TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Hayder Hasan Ali Al-Sabti - 184072IVSB

Guidelines for Developing Coronavirus Tracing Applications: the Case of HOIA

Bachelor's thesis

Supervisor: Kaido Kikkas

Doctor of Philosophy
(PhD) in Engineering

Tallinn 2021

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Hayder Hasan Ali Al-Sabti - 184072IVSB

Juhised koroonaviiruse jälgimise rakenduste arendamiseks HOIA näitel

bakalaureusetöö

Juhendaja: Kaido Kikkas

Tehnikateaduste doktor

Tallinn 2021

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Hayder Hasan Ali Al-Sabti

17.05.2021

Abstract

There are currently insufficient general guidelines and best practices for developing Coronavirus tracing applications in a way that maximizes their effectiveness. This thesis aims to fill that gap by providing development guidelines for software developers and cyber security specialists.

A quantitative online survey was conducted using a non-probability voluntary response sampling method. The author analyzed the possible missing features within the HOIA application and researched the most common issues people have with Coronavirus tracing applications in general, and the questions were composed based on that. The survey consisted of nine Dichotomous Questions, five 4-Point Likert Scale System Questions and 3 Multiple Choice Questions, and a total of 180 responses were received.

Analysis of the responses revealed that 30% of the participants who used the application and 70% of the participants who thought it was not useful were willing to provide their location if that improved the accuracy of the application and therefore help contain the virus and save more lives, contrary to the belief that people would value their privacy over everything.

This thesis claims that using the guidelines proposed by the author, such as adding statistics and optional location tracking feature, software engineers and cyber security specialists can ensure that the maximum number of people would use the exposure notification technology and thus, helping with tackling the pandemic more effectively.

This thesis is written in English and is 59 pages long, including 7 chapters, 26 figures.

Annotatsioon

Hetkeseisuga ei leidu piisavalt üldisi juhiseid ja häid tavasid koroonaviiruse jälgimise rakenduste arendamiseks efektiivsel viisil. Käesolev diplomitöö püüab seda lünka täita, pakkudes välja arendussuunised tarkvara loojatele ja küberturbe asjatundjatele. Töös viidi läbi veebipõhine kvantitatiivne uuring, kasutades mugavusvalimit. Autor analüüsis võimalikke HOIA rakendusest puuduvaid funktsioone ning kasutajate tagasiside põhjal selgunud peamisi probleeme selle kasutamisel, koostades uuringu küsimused selle põhjal. Küsimustik koosnes 9 kahe alternatiiviga küsimusest, 5 neljapunktilise Likerti skaalaga küsimusest ning 3 valikvastustega küsimusest, vastused saadi 180 inimeselt. Vastuste analüüs näitas muuhulgas, et 30% rakenduse kasutajatest ja 70% neist, kes pidasid rakendust vähetõhusaks, olid nõus avaldama enda asukohta, kui see aitab rakendust täpsemaks muuta ning nii haigust piirata ja elusid päästa. See ei vastanud eeldusele, et inimesed hindavad enda privaatsust üle kõige. Diplomitöös leitakse, et väljapakutud juhiste (nagu täiendav statistika ja vabatahtlik asukohatuvastus) abil saaksid tarkvaraarendajad ja küberturbespetsialistid maksimeerida nakatumisteavituste tehnoloogia kasutamist ning nõnda pandeemiaga tõhusamalt võidelda.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 59 leheküljel, 7 peatükki, 26 joonist.

