarantine stated [36]. It is worth noting that my model does not account for false positives or false negatives.

4.2.3 Results

After each day the number of infected people is counted and added in a list of results, netting us both the number of infected people on a daily basis, and the number of total infected after summing over all days. In this section I represent some visual plots of some selected data based on varying parameters. The plots are an average from "runNumb" simulations for the different variables, this is usually 1000 but will be lower values for larger N's, as running 1000 simulations for large N's take substantial time. For further reading on the results I have added an appendix B with some data sets of different simulations.

Figures

In this subsection we will display some figures to show different characteristics I feel are important to showcase. All the data generated is in Appendix B where you can replicate or create new figures. The figures follow the following format:

- There are different scenarios where the population has different uptakes (Users with a phone). These are separated by 10% increments starting from 0 all the way to 100.
- These scenarios are color gradients where the lesser blue colors have higher uptake than the darker blue in the same group.
- The red plot signifies a scenario with no isolation
- The green plot means 100% adoption rate
- The orange plot means 0% adoption rate
- The plot follows daily new cases by day

In order to maintain *some* readability the plots end after the first 20 days even though there are data for the first 50. I have yet to find occurrences of second waves that are worth plotting, and plotting all 50 days for every scenario would make the peak values less readable. Some of the later plots will just showcase the peak values for readability as well.

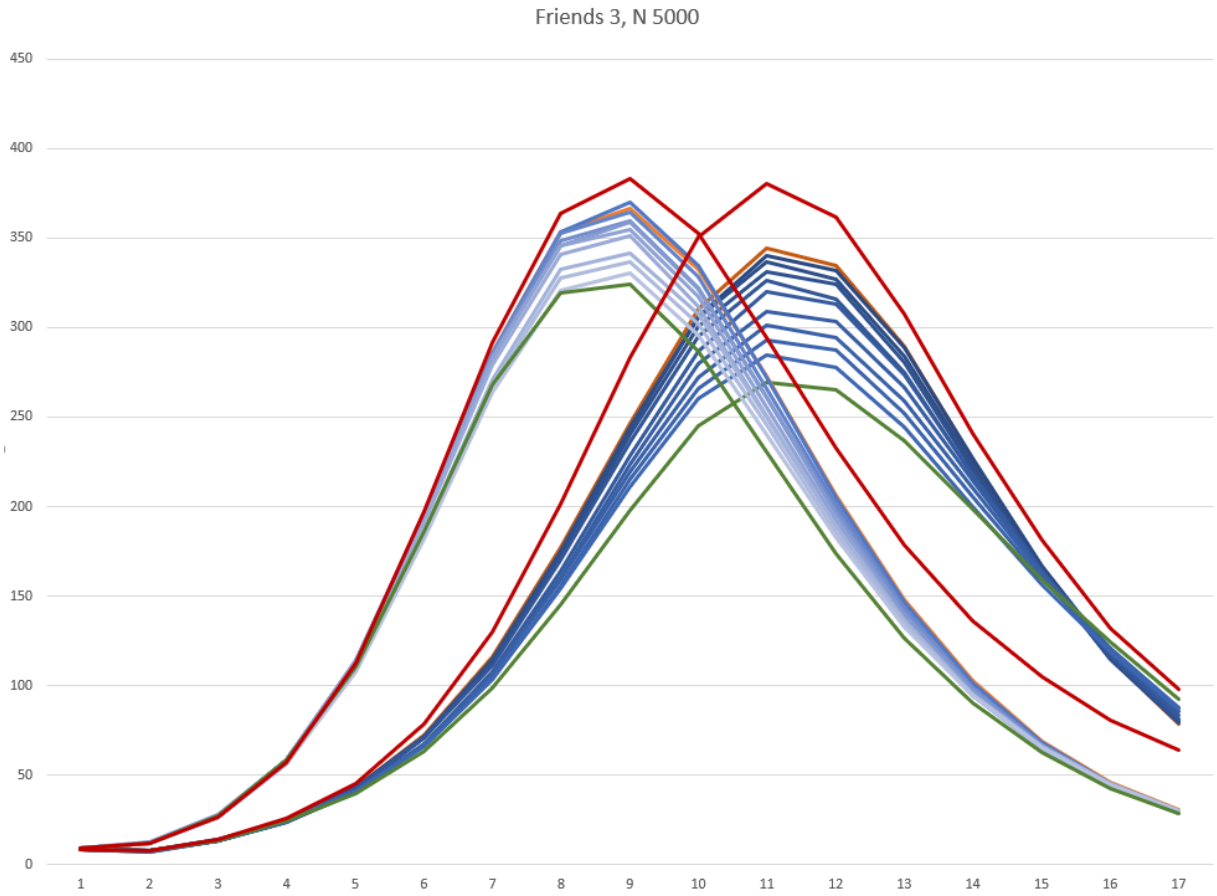


Figure 4.2: Friends 3, Distributed left group, uniform right group

In 4.2 we observe that in both groups the tendency of daily cases decreases the higher the uptake. The efficacy of the DCT solution is significantly higher in the uniform group compared to the distributed. We also observe that by comparing the data set for Friends3 yields us a "linear" effect of 10-30 cases avoided per 10% uptake, which is a lot less than we would expect with the more "exponential" curves displayed in [17] or the "1% uptake gives 0.8-2.3% decreased infection size" statement by [36]. Some reasons as to why this could be will be discussed in the critique section of this chapter. What is interesting though is that the distributed group spikes faster than the uniform by 2-3 days in every scenario.

If we increase the cliqueness we get this figure 4.3:

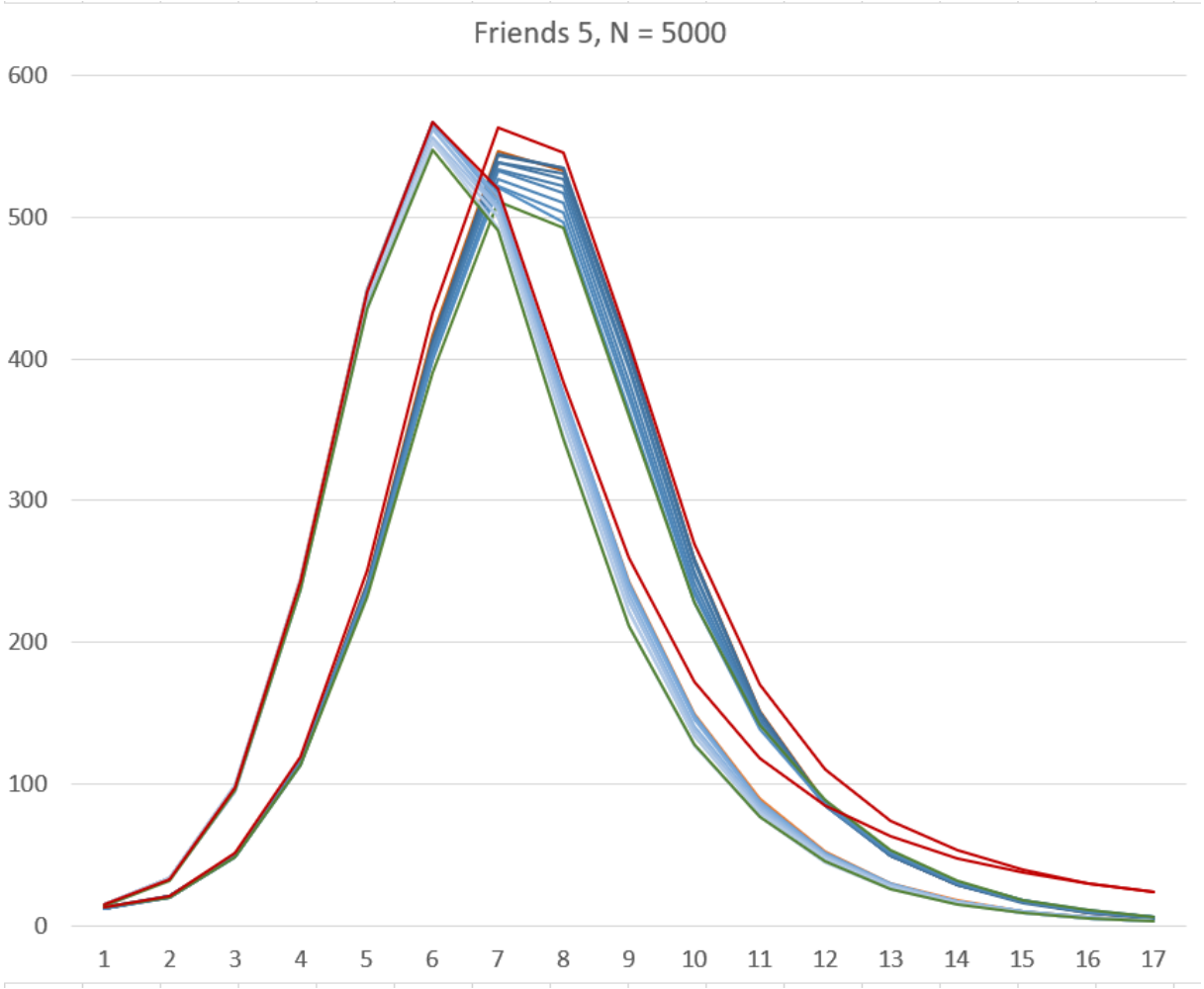


Figure 4.3: Friends 5, Distributed left group, uniform right group

As we can see when the number of friends increases the graph gets narrower both in height and length. The number of new daily cases spikes faster and higher compared to 4.2, but the total infected after 50 days is similar in size as in the simulations ran with friends = 3. The DCT effect is still notably higher in the uniform group than the distributed. Here is the same plot with just the scenarios no isolation, 0% and 100% uptake 4.4:

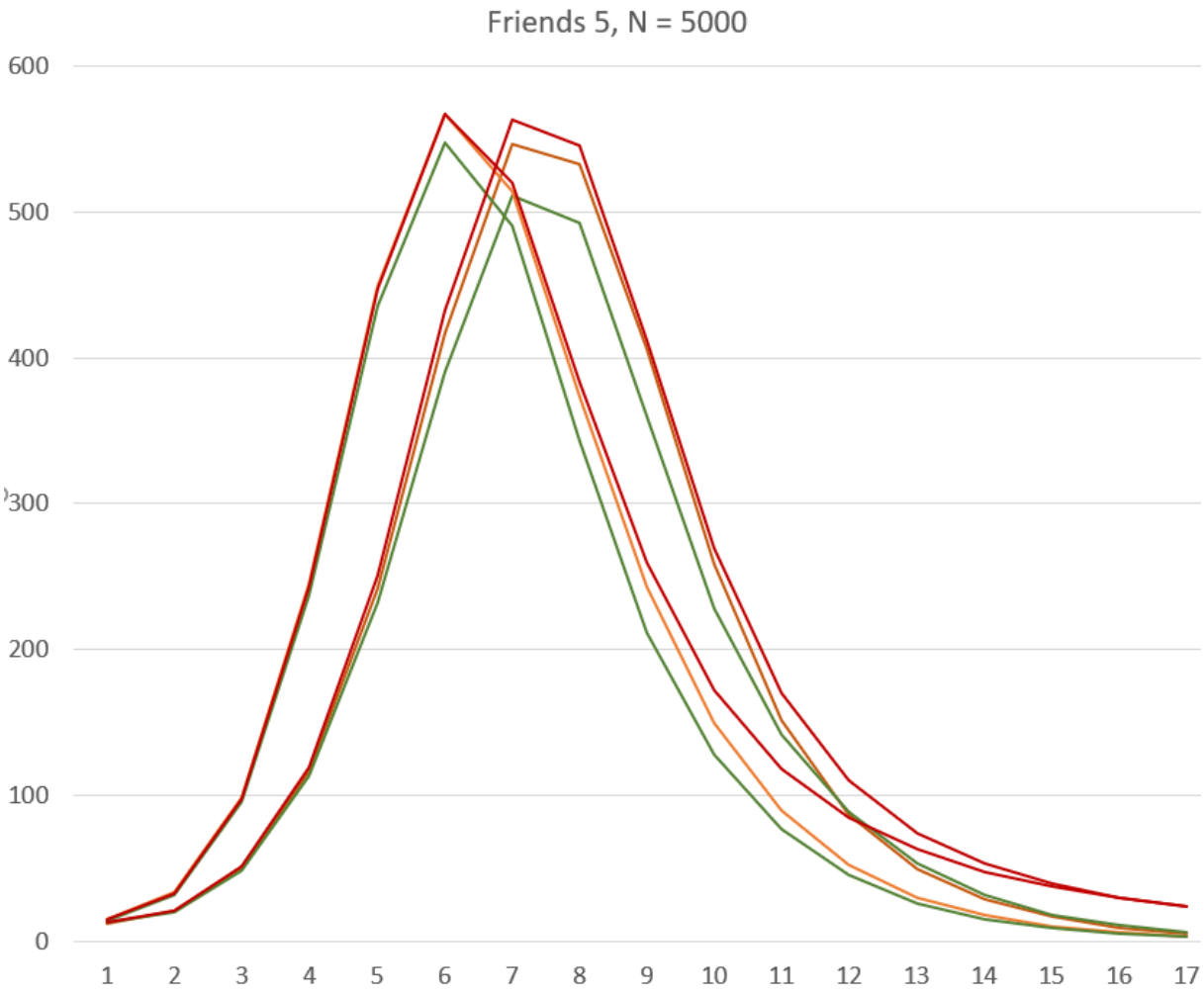


Figure 4.4: Friends 5, Distributed left group, uniform right group

We will notice that the infection size is in the range of 54.6% of the population at the lowest (with friends = 3, non-distributed and 100% uptake) and 66.7% at its highest (friends = 8, distributed no isolation). The infection size generally with no isolation is in the range of 65-66% of the population after 50 days. The effect of uptake is usually 0.1-0.5% reduction in infection size per 10% increase. Varying on friends number (or cliqueness). Although friends3 varies from the norm and sees some higher reductions (up to 0.8% difference between one scenario and the other). These are some simple computations done by examining the sample data from the CSV files in Appendix B.

If we increase the cliqueness even further shown in4.5:

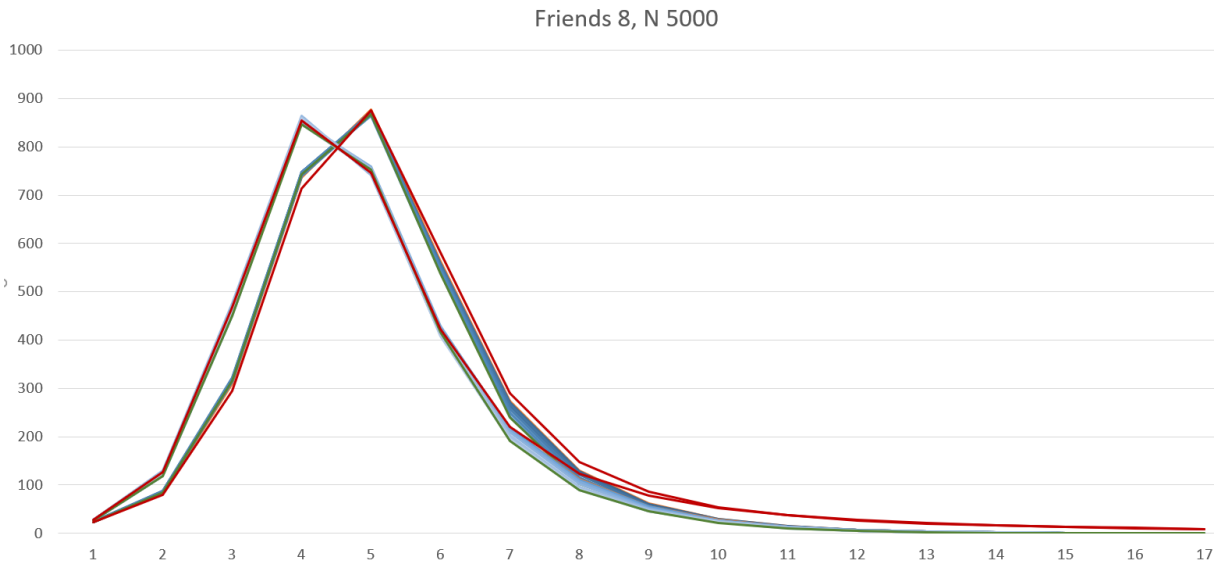


Figure 4.5: Friends 8, Distributed left group, uniform right group

we observe that the daily infections spike even faster than before. Thus the tendency of higher cliqueness gives a faster spike can be observed.

If we go back to a more standard cliqueness of 5 we can observe if changing N to be larger values than 5000 has an effect 4.6:

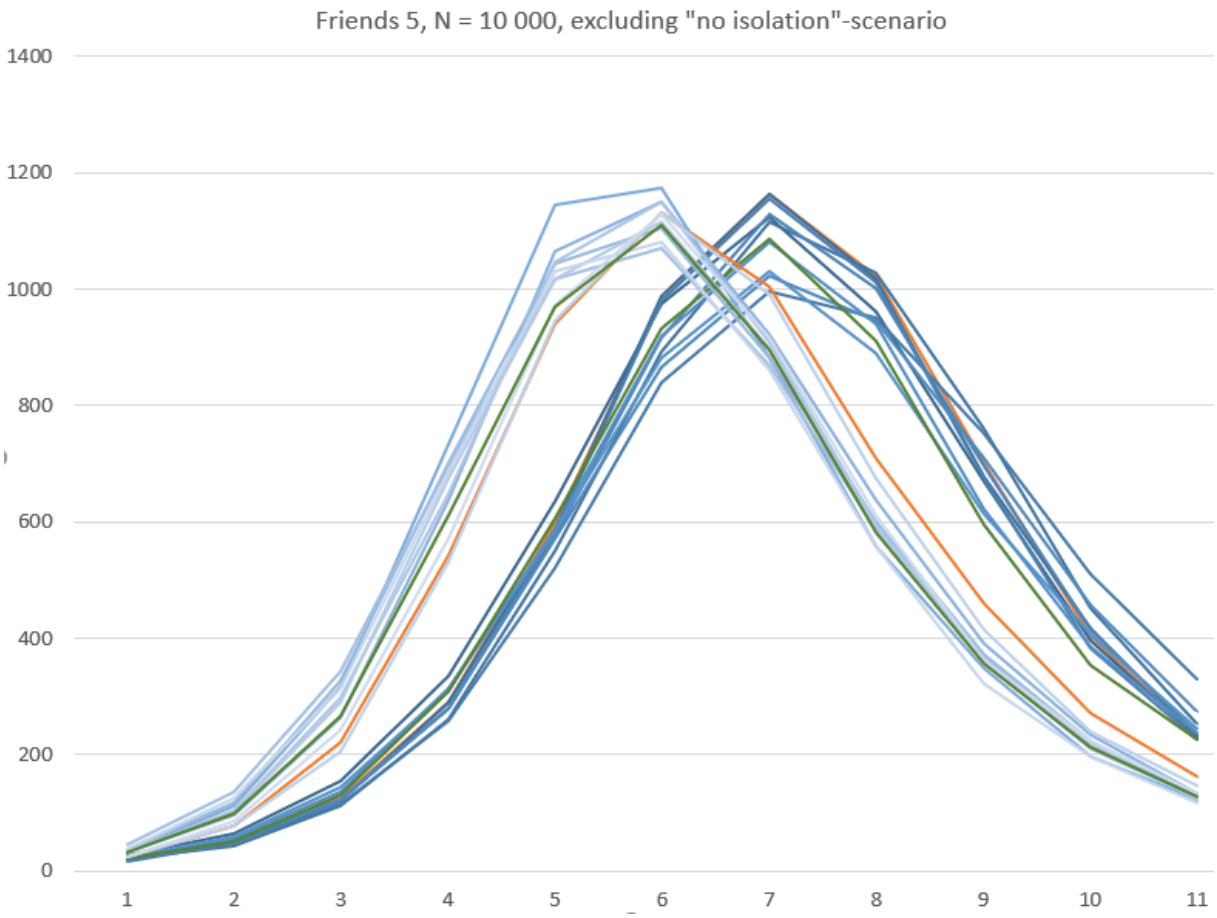


Figure 4.6: Friends 5, Distributed left group, uniform right group

A phenomenon occurs where the simulations with the highest uptake marked in green, no longer has the lowest spike. This is not reflected in infection size after T days where we still see that higher uptake yields lower infection totals. The plots are also more "chaotic" in the sense that they are not as neatly on top of each other as in 4.3. I believe this is mainly due to the cliqueness being affected by increasing the total number of nodes, giving us a more sporadic and random infection than with lower values of N. The daily infection cases are at its highest in days 5-8 in similarity as with 4.3.

I also wanted to showcase the largest number of N I ran for 4.7: Where it is still similar to

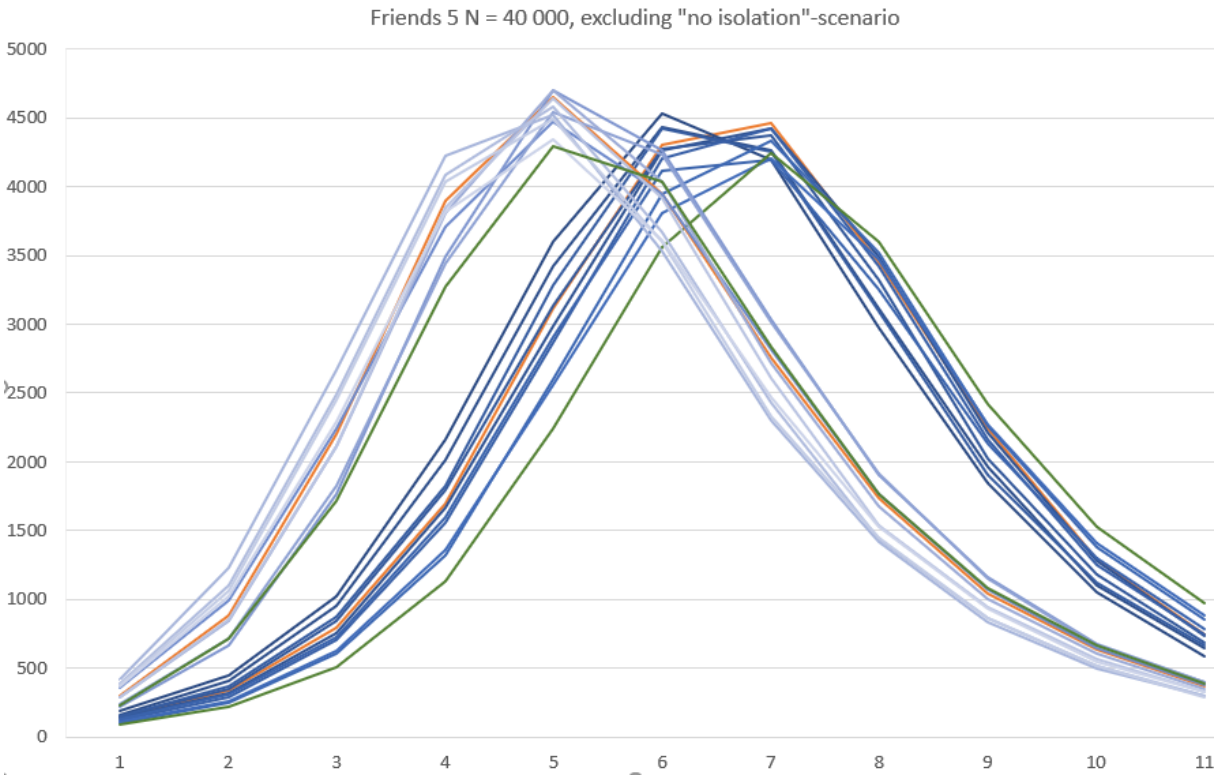


Figure 4.7: Friends 5, Distributed left group, uniform right group

what we saw when N was 10 000, but the infection spikes slightly faster here compared to the $N = 10\ 000$ scenario. This might hint towards increasing N to larger values leads to an even earlier infection spike.

Critique

Firstly; my main issue with the current approach is that the population obtains herd immunity relatively quick when isolation is implemented. I think this is mainly due to the cliqueness of my program. In reality a node would be part of several social cliques, such as; work, family, multiple friend groups, school, and sport, to mention a few. Which means that given a node in my network is at best part of its friendlist-size number of cliques, the possibility of a node being infected significantly decreases as larger portions of the network gets infected.

Secondly; In most social networks, triadic closure is an important property which is not implemented in my program. To repeat, if A-B and B-C exists, A-C is more likely to exists. Which effectively means that in my simulation, cliques should be made with tendencies of friends where they also are friends with each other.

Thirdly; The model does not implement properties such as age and location. Population density would have an affect on how the disease would spread, e.g. urban vs rural. Age would affect probabilities of nodes meeting other nodes. I imagine the number of contacts can be altered to fit some statistical approximation of how many contacts people have per age group. Also probability of people being more likely to meet people their age, or with common interests could be implemented.

Fourth; The value T for iterating over time is in this simulation in days. Which is a intended choice, however if this was implemented in a format of hours or minutes one could do interactions such as probability of infection over time. And delay from notification to actual quarantine. Even though my implementation is implemented in days, activity varying from weekends to weekdays is not implemented either.

Fifth; In real-life epidemics there are cases that go undetected, this is not implemented in my model. Mainly due to attempting to keep the model more simplistic, since estimations of undetected cases is not consistent throughout papers as it is a hard thing to determine in its nature of being undetected. The number of undetected cases can range from 20% [37] as far up to 86%, [38] and the effect it would have if implemented is unknown.

Sixth; Given that T is implemented in days, as mentioned in the fourth point interactions that goes on the hourly basis is not accounted for. However the delay from when a person is infected and should upload their codes to notify other users varies from 0-8 days, [39] which in my model is not accounted for as this happens instantaneously.

Conclusion

To conclude this section we had the following two questions of interest: " 1) The difference in a uniformly distributed population versus a non-uniformly distributed. 2) Different adoption rates of digital contact tracing and their results."

Based on the results I have found through my methods I would say that there could be significant difference between a uniformly distributed population vs a non-uniformly distributed population. Where the main difference between uniform and non-uniform in scenarios where friends=3 and friends=4 I think comes from the aforementioned exclusive cliques. Which to

repeat gives a fake immunity in the scenarios where the occurrence of a node having 0 and 1 friends. And this occurrence is more likely in scenarios where the distribution revolves around a lower mean of 3 and 4, comparatively to when the distribution is revolved around higher values of 5 and above. However in the scenario where friends = 5 we see a notable difference in the uniform compared to the distributed, and the statement of these "fake immunity's" were to be true it would make sense that the difference would have been inverted, that the distributed were lower in infection numbers than the uniform. The difference between the two is then majorly due to the effect DCT has on a uniform distribution compared to the non-uniform.

Different adoption rates shows us what we already expected to see, in that earlier isolation of nodes reduces the infection size of the population. And by increasing the number of users that use a phone, given the assumption that the phones will give out faster notifications than manual contact tracing, we increase the number of nodes that isolate early. We would have liked to have gotten larger variances in efficacy to the relative sizes, which would give more distinct results compared to the ones achieved. With all probabilistic studies there will be deviations and especially with the small changes in efficacy it is hard to tell whether it is probability deviation or actual efficacy change.

Chapter 5

Discussion

Now that we have a baseline understanding of how digital contact tracing works, we can discuss some of the major controversies around it. Given that digital contact tracing was not in use before the SARS-CoV-19 virus, the short development timeframe over the past 18 months leaves the discussion surrounding digital contact tracing up to ambiguity. Most research is based on assumptions given the lacking statistics, or they attempt to generate more data using a simulating approach similar to mine. I am sure this field will just evolve, and many of the issues we are struck by now will hopefully be gone when the next pandemic comes around. In this discussion section we will debate around some of the points of interest regarding digital contact tracing. Firstly we will dive into some statistics published from different countries in an attempt to determine if we get the results we would expect, in a similar fashion as when we observed the results from our program earlier. After that we will take a look at the different critiques that surround digital contact tracing, such as; socio economical issues, trust issues, juridical issues, and perhaps discuss potential future solutions or ideas.