List of abbreviations and terms

AEM	Associated Encrypted Metadata
API	Application Programming Interface
AU_RAND	128-bit Random Challenge Number
BD_ADDR	Bluetooth Device Address
BLE	Bluetooth Low Energy
CA	Certificate Authority
CCPA	California Consumer Privacy Act
CDC	Centers for Disease Control and Prevention
CIA	Central Intelligence Agency
COVID-19	Coronavirus Disease 2019
DoS	Denial of Service
eSCO	Enhanced Synchronous Connection-Oriented
FAQ	Frequently Asked Questions
FDA	Food and Drug Administration
GAEN	Google Apple Exposure Notification
GBP	Great Britain Pound
GDPR	General Data Protection Regulation
HIPPA	Health Insurance Portability & Accountability Act
ID	Identifier
iOS	iPhone Operating System
MAC	Media Access Control
OWASP	Open Web Application Security Project
PC	Personal Computer
PIN	Personal Identification Number
QR	Quick Response
SCO	Synchronous Connection-Oriented
SSL	Secure Sockets Layer
UK	United Kingdom
USD	United States Dollars
UUID	Universal Unique Identifier
VPN	Virtual Private Network
WHO	World Health Organization

Table of contents

List of figures

1 Introduction

The problem addressed in this thesis is that there are insufficient general guidelines and best practices in developing Coronavirus tracing applications. Without a pre-defined set of guidelines, software developers and security specialists end up taking different approaches in developing these applications, some of which are less effective than others [1].

It is no secret that the COVID-19 pandemic has altered our lives drastically since its first discovery in Wuhan, China in December 2019. What makes the virus particularly dangerous is the fact that it has a high infection rate and a relatively long incubation period [2]. One of the many solutions that were implemented to mitigate the damage of the virus was Coronavirus tracing applications. These applications aim to reduce the danger of the infection rate property of the virus by notifying people who may have been in close contact with an infectious person and asking them to self-isolate before their symptoms start manifesting [3].

The goal of this thesis is to formulate a set of guidelines for the development of Coronavirus tracing applications with cyber security and efficiency in mind. These guidelines aim to maximize user satisfaction and application effectiveness by ensuring a combination of privacy, usability and security of these applications. The more people use them, the more we can reduce the infection rate of the virus and thus combat the pandemic more effectively.

A lot of people are understandably skeptical about these applications, as they are concerned about a variety of issues, ranging from privacy concerns to battery drainage concerns to Bluetooth security [4]. Different countries have gone in different directions regarding how to implement these applications, with some countries valuing privacy more and some valuing effectiveness more. Some have opted for a centralized privacy model while others opted for a decentralized approach. In order to reduce the number of differences between these applications and speed up their development and cross-border compatibility, Google and Apple collaborated to create the GAEN API which proposed

an efficient model in developing Coronavirus tracing applications with privacy as a priority [5].

The author focused on the case of the HOIA application, the official Coronavirus tracing application in Estonia, but some principles can possibly be applied to other Coronavirus tracing applications developed in other countries.

The first part of this thesis describes the impact the pandemic has had on cyber security as well as the security and privacy of the technologies used in Coronavirus tracing applications. The second part explains the rationale behind using a quantitative survey for gathering the data necessary to formulate the guidelines and provides an analysis of the data set. The final part includes the proposed set of guidelines as well as suggestions for future research on the topic.

2 Theoretical Background

This chapter covers the basics of information security and user behavior, the impact of COVID-19 on cyber security, Bluetooth security, the Exposure Notification API proposed by Apple and Google as well as the difference between centralized and decentralized privacy approaches.

2.1 COVID-19 and Its Impact on Cyber Security

COVID-19 is an infectious disease caused by severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2). It was first identified in Wuhan, China in December 2019. Since then, the virus has spread worldwide causing a global pandemic. The virus is characterized by its high infectious rate as well as its relatively long incubation period, meaning that you can carry the virus and infect other people without showing any symptoms for a few days.