5.1 Statistics

When researchers are trying to determine the efficacy of digital contact tracing confusion might be had with how you define efficacy itself. Some common measures could be the number of contacts identified through DCT, or the reduction in effective reproduction number, or the number of infected individuals. In our model earlier we gauged the efficacy by comparing the number of infected individuals. The majority of solutions globally run, are based on the GAEN framework. As this is tied heavily with that Google and Apple statistics seems to be harder to come by, making it more difficult to gauge the effect of digital contact tracing. The ENX which we saw earlier lets the health authorities aggregate statistics such as:

- the number of exposure notifications sent in a public health authority's region
- the number of user interactions (e.g., taps, dismissals) with exposure notifications in the region
- histograms of the risk scores computed for users of the Exposure Notifications System in the region
- the number of exposure notifications sent in the past 14 days when verification codes are used in the region

- the number of exposure notifications sent in the past 14 days when temporary exposure keys are shared in the region
- histograms of the number of days between having an exposure and receiving a notification

In addition to these I would be interested in; an estimate on number of individuals that get exposure notifications through GAEN that did *not* get one through manual contact tracing. Number of active users per day/week/month. Percentage of users that get a positive diagnosis, but declines uploading their infected TEKs. My main assumption to why the data is not publicly accessible is data privacy. Which again is why this topic in general is "bittersweet", by that I mean that the data we wish to observe to determine if the solution is "good", cannot be seen due to it then being "bad" (less privacy preserving).

Countries that do release their numbers typically have less than ideal percentages of uptake. Researchers are not united on what uptake is needed to ensure an effect, however 60% [40] is a number that is seen floating around referred to in multiple papers. On the contrary, cases are made that low-uptake countries can benefit from DCT as well. This might be hard to imagine due to the numerical fact that the probability of a person engaged with DCT meeting another person engaged in DCT is uptake squared. E.g. in a country with 20% uptake, like Norway, the expected number of cases one could detect is around 4.5%. Is this reflected in the data released?

If we look at the official NIPH statistics for cases detected by the application [41] the past 100 days (26th October to 24th of July). Compare those to contact tracing in general, accounting for both MCT and DCT, [42] showcased in Figure 5.1: showing a histogram over the past 100 days where the y-axis is the number of days that corresponded to the following percentage shown by the x-axis. The expected number based on uptake and cases would 4.5%, however only a single day out of the past 100 reflect this, showcased on the far right in the histogram. Additionally, by the number of cases found in the app, it is not stated which of those, if any at all, was not found through MCT. One caveat here is that the number of detected cases by the app does not nescessarily mean that they are in fact positively infected.

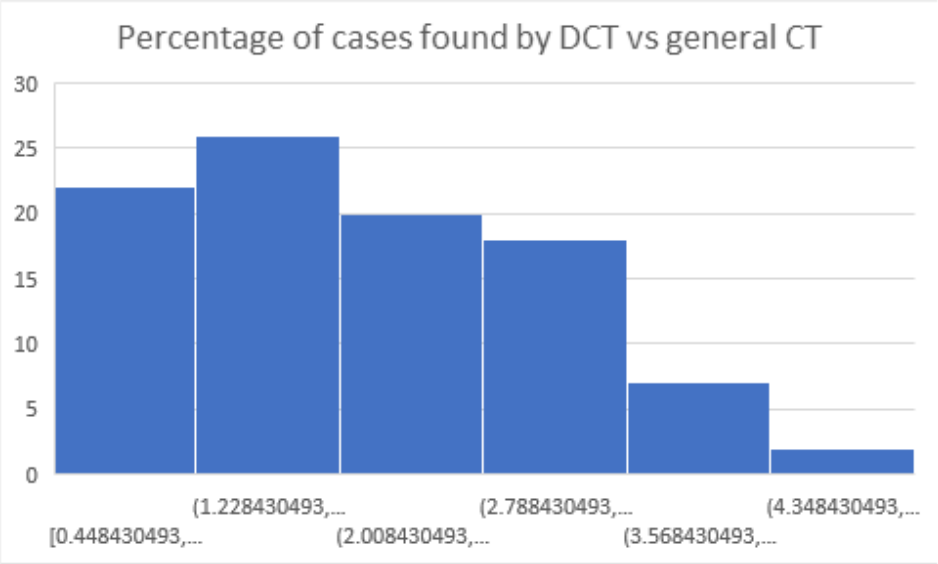


Figure 5.1: FHI stats

An analysis done 8th of July 2021 showcases some statistics and other statements regarding uptake and efficacy in western countries The analysis bases some of its arguments on statistics from this chart 5.2[43]:

		Population size	App name	Launch date	Downloads (% of population)
	Slovenia	2.1 mil	#OstaniZdrav	17 Aug 2020	372,464 (19%, Feb. 21)
	Lithuania	2.8 mil	Korona Stopp LT	06 Nov 2020	300,000 (10%, Feb. 21)
	Estonia	1.3 mil	HOIA	20 Aug 2020	265,093 (20%, Feb. 21)
	Malta	0.5 mil	COVIDAlert	18 Sep 2020	94,215 (19%, Feb. 21)
	Croatia	4.1 mil	Stop COVID-19	27 Jul 2020	83,191 (2%, Feb. 21)
	Hungary	9.8 mil	Virus Radar	13 May 2020	75,000 (0,8%, Sep. 20)
	Bulgaria	7 mil	ViruSafe	07 Apr 2020	63,577 (0.8%, Sep. 20)
	Germany	83 mil	Corona-Warn-App	16 Jun 2020	27 mil (33%, Apr. 21)
	England and Wales	59.1 mil	NHS COVID-19	24 Sep 2020	21.7 mil (36%, Feb. 21)
	France	67.2 mil	StopCovid / TousAntiCovid	02 Jun 2020	13.5 mil (20%, Mar. 21)
	Italy	59.8 mil	Immuni	15 Jun 2020	10.4 mil (17%, Apr. 21)
	Cyprus	0.9 mil	CovTracer / CovTracer-EN	05 Apr 2020 / 01 Feb 2021	8.000 (9%, Feb. 21)
	Spain	46.9 mil	RadarCOVID	21 Aug 2020	7.3 mil (16%, Apr. 21)
	Netherlands	17.3 mil	CoronaMelder	10 Oct 2020	4.5 mil (26%, Feb. 21)
	Portugal	10.3 mil	Stayaway COVID	01 Sep 2020	2.9 mil (25%, Jan. 21)
	Finland	5.5 mil	Koronavilkku	31 Aug 2020	2.5 mil (45%, Feb. 21)
	Ireland	4.9 mil	COVID Tracker	07 Jul 2020	2.4 mil (49%, Feb. 21)
	Belgium	11.5 mil	Coronalert	30 Sep 2020	2.3 mil (20%, Jan. 21)
	Denmark	5.8 mil	Smittestop	18 Jun 2020	2.2 mil (38%, Feb. 21)
	Czech Republic	10.6 mil	eRouška	20 Apr 2020	1.5 mil (14%, Feb. 21)
	Poland	38 mil	STOP COVID - ProteGo Safe	09 Jun 2020	1.5 mil (4%, Nov. 20)
	Austria	8.9 mil	Stopp Corona	25 Mar 2020	1.4 mil (16%, Feb. 21)

Figure 5.2: EU Uptake

Of importance here is the percentage under the "Downloads" section on the far right side. Which gives us an upper bound of the active users. We can assume that the actual number of active users is substantially less. Together with this chart of US state uptake shown in figure 5.3[44]:

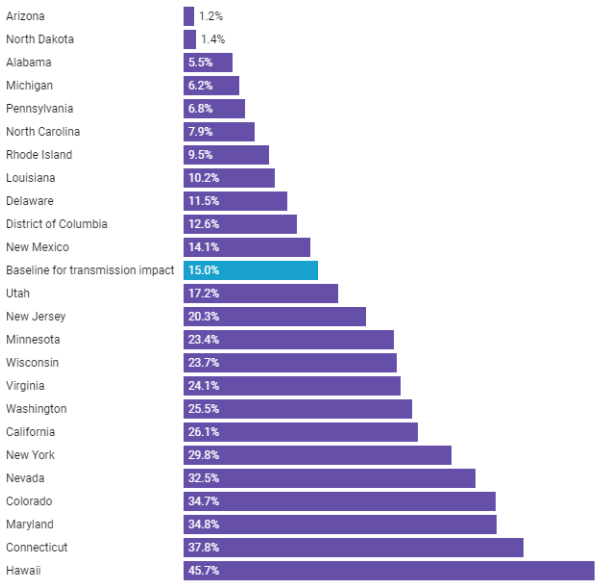


Figure 5.3: US Uptake July 2021

We can see that the uptake percentages of western countries lie in the range from 1 and 40% with a few exceptions that are above 40% (Finland, Ireland and Hawaii). Which is far below the speculated threshold of 60%. The European uptake statistic time of retrieval is noted in 5.2 and the US state uptake is from July 2021. Thus we can imagine the uptake could only be lower the closer we get to the time of outbreak (March 2020). However as stated before lower uptakes can still have an effect to the extent where it can notify people that would not get notified by MCT, thus reducing the number of total infected.

Here are some individual cases:

Spain has released a website [45] for their application, where we note the 19% uptake, and observe for October that the number of codes sent last week (week 42) was 105, while the [46] shows that the total number of positive cases that week was a little over 13000. Also stated on the application website is the ratio of sent codes to confirmed cases, which is around 1.5%, meaning that 98.5% of quarantined cases were not in fact infected.

In Australia the COVIDSafe reported 561 cases by the app, whereas 544 were also identified by MCT effectively giving 17 new cases over a 6 month period[47]. The COVIDSafe project is estimated to have a monthly upkeep cost of between 60 000$ and 70 000$.

The app Immuni in Italy peaked at 197 on a day where MCT found some 34 000 cases[47].

Covid Alert in Canada through April 2021 had 35 000 app notifications, which resulted in 400 confirmed cases found, in the same time 165 000 cases was found through MCT[48].

In Switzerland the app has about 1.6 million active users per day [39] and with a population of almost 8.7million [49] that gives us a app uptake of around 18%. During the first five days of November, the app had 73, 122, 111, 136 and 137 covid codes entered into the app respectively. On the same days the number of total covid cases was 2326, 2966, 2820, 2840, and 2691 [50] respectively. There is no information to be found on whether or not a entered covid code leads to the finding of a case or not. But if we assume that every code leads to one case, which is very unlikely, but nonetheless; the percentages for the same five days would be 3.1%, 4.1%, 3.9%, 4.7%, and 5.1% at best. Additionally a survey was done in Switzerland asking people at the test stations the reason to why they went and got tested. Of 7842 people 41 were there through DCT notifications[48].

Even though Ireland has a high uptake rate compared to the other countries mentioned, the expected number of uploads has only been in the range of 15-25% of what it should have been [48], implying that high uptake rates does not nescessarily mean high efficacy.

As we can see the current DCT situation in the countries we looked at above has shown varying results. Obviously these results does not nescessarily conclude if something works or not, or if it is gonna work or not in the future, but it does give us some insight in the situation so far. Different governments seems to be reluctant to release statistics on the app they are using, this is the case for most or not all of the European countries that were not mentioned in this chapter, and the US as a whole. Following figure 2.D from this report [17] made by the Norwegian research group Simula metropolitan, we can compare the theoretical efficacy given the app uptakes with the actual results. If we draw some comparisons using this function as a baseline the results looked at previously shows that e.g. in Switzerland the app gives results as we would expect with a 20% uptake and 4% detection, while other countries such as Italy, Canada and Spain show very low numbers in detected based on what should be expected given their uptake. Which was also what was stated for Ireland in [48]. One thing to keep in mind with these statistics in general is that the number of *actual* infected users is not nescessarily the same as the number released by the governments. But we cannot know the presumed number, and will look more on the factual data released by the different governments. Generally this number of "invisible nodes" or "true infected", which are synonyms for the same phenomenon, implies that there are undetected cases which does not get discovered through contact tracing in general (both manual and digital).

5.1.1 Uptake and efficacy

When debating the "efficacy" of DCT in different countries, it is vital to understand the difference in what the different numbers really represent. Even though countries might have similar uptakes, the ratio between rural and urban users might differ from one country to another. Population density among users living in these urban areas will also differ from one country to another, as this is already an issue when simulating data on a nation-level with different provinces having different density levels. Usually uptakes is only a percentage representing the fraction of the total population that has downloaded the application. Some problems with this is that it does not nescessarily mean that a user that has downloaded the app, is an active user of the solution. Another problem is that individuals are not the same. By that I mean that given a population size and saying that 20% of them are partaking in DCT says nothing of who actually is using it in the population. If only established middle-class persons use the app, and still make out to be 20% of the population, that does not nescessarily give a "good" DCT efficacy compared to a distribution where the users of the apps are more spread based on age, religion and economic background. This is a problem we will discuss in more detail later.

As mentioned regarding the US situation lacking a standard of what information to release, I also discovered the same issue trying to find information regarding statistics on the different solutions. Given the aforementioned problem with the lacking definition of "efficacy" in DCT, it does not seem to be a unified answer as to what information should be collected and displayed to represent a "good" solution. There might be a underlying reason as to why different governments publish different properties from their app, I am speculating on that the efficacy is too low, and thus having full transparency in terms of statistics might lead to distrust, and then even less efficacy than before. Some characteristics would be unattainable based on the nature of decentralized digital contact tracing, such as identifying if a case was found through DCT or MCT first, if the index case does not answer that question themselves. But regardless I am sure that governments can only improve on releasing good statistics based on the current situation. It seems to me to be obvious that statistics regarding DCT is not particularly in focus, compared with numbers regarding MCT and death rates. The norm regarding DCT statistics currently seems to be to only release numbers such as; "daily and total number of downloads, daily and total number of app notifications." In my mind daily active users is of greater interest than daily downloads when trying to deduce the efficacy of the solutions. Giving us a better grasp of actual uptake percentages. Also the rate of app notifications that actually led to a confirmed case. We know that false positives and false negatives is an issue in DCT due to the inaccuracies in bluetooth, but having real numbers on the fraction of correct quarantines would reflect the reality on this issue and could be interesting.

On the topic of uptake being a difficult property to describe properly. I have an assumption that users that do download the application and partake in a DCT scheme, would be less likely to put themselves in contagious situations, than people who generally does not care about DCT or the pandemics at all. I have not found any studies to confirm or deny this, but I think it is a logical assumption to make. How it affects DCT is hard to gauge, but will probably have a small negative effect on the efficacy of DCT.

Human behaviour in general is a critical aspect of contact tracing, both manual and digital. A quote: "The drawback is that individuals may gain a false sense of security at a low uptake level if they do not receive exposure notifications, which may encourage a more relaxed behaviour" [51] from an analysis speaking about low uptakes, this quote is referring to Spain in particular, which we also have seen. When given notifications to quarantine, whether or not this is through

manual or digital, the adherence to quarantine differs from individuals. [36] How much or little individuals quarantine might also differ from one country to another, one factor to this could be based on the individuals trust to the government.

The point I am trying to make here is that there are so many factors in how well a digital contact tracing scheme could work in any given country. And therefore the solutions should be tailored towards fitting the country they are in based on movement statistics and population density and so on. Comparisons between countries that does not have similar demographics has to be taken with large grains of salt. That does not nescessarily mean that they cannot give us a tendency or vague understanding of the situation though.

Based on the previous stated fact that the probability of two individuals partaking in DCT meets each other is uptake squared. And in unison with the tendencies we have seen from our model program, which then again is in similar to the results of other studies. We can with high certainty say that higher uptake nets to better "efficacy". However, we have also seen that lower uptakes does have an effect although smaller than if the uptake would have been higher. How to achieve high uptake is a ongoing problem, governments are continuously trying to overcome this problem with different marketing strategies, and generally advising the population to partake in DCT. So why is it that the uptakes are so low?

Even though digital contact tracing solutions so far has not given the results we hoped for. They are still doing something. The few cases that are in fact detected through DCT would stop those chains of infection early, how many less cases we get after the second and third step in those chains is a hard number to estimate. It is this number we are weighing up to the cost, both economic and the privacy loss cost.

5.2 Purpose

5.2.1 User Purpose

One reason for the low uptake in these applications might be that the users lack incentive to download and use the application. We have seen that solidarity in itself is not a big enough push for everyone. Many US states are giving out money or marijuana as an incentive to get vaccinated [52]. In the earlier days when the GAEN framework was not yet implemented, users were advised to go around their day with the application in the foreground [19]. I do not believe that the majority of users were following this advice on the daily. Other reasons I can imagine for users not using the solutions; either being scared or sceptic to technology in general, mistrust in the government, or scepticism towards pandemics. It has to be noted that lack of cooperation is an issue for contact tracing as a whole, and not just for digital contact tracing. As with both cases human behaviour and social issues is the foundation contact tracing builds upon. Some of the scepticism might have been brought up early on by the debate surrounding centralized and decentralized solutions. With the signed letter on critiquing centralized solutions [53] basically acting as a "cease and desist" for the PEPP-PT organization being in focus here. Not too long after the PEPP-PT disbanded the first Norwegian app "Smittestopp" was taken down by the Norwegian cyber-rights department. This is one of the first and only regressions of digital contact tracing apps so far.

5.2.2 Developer purpose

The reason why Smittestopp was taken down was majorly due to low uptake, and the cyber-rights department deemed the privacy cost higher than the effect the application gave. Smittestopp was a custom solution, and with a higher privacy cost than other protocols discussed in this thesis. Smittestopp tried to both do what other DCT solutions do; reduce the number of infections in the population. But at the same time they wanted to collect data for epidemiologists to do research and get better understanding regarding the pandemic. What purpose do we want to achieve? Was a question that rose amongst the protocol developers. Every protocol now has its own section stating the specific purposes of the protocols, and it is amongst these purposes that one should weigh the privacy cost towards. I think this quote from the DP-3T documentation[20] fits well to represent the thought process of the developers: "We strongly believe, however, that it is not the time to conflate novel, untested technologies with the understandable desire to collect new epidemiological insights".

5.3 Socio-economic problems

Another group of issues is tied under what we will refer to as "Socio-economic" problems. Susan Landau[54] brings up many perspectives on the socio-economic problems.