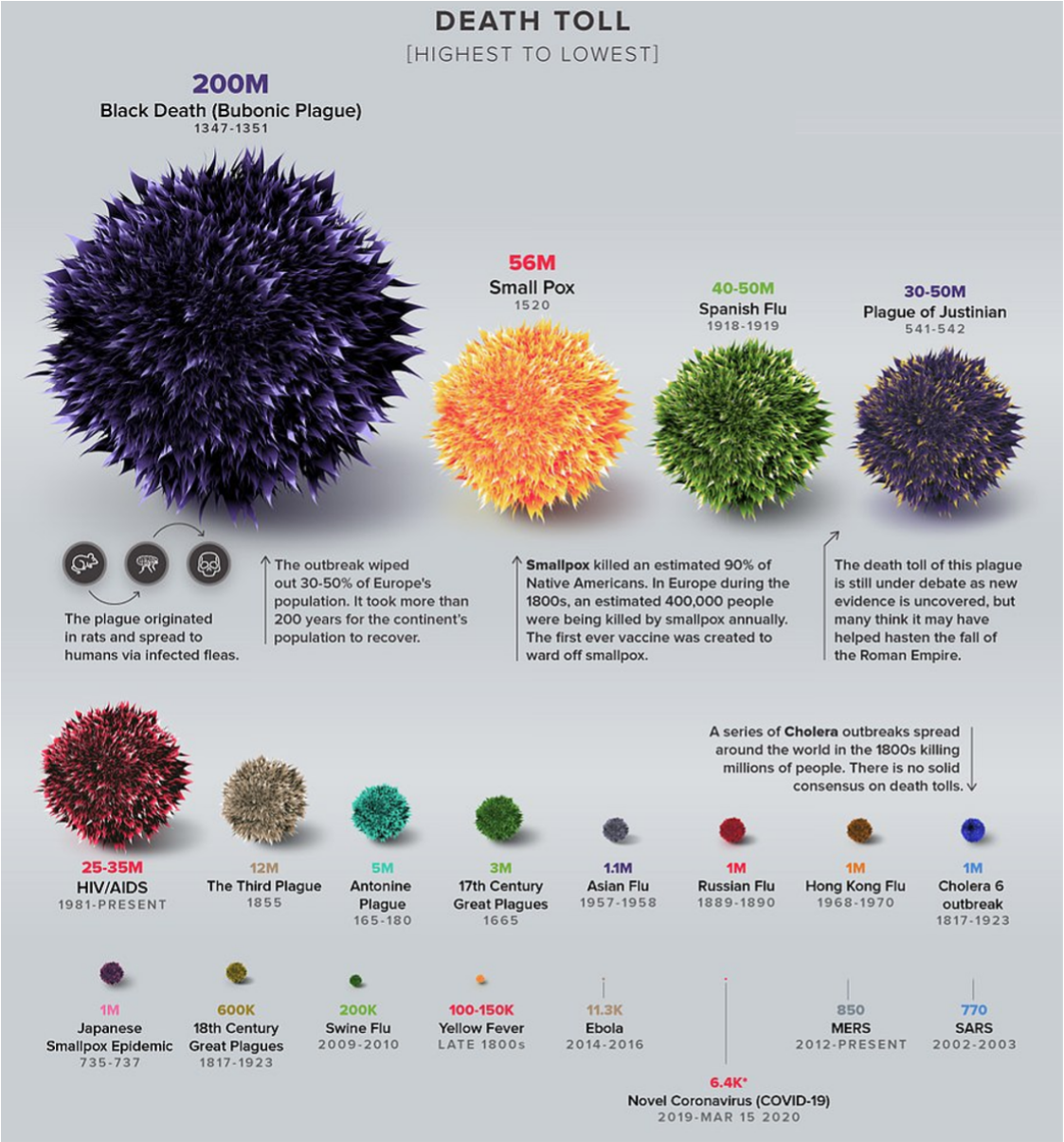

Figure 1. History of death tolls during different pandemics [6].

The current global pandemic provides the perfect environment for cyber criminals to execute their malicious intentions. This is due to several reasons [7], some of which are:

1. People are relying a lot more on digital infrastructure than ever before.

2. People are spending most of their time anxiously and curiously consuming media regarding the pandemic, making them more susceptible to click on malicious links, fall for social engineering traps and indulge in risky online activities.