5.3.1 Access to smartphones

First among these is the underlying property that digital contact tracing is ran on a handheld device, which then requires an individual to own and use said device. And the probability of owning a smartphone relies on factor such as your income and culture which can result in inequity when determining the usefulness for digital contact tracing on a population as a whole. We know that a population is broken down into different demographics, and some of these minorities might have less gain than others. A fair assumption to make is that a middle-class individual has a higher chance of owning a smartphone compared to a low-income or no-income individual. Which is in my mind logical, but also already a tendency seen: "A study of the UK app showing decreased incidence of SARS-CoV-2 spread noted that regions of higher app use had "lower levels of poverty, are more rural, and have higher local GDP" (2). In those regions, the populations that use the apps are more likely to be white, but also more likely to be elderly." [54].

5.3.2 Isolation-inequity

On the topic of income Landau [54] brings up the point of isolation not being equally accessible for someone that has a job which requires physical attendance, compared to someone that can do their work from home. This is an interesting thought, but is similarly a problem in manual contact tracing as in digital contact tracing. In terms of digital contact tracing this isolation-inequity suggests that the applications should have a high rate of quarantine cases that were in fact positive. This goes back on the topic earlier of distance-policy making, as to which cases to trace or ignore. If the applications spews out notifications for its users to quarantine with a low rate of "actually infected", the economic sacrifices made might be more of a societal hindrance than if left ignored. And individuals that belong to more economically-exposed minorities cannot

afford to isolate and perhaps lose out on income compared to someone that can do their work from home.

5.3.3 Mistrust in government

The mistrust in government that was mentioned previously is also hugely dependent on which demographic the individuals belong to. "For example, a history of mistreatment has created great distrust of both the government and public health by many in the US Black community." [54]. Probably making them less likely to partake in DCT. The same idea can probably be made for other countries where there are smaller tendencies to racism, or other similar discrimination towards minorities.

5.3.4 Contact tracers

Decentralized digital contact tracing solutions have become the solution the majority of countries have adopted. This has as we have seen several pros and cons. One of the cons not mentioned yet is that these "contact tracing solutions" are actually just exposure notification solutions. The difference being that a contact tracer is usually a health care official that follows up on infected cases. This is not possible in a "DCT" or exposure notification solution, as the health care official is subtracted from the equation. The nature of the decentralized solutions is that the health care should not be able to identify you, which then results in you not being able to be contacted if you need help. To combat this issue some solutions have implemented a feature where you can register a phone number to be contacted on, but this obviously increases the privacy cost, and removes some of the principles behind a decentralized solution. Users could also reach out themselves and ask for help, but some individuals find this harder than others.

5.4 Cost of using DCT

This "cost" that has been mentioned previously. What is it really? I think of the cost as separated in two parts, economic and privacy.

5.4.1 Economic

The economic numerical value of developing and maintaining a contact tracing solution varies by country and solution. We have seen previously that the Canadian solution costs about 60-70 thousand Canadian dollars a month to maintain. This article[55] states that in America the "States each spent hundreds of thousands of dollars to develop these contact tracing apps...", with some cited values with 700 000 USD for the New York solution, and 229 000 USD for the one in Virginia. The Norwegian app also has an official price tag of 14.4 million Norwegian kroner, or 1.645 million USD[53]. A caveat here is that manual contact tracing is *not* free either, with a underlying cost of employing contact tracers and the cost in waged hour whereas DCT would be autonomous.

5.4.2 Privacy

The privacy cost is as mentioned a hard value to determine. I believe that companies even though they seem transparent are not 100% honest. They earn money by selling data and we have seen with Snowden and such that even the "most trusted" authorities can do foul play. If we mainly have the GAEN solution in focus, as this is the by far most adopted solution out there. The data gained by Google and Apple through the digital contact tracing is probably not that much in my opinion. I presume that people in general are not willing to go out of their way to keep their data safe by e.g. not using google play services, or doing actions like shutting their phones off when they are out and about. Most people use their smartphones when they are out doing their daily activities, and through that use I already assume that Google/Apple already collects most of what they would gain through DCT. Through DCT more people are exposed to having their data collected, however I also think that the more pessimistic users that go out of their way to preserve their privacy would not partake in a DCT system.

In terms of the health authorities, they already have a lot of data on us, and given that you need to get a confirmed lab test before you can upload your infected seeds/TEKs the health authorities have little to gain through the DCT schemes. Implying that you need to go through health authorities to partake in DCT and cannot go the other way around. The most concerned privacy cost around DCT is the deduction of your social graph. Which in short means that someone can more easily with DCT than without, determine who you are surrounding yourself with. This can in turn expose potential business partners, cheating spouses and in general just an intrusion of privacy. This "deduction of social graph" is a issue stated by all the protocols we have seen in this thesis, as there is no real solution to how one can fix this. Some properties the protocols hold to try and make this harder to do, is that your temporary contact tracing identifier must have the same expiration date as the bluetooth MAC address, so these cannot be linked. Serge Vaudenay[56] suggests that even this is not a fix to the issue, where you can observe the disappearance of the identifier and MAC with the emergence of the new ones. In practice this would mean that you follow someone for 20 minutes or longer, (or are somehow able to trace them in some other way with bluetooth observers left around a large area or something) which to me is a rather dumb point because if you go to those lengths you can always just stalk someone to intrude their privacy.

When the manual contact tracers call you it is this "deduction of social graph" that they really want to know. It is the only way to determine your close contacts and find places of potential exposure. Cooperation is an issue here, where many individuals does not even answer the phone from the contact tracers. In an era of spam-callers and "stranger danger" picking up calls from unknown numbers are less likely than before. Many American states and county's have made campaigns to increase the rate of answered calls as it hinders the effectiveness of MCT if you do not pick up the phone. If the call happens though, it is vital for the manual contact tracers that the index case can recollect as many of the contacts as they can. It is this memory gap where DCT stands to gain over a manual solution. But DCT would most certainly hand out significantly more notifications that were not real infections. With this in mind the current solution that most governments rely on is a mix of the two, where the manual contact tracer consults with the index case and compares the information they get there with the DCT data to determine potential quarantines.

With GAEN, the user needs to consent to upload their infected TEKs when diagnosed. If they say no, the effectiveness of the solution is hindered. The more individuals that do not consent to the upload, the larger the pool of "invisible nodes" there are in the network. We

already know that individuals usually get asymptomatic, but are still infectious, which makes contact tracing as a whole hard. These "no-consent" users would contribute to the size of these undetected invisible nodes.

5.4.3 Price of life

In terms of digital contact tracing these costs can be seen as the investment, and the lives saved as the return. Now we already know that the amount of actual lives saved is not a easy number to estimate, and we know that the economic drawbacks of quarantining and isolating might be just as big or bigger than the damage caused by the disease. Nonetheless in an attempt to answer the question of how much efficacy does one need to justify the cost we need to put a value on life. There is a criteria called the "cost-effectiveness threshold" [57] which governments usually use to determine if a costly project is worth investing in, in terms of how many "quality-adjusted life-years" it affects in the population. Where these "QALY" (quality-adjusted life-years)[58][57] refers to what number you are willing to pay for a year of perfect health. Obviously this is a subjective subject, where each has their own opinion, although there are some critical points we can draw from this. If we take the Norwegian application as an example. The cost was around 1.675million USD, and the number of total notifications to date is around 6000, where there is no statement towards these notifications actually leading to new discoveries or not. By other countries the estimation of a code to a new discovery is 5% and less. Which gives us a *very* optimistic 300 new cases for the cost of 1.675 million USD. The optimistic 300 new cases is an upper bound of the possible individuals found through DCT alone, this number is probably significantly lower. Additionally in the scenario where these cases were to go undetected, it is hard to compute the additional cost these chains of infections would add. The size of the chains would be a function of the reproduction number. Now, had we used the same amount on manual contact tracing to hire more contact tracers, or invest in upgrading the current systems or other health care would the effective "QALY" number be bigger than with DCT?

With the ENX, the adoption of new solutions are believed to be easier and probably cheaper than before, so we have some redeeming qualities in all of this. Additionally we have to keep in mind that the decision to adopt or use DCT solutions needs to be kept with a futuristic point of view. To expect a high efficacy straight out of the gate is unreasonable, when we know that uptake is such an important factor. We have seen that the uptake will not sky rocket from zero to the wanted 60% in a matter of weeks or months, but that does not mean that we should not have tried, or that we should stop trying either.

5.5 Politics

5.5.1 Interoperability gateway EOC

Interoperability in terms of contact tracing is the working foundation of back-end to back-end communication between countries. After the early development of the protocols we have looked at during spring 2020, this interoperability update came rolling for most of the working protocol/solutions around autumn or winter 2020. One of the major constraints for interoperability to work was, and still is, that the country you are trying to cooperate with needs the same contact tracing solution as you. However, when most countries now have adopted or switched to a GAEN based solution, this interoperability feature could be used more efficiently. The

European union agreed to this thought, and made what they refer to as the "Federation Gateway Service"[59] which is a working architecture of how countries should modify their backends to allow for communication across borders. Effectively letting any GAEN user have a working contact tracing solution given that they are in a country also having a nation wide adopted GAEN solution.

On the 15th of May 2019[60], almost a full year before the outbreak of covid, the president of the US Donald. J. Trump put an executive order which "bans the use of telecommunications equipment from foreign firms deemed a national security risk". [60] The reasoning in short was a worry that foreign countries are spying on US citizens[61]. In terms of digital contact tracing newer Huawei phones that do not have access to the google play store are not eligible for interoperability. China has made "Contact Shield" [1] as a solution for their citizens. However chinese travelers using contact shield would not be able to have working solutions in European countries or US states.

It is an interesting consequence that what effectively started as a trade ban has secondary effects to contact tracing interoperability. It is also interesting that the US government is so hypocritical about surveillance and spying when we know some of what they have been doing on that front after 9/11[62]. Again, how much data is *really* being collected on foreign users by companies such as Google and Apple is a hard number to estimate. But I would not be surprised that a similar whistleblower case could happen in later years regarding all the data being processed by contact tracer providers.

5.6 Conclusion

One of the main goals of this thesis was to determine if digital tracing solutions is worth using compared to the cost. We have now seen that the privacy cost is subject to change on a country level, and additionally on an individual level. Digital contact tracing cannot solve societal issues such as oppression, economic inequalities, racism and such. However, in terms of economical cost an application is very cost-efficient compared to the cost of e.g. hospitals and other health care facilities where hourly wages, electricity, equipment and such outweigh the economic cost of the application maintenance and development. The privacy cost is only "paid" by users partaking in DCT, where they are required to give consent. This freedom of choice leads to questions as to why one should use these solutions, and as a result the current situation globally is that the uptake is lower than expected. This lower uptake has yielded questionable results at best, so far. But given that *most* of the economic cost is paid during development, which now is done, and the privacy cost would be a function of use, which is currently rising. The cost-efficiency of the solutions should be increasingly improved by the years to come.

The questions we started this thesis with: "What % of the population would need to actively use a digital solution for it to have a reasonable affect?" and "What amount of benefit is required to justify the privacy intrusion it surmises?" goes hand in hand. They are prone to subjective opinions, and the answer would differ based on the scope used. Currently, with the low uptakes I do think that countries in the 10-30% uptake range yield incredibly low results with the actual new cases found through DCT. Where as we saw around 1-5% of cases a day is reflected in the number of notifications sent by the app, and then again the ratio of actually infected to notification is around 1-5% also. Immediately I do not have any good fix as to how one should improve the uptakes significantly enough from a 10-30% level, to a 50-80% level, that in my opinion would be a expected level in 5-10 years to come. Personally as a Norwegian citizen, I

do think that advertisement and marketing is probably a field that the Norwegian government should work on, and I believe that could be the case in other countries as well.

If you share my futuristic mindset on the matter, I think that the policies, methods and uptake could only improve, and should maneuver this technology in to a place where it is of significant use to combat pandemics in the future. With increased vaccination rates and development, together with herd immunity it is inevitable Covid-19 will die out. When this happens the interest and need regarding development and improvement in the DCT field is surely to decrease, this is worrisome as this is usually the tendency with pandemic research where the economical backing is decreased when the need is not there anymore.

5.6.1 Future Work

For future work along the lines of this thesis, I think the main focus should be on improving the model program.

Implement the features presented in the Model-program critique section, and perhaps additionally try adopting other simulation tools as a comparison. Other improvements that could be done to the program:

- Generally improve current parameters to better reflect real-world scenarios.
- Improve the logic of how and why people meet, based on factors such as age and geographical location, but also perhaps common. interests/hobbies and such.
- Other parameters that would make sense to implement on a node-level: Race, religion, age, location, interest, extroverted/introverted, work, school.
- Logic on providing nodes with phones should be based on some algorithm based on age and economical statistics.
- Probability of partaking in DCT dependant on triadic closure, e.g if your friends use DCT and talk to you about it that way you would be more likely to be persuaded to also use DCT.
- Attempt to test impact of invisible nodes to account for the no-consent GAEN users, and invisible nodes in general.
- Dynamic contagion probabilities depending on incubation time and if asymptomatic.
- Probability of movement dependant on sickness.
- Implement indoors and outdoors

Many of the topics discussed in this thesis is prone to change, given the modernity and actuality of the topics. Especially the technological aspects, where future digital contact tracing solutions might circumvent many of the issues presented or discussed in this paper. Advancements in the device-to-device field would have significant impacts on the DCT solutions, with either new and improved bluetooth specifications, or new technology in general.

Bibliography

[1] *Document*. URL: https://developer.huawei.com/consumer/en/doc/development/HMS-Plugin-Guides-V1/faq-0000001072130397-V1 (visited on 11/17/2021).

[2] *Glossary of Terms for Infectious Disease Modelling: A Proposal for Consistent Language*. canadian. Apr. 29, 2020. URL: https://nccid.ca/publications/glossary-terms-infectious-disease-modelling-proposal-consistent-language/ (visited on 11/17/2021).

[3] *Smittesporing*. Feb. 2020. URL: https://www.fhi.no/nettpub/coronavirus/testing-og-oppfolging-av-smittede/smittesporing/ (visited on 11/17/2021).

[4] Hinta Meijerink et al. "The first GAEN-based COVID-19 contact tracing app in Norway identifies 80% of close contacts in "real life" scenarios." In: *medRxiv* (2021). DOI: 10.1101/2021.05.06.21253948. eprint: https://www.medrxiv.org/content/early/2021/05/07/2021.05.06.21253948.full.pdf. URL: https://www.medrxiv.org/content/early/2021/05/07/2021.05.06.21253948.

[5] Matt J Keeling, T Deirdre Hollingsworth, and Jonathan M Read. "Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19)." In: *Journal of Epidemiology & Community Health* 74.10 (2020), pp. 861–866. ISSN: 0143-005X. DOI: 10.1136/jech-2020-214051. eprint: https://jech.bmj.com/content/74/10/861.full.pdf. URL: https://jech.bmj.com/content/74/10/861.

[6] Sebastian Contreras. *The challenges of containing SARS-CoV-2 via...* Jan. 2021. URL: https://www.nature.com/articles/s41467-020-20699-8.

[7] L.S Small et al. *Summary of Bluetooth Contact Tracing Options*. Tech. rep. May 2020. URL: https://www.dta.mil.nz/assets/Publications/Bluetooth-Contact-Tracing-Options.pdf (visited on 11/17/2021).

[8] B.R. Rowe et al. "Simple quantitative assessment of the outdoor versus indoor airborne transmission of viruses and COVID-19." In: *Environmental Research* 198 (2021), p. 111189. DOI: 10.1016/j.envres.2021.111189.

[9] G. Cencetti. *Digital proximity tracing on empirical contact...* Mar. 2021. URL: https://www.nature.com/articles/s41467-021-21809-w?error=cookies_not_supported&code=23132e9e-65fd-4b9e-af05-2bffddd22da4 (visited on 11/17/2021).

[10] Shamshul Bahri. "Enhancing quality of data through automated SARS contact tracing method using RFID technology." In: *International Journal of Networking and Virtual Organisations* 4.2 (2007), p. 145. DOI: 10.1504/ijnvo.2007.013540.

[11] Tang See Kit. *Singapore launches TraceTogether mobile app to boost COVID-19 contact tracing efforts*. Mar. 2020. URL: https://www.channelnewsasia.com/singapore/covid19-trace-together-mobile-app-contact-tracing-coronavirus-773571 (visited on 11/17/2021).

[12] Surabhi Agarwal. *Aarogya Setu app launched by GoI in public-private partnership mode*. Oct. 2020. URL: https://economictimes.indiatimes.com/tech/information-tech/aarogya-setu-app-launched-by-goi-in-public-private-partnership-mode/articleshow/78916970.cms (visited on 11/17/2021).

[13] Coco Feng. *China launches coronavirus 'close contact detector' in effort to reassure public over health risks*. Feb. 2020. URL: https://www.scmp.com/tech/apps-social/article/3050054/china-launches-coronavirus-close-contact-detector-effort-reassure (visited on 11/17/2021).

[14] Alessandro Blasimme. *Digital Contact Tracing Against COVID-19 in Europe: Current Features and Ongoing Developments*. June 2021. URL: https://www.frontiersin.org/articles/10.3389/fdgth.2021.660823/full (visited on 11/17/2021).

[15] Fan Liang. "COVID-19 and Health Code: How Digital Platforms Tackle the Pandemic in China." In: *Social Media + Society* 6.3 (2020), p. 205630512094765. DOI: 10.1177/2056305120947657.

[16] Wikipedia contributors. *Pan-European Privacy-Preserving Proximity Tracing*. May 2021. URL: https://en.wikipedia.org/wiki/Pan-European_Privacy-Preserving_Proximity_Tracing (visited on 11/17/2021).

[17] Ahmed Elmokashfi et al. *Nationwide rollout reveals efficacy of epidemic...* Oct. 2021. URL: https://www.nature.com/articles/s41467-021-26144-8.

[18] TCNCoalition. *GitHub - TCNCoalition/TCN: Specification and reference implementation of the TCN Protocol for decentralized, privacy-preserving contact tracing.* Aug. 2020. URL: https://github.com/TCNCoalition/TCN#tcn-sharing-with-bluetooth-low-energy (visited on 11/17/2021).

[19] Government Technology Agency et al. *BlueTrace: A privacy-preserving protocol for community-driven contact tracing across borders*. Tech. rep. Apr. 2020. URL: https://bluetrace.io/static/bluetrace_whitepaper-938063656596c104632def383eb33b3c.pdf (visited on 11/17/2021).

[20] DP-3T and C.T Troncoso. *Decentralized Privacy-Preserving Proximity Tracing*. Tech. rep. May 2020. URL: https://github.com/DP-3T/documents/blob/master/DP3T%20White%20Paper.pdf (visited on 11/17/2021).

[21] B.F Fan et al. *Cuckoo Filter: Practically Better Than Bloom*. Tech. rep. Dec. 2014. DOI: 10.1145/2674005.2674994. URL: https://dl.acm.org/doi/pdf/10.1145/2674005.2674994.