3. People who are not very technically literate are now forced to get accustomed to a significant shift in their work environment. A lot of people were underprepared for having to suddenly do all their work from home.

4. Businesses are urgently tasked to implement working from home policies, employees are under more danger in their home offices as the company firewalls are not there to protect them. Some enterprise VPNs are not handling the load to access the network, leading to less productivity when working from home.

All these factors have caused a serious increase in cyber attacks. It is worth noting that the previous pandemics also had an effect on cyber security; Ebola, for example, has caused losses of over 53 billion USD in both social and economic aspects in West Africa [8].

The current pandemic provides a great environment for social engineering attacks. Social engineering is defined as "the science of using social interaction as a means to persuade an individual or an organization to comply with a specific request from an attacker where either the social interaction, the persuasion or the request involves a computer-related entity" [9]. The vulnerable emotional state of individuals during this pandemic leaves them as easy prey for this type of attack. The human element has always been the hole within security systems, thus exploiting this element has been of utmost importance to cyber criminals in order to compromise both individuals and enterprises.

One of the most prominent social engineering techniques that has been used is phishing. Phishing can take many forms, for instance as emails from tax authorities offering people tax refunds to help them cope with the effects of the pandemic. Usually, these phishing scams happen in tandem with global trends and news [10]. An example is the smishing scam that happened in the US offering 1000 USD to each citizen to support them during the pandemic [11].

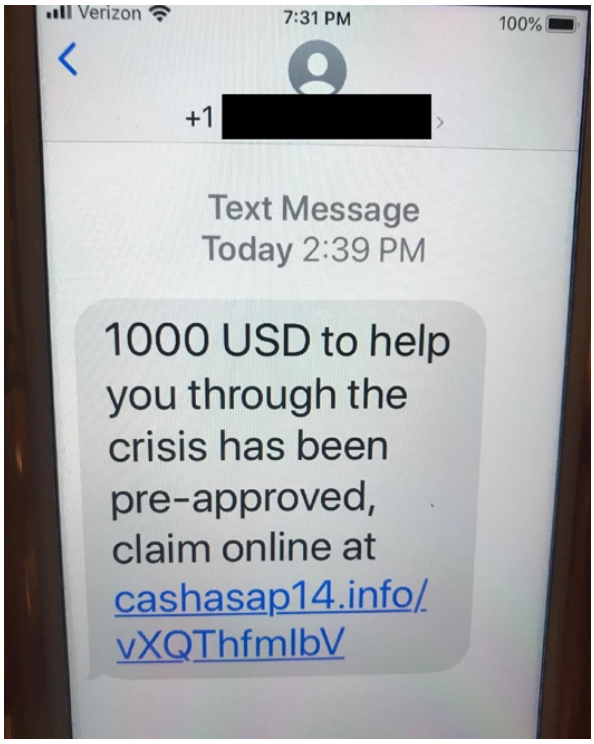

Figure 2. Example of a Smishing attack [11].

Possible smishing attacks to expect in the current period are:

– Your COVID-19 tests are here! Click on this link to download the results!

– It is unbelievable what this country did to stop the outbreak, click here to know more!

– Your free COVID-19 testing center is finally opened. Click here to make your free booking!

The abovementioned attacks are fictitious, but similar attacks can be found online.

Another technique social engineers have been using during this pandemic is setting up donation webpages where people can donate money to fund the research in finding a cure for COVID-19. Since people are desperate and it is in our good nature to want to help and assist, this has been very effective. One of the notable attacks was one where almost 2 million USD were stolen through cryptocurrency donation scams. The attackers asked the victims to donate in bitcoins, which made it nearly impossible to trace where that money ended up eventually [12].

Furthermore, at the beginning of the global outbreak, there has been an increase in claiming domains that have COVID-19 related vocabulary and domains that contain typos such as “coronovirus”, and then turn these websites into malicious sites for information theft. Examples of domains that have been registered in March 2020 are buycoronavirusfacemasks.com, beatingcorona.com, combatcorona.com, corona-emergency.com, coronadatabase.com, corona-crisis.com, coronadetection.com [13].

It is unfortunate that useful domain names such as ‘beatingcorona’ and ‘coronadetection’ have already been claimed by hackers with malicious intentions. These domains could have provided a quick and easy way for people to locate important information.

In addition to that, due to the increased interest in Coronavirus statistics maps, there has been a rise in malicious websites pretending to provide this type of information. An example is the maps that mimicked the Johns Hopkins University COVID-19 map, and it required the user to download a malicious plugin in order to show the information. That plugin allowed the attacker to gain remote access to the victim’s system [14].

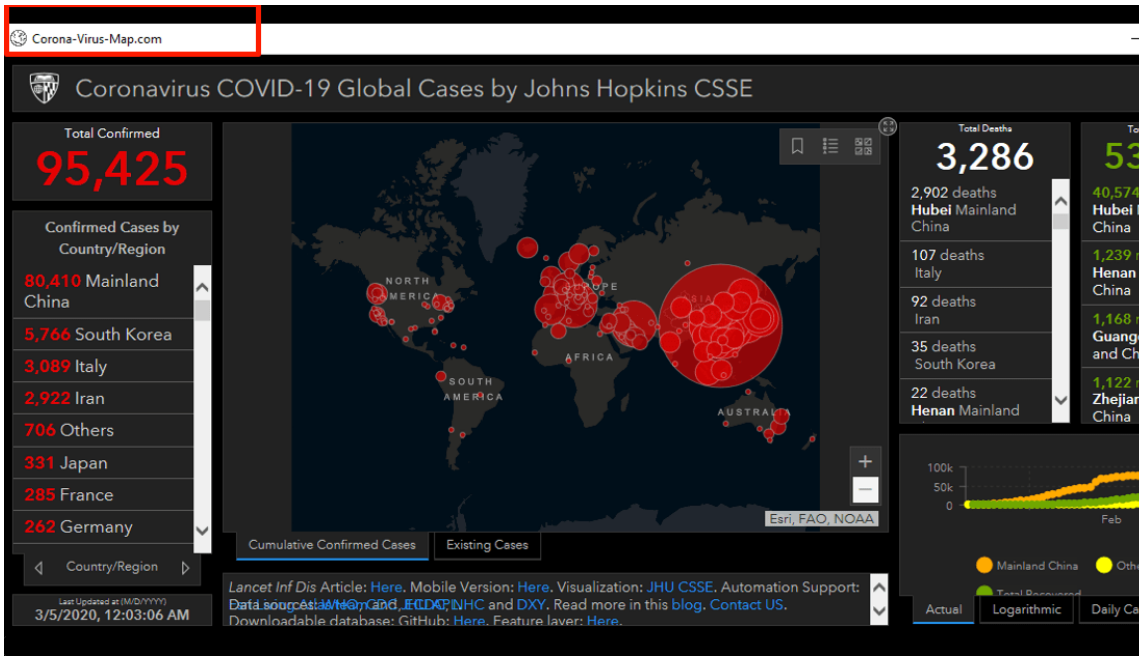

Figure 3. Fake COVID-19 map [14].

Another example is the websites pretending to be official communication channels for the pandemic, such as the WHO or the CDC. These websites prompted visitors to download documents that claimed to contain safety tips. These documents contained malware that would steal banking details and keylog users’ passwords [15].

Figure 4. Example of Phishing during the COVID-19 pandemic [11].

Finally, one of the driving factors of the success of the abovementioned attacks is the widespread of misinformation [16]. People generally tend to share any news that has an alarming, urgent or hopeful tone to them. This gives social engineers the ability to easily manipulate people by publishing this fake news in a way that would capture the public's attention. The problem is serious enough that the Humanitarian-to-Humanitarian Network has invested 500000 GBP to fight against misinformation [17]. Other countries, such as South Africa, have taken more extreme measures by placing a legal requirement not to spread fake news. It has been labeled as a criminal offence as of the 18th of March 2020 [18].