[22] TCNCoalition. *GitHub - TCNCoalition/TCN: Specification and reference implementation of the TCN Protocol for decentralized, privacy-preserving contact tracing.* Aug. 2020. URL: https://github.com/TCNCoalition/TCN (visited on 11/17/2021).

[23] TCNCoalition. *GitHub - TCNCoalition/TCN: Specification and reference implementation of the TCN Protocol for decentralized, privacy-preserving contact tracing.* Aug. 2020. URL: https://github.com/TCNCoalition/TCN#a-strawman-protocol (visited on 11/17/2021).

[24] *OpenTrace*. Apr. 2020. URL: https://github.com/opentrace-community (visited on 11/17/2021).

[25] Apple and Google. *Privacy-Preserving Contact Tracing - Apple and Google*. URL: https://covid19.apple.com/contacttracing (visited on 11/17/2021).

[26] By Leo Kelion. *UK virus-tracing app switches to Apple-Google model*. June 2020. URL: https://www.bbc.com/news/technology-53095336 (visited on 11/17/2021).

[27] Google. *Check if an Exposure Notifications app is available in your area - Android Help*. URL: https://support.google.com/android/answer/10289696 (visited on 11/17/2021).

[28] Apple and Google. *Exposure Notification Bluetooth® Specification*. Tech. rep. Apr. 2020. URL: https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ExposureNotification-BluetoothSpecificationv1.2.pdf?1 (visited on 11/17/2021).

[29] Apple and Google. *Exposure Notification Cryptography Specification*. Tech. rep. Apr. 2020. URL: https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ExposureNotification-CryptographySpecificationv1.2.pdf (visited on 11/17/2021).

[30] Nina Wu. *COVID-19 app that warns of Hawaii exposure now available for older iPhones*. Feb. 2021. URL: https://www.staradvertiser.com/2021/02/15/breaking-news/covid-19-app-that-warns-of-hawaii-exposure-now-available-for-older-iphones/ (visited on 11/17/2021).

[31] Google and Apple. *Exposure Notification Privacy-preserving Analytics (ENPA) White Paper*. Tech. rep. Apr. 2021. URL: https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ENPA_White_Paper.pdf (visited on 11/17/2021).

[32] D.J.L Leith, S.F Farrell, and School of Computer Science & Statistics, Trinity College Dublin. *Contact Tracing App Privacy: What Data Is Shared By Europe's GAEN Contact Tracing Apps*. Tech. rep. July 2020. URL: https://www.scss.tcd.ie/Doug.Leith/pubs/contact_tracing_app_traffic.pdf.

[33] Birgitte Freiesleben Blasio. *Coronavirus modelling*. Apr. 2020. URL: https://www.fhi.no/en/id/infectious-diseases/coronavirus/coronavirus-modelling-at-the-niph-fhi/ (visited on 11/17/2021).

[34] E. Bonabeau. "Agent-based modeling: Methods and techniques for simulating human systems." In: *Proceedings of the National Academy of Sciences* 99.Supplement 3 (2002), pp. 7280–7287. DOI: 10.1073/pnas.082080899.

[35] *List of COVID-19 simulation models*. Oct. 2021. URL: https://en.wikipedia.org/wiki/List_of_COVID-19_simulation_models (visited on 11/17/2021).

[36] Chris Wymant. *Contact-tracing app curbed the spread of COVID in England and Wales*. May 2021. URL: https://www.nature.com/articles/s41586-021-03606-z (visited on 11/17/2021).

[37] Billy Gardner. *Contact tracing efficiency, transmission heterogeneity, and accelerating COVID-19 epidemics*. June 2021. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009122.

[38] C. Rippinger. *Evaluation of undetected cases during the COVID-19 epidemic in Austria*. Jan. 2021. URL: https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-020-05737-6.

[39] Federal Statistical Office Experimental Statistics. *SwissCovid App Monitoring*. URL: https://www.experimental.bfs.admin.ch/expstat/en/home/innovative-methods/swisscovid-app-monitoring.html (visited on 11/17/2021).

[40] fraser group. *Digital contact tracing can slow or even stop coronavirus transmission and ease us out of lockdown*. Apr. 2020. URL: https://www.research.ox.ac.uk/article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown.

[41] FHI. *Nøkkeltall fra Smittestopp*. Dec. 2020. URL: https://www.fhi.no/om/smittestopp/nokkeltall-fra-smittestopp/ (visited on 11/17/2021).

[42] FHI. *Daily report and statistics about coronavirus and COVID-19*. Mar. 2020. URL: https://www.fhi.no/en/id/infectious-diseases/coronavirus/daily-reports/daily-reports-COVID19/ (visited on 11/17/2021).

[43] LibertiesEU. *COVID-19 Contact Tracing Apps in the EU*. URL: https://www.liberties.eu/en/stories/trackerhub1-mainpage/43437 (visited on 11/17/2021).

[44] Betsy Ladyzhets. *We investigated whether digital contact tracing actually worked in the US*. June 2021. URL: https://www.technologyreview.com/2021/06/16/1026255/us-digital-contact-tracing-exposure-notification-analysis/.

[45] *App RadarCOVID*. URL: https://radarcovid.gob.es/estadisticas/codigos-introducidos-a-casos-confirmados (visited on 11/17/2021).

[46] Hannah Ritchie. *Coronavirus Pandemic (COVID-19) â the data Statistics and Research*. Mar. 2020. URL: https://ourworldindata.org/coronavirus-data?country=~EESP (visited on 11/17/2021).

[47] F.C Chiusi. *Digital contact tracing apps: do they actually work? A review of early evidence*. July 2021. URL: https://algorithmwatch.org/en/analysis-digital-contact-tracing-apps-2021/.

[48] T.D Daigle and P.Z Zimonjic. *Federal COVID Alert app caught 400 cases of COVID-19 in April*. May 2021. URL: https://www.cbc.ca/news/politics/federal-covid-app-first-numbers-data-1.6040360 (visited on 11/17/2021).

[49] Federal Statistical Office. *Population*. URL: https://www.bfs.admin.ch/bfs/en/home/statistics/population.html (visited on 11/17/2021).

[50] Srf Data. *Coronavirus: the latest numbers*. Nov. 2021. URL: https://www.swissinfo.ch/eng/swiss-stats_coronavirus--the-latest-numbers-/45674308 (visited on 11/17/2021).

[51] G.G Grekousis and Y.L Liu. *Digital contact tracing, community uptake, and proximity awareness technology to fight COVID-19: a systematic review*. Tech. rep. May 2021. DOI: 10.1016/j.scs.2021.102995. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8114870/.

[52] National Governors Association. *COVID-19 Vaccine Incentives*. Oct. 2021. URL: https://www.nga.org/center/publications/covid-19-vaccine-incentives/ (visited on 11/17/2021).

[53] *Joint Statement on Contact Tracing*. Tech. rep. Apr. 2020. URL: https://giuper.github.io/JointStatement.pdf (visited on 11/17/2021).

[54] Susan Landau. "Digital exposure tools: Design for privacy, efficacy, and equity." In: *Science* 373.6560 (2021), pp. 1202–1204. DOI: 10.1126/science.abi9852.

[55] Nicole Wetsman. *Coronavirus contact tracing apps promised big and didn't deliver*. Dec. 2020. URL: https://www.theverge.com/22168473/coronavirus-contact-tracing-apps-exposure-notification-covid-google-apple (visited on 11/17/2021).

[56] Serge Vaudenay. *Centralized or Decentralized? The Contact Tracing Dilemma*. Cryptology ePrint Archive, Report 2020/531. https://ia.cr/2020/531. 2020.

[57] Laura Vallejo-Torres et al. "On the Estimation of the Cost-Effectiveness Threshold: Why, What, How?" In: *Value in Health* 19.5 (2016), pp. 558–566. ISSN: 1098-3015. DOI: https://doi.org/10.1016/j.jval.2016.02.020. URL: https://www.sciencedirect.com/science/article/pii/S1098301516000693.

[58] Julian Jessop. "The UK lockdown and the economic value of human life." In: *Economic Affairs* 40.2 (2020), pp. 138–147. DOI: 10.1111/ecaf.12417.

[59] eHealth Network. *European Proximity Tracing An Interoperability Architecture for contact tracing and warning apps*. Tech. rep. Sept. 2020. URL: https://ec.europa.eu/health/sites/default/files/ehealth/docs/mobileapps_interop_architecture_en.pdf.

[60] Scott Brown. *The Huawei ban explained: A complete timeline and everything you need to know*. Aug. 2021. URL: https://www.androidauthority.com/huawei-google-android-ban-988382/ (visited on 11/17/2021).

[61] Deutsche Welle (www.dw.com). *US labels five Chinese tech firms security risks*. URL: https://www.dw.com/en/us-designates-huawei-four-other-chinese-tech-firms-national-security-threats/a-56860474 (visited on 11/17/2021).

[62] Guardian Interactive Team and Ewen MacAskill. *NSA files decoded: Edward Snowden's surveillance revelations explained*. Nov. 2013. URL: https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded#section/1 (visited on 11/17/2021).

Appendix A

Source Code

Source code can be found on this github repository:
https://github.com/BernhardUIB/Thesis/tree/main/1.0

Appendix B

CSV

Example code executed from the program, which is also used to create the figures in section 4.2.3 can be found here: https://github.com/BernhardUIB/Thesis/tree/main/CSV