There are various ways of spreading misinformation, but the trending ones were regarding fake COVID-19 medications. Some attackers claimed to be doctors oppressed by global governments who already have the vaccines available. Another case was when Donald Trump, the US president at the time, has made claims that a medication is available and approved by the FDA, which was later announced to be false [19]. Most of the misinformative websites require users to register to view their content and news, thus, the hacker can gain personal information from the individual [7].

2.2 Users' Behavior Towards Smartphones and Notifications

Users tend to have preconceived perceptions about their smartphones that can subconsciously affect the way they use them or not use them. One study by Chin et al [20] shows that people are more concerned with privacy on their smartphones than they are on other devices such as their PC or laptops. Users tended to avoid carrying out sensitive tasks on their smartphones such as making important money transactions. Other smartphone misconceptions originated from misconceptions regarding the security of wireless connections.

When it comes to smartphone applications, users tended to be more likely to download games or entertainment applications without verifying the legitimacy of the application developer.

With regards to smartphone notifications, another study by Pielot et al [21] showed that receiving a large number of notifications, especially from social media and email applications, was correlated with feelings of stress and overwhelm. However, instant messaging notifications in particular were associated with feelings of being connected with other people. Moreover, while the initial thought is that notifications should be sent out immediately to grab the user's attention when they are busy doing other tasks, this study revealed that users tended to view notifications that are sent to them when they start interacting with their phones.

The authors of that study suggest the following solutions to help improve user's confidence and trust in smartphones:

1. User education: it is vital to educate users about various concepts such as wireless security, application security, end-to-end encryption and data privacy. There needs to be more effort put into simplifying these concepts for the average user to understand them clearly. Some lab studies showed that user education and raising awareness when it comes to security has proven to be an effective procedure.

2. Security indicators: the authors recommend adding security indicators within application marketplaces or stores to help users understand what data these applications collect as well as what permissions they would request. It is worth noting that as for now, the App Store on iOS and Google Play on Android have these security indicators as there has been a large shift in moving towards transparency after the scandals that Facebook and other companies have faced for their user agreement policies.
3. Improving user experience in applications dealing with sensitive data or tasks: more effort should be put into improving the usability of applications that deal with critical tasks such as banking applications. For instance, adding indicators within the application that the connection is secure and have multiple verifications for the user before executing a transaction. A good example of that is banking applications in Estonia requiring users to verify a transaction with Smart-ID [21].

2.3 The Current State of the HOIA application

As of March 2021, the HOIA application has been downloaded 271,024 times but only 5,894 people have registered themselves as infected. That is a low number as it potentially means very few people are reporting themselves as infected when they contract the infection. The insignificant help the HOIA application has provided can be attributed to multiple factors. Firstly, the development process for the application has been noticeably slow. The application was developed in cooperation with 12 Estonian companies in the private sector with the assumption that the state would continue developing the application after its initial release. However, that was not the case as the bureaucracy of procurements has hindered the development process. Secondly, there is an unclear definition of who owns the application. Officially, the technical administrators are the Health and Welfare Information Systems Center (TEHIK), but the social ministry states that the application is under the responsibility of the Health Board. Finally, the identity verification within the application can only be done through Smart-ID and Mobile-ID. Therefore, if a person is not using those two technologies, then there is no way for them to register themselves as infected. It is argued that there are already 500,000 Smart-ID users plus 250,000 Mobile-ID users, but that still leaves a significant portion of the population of Estonia unable to register their infection status, especially if we consider

that some people use both technologies, so those numbers do not reflect that 750,000 users are able to register [22, 23].

2.4 Comparison Between COVID-19 Tracing Applications

Patrick Howell O'Neill et al [1] have compiled a list of COVID-19 tracing applications around the world and compared them based on five parameters:

1. Is downloading the application and using it voluntary?
2. Are there limitations imposed on the usage of the collected data? Meaning that it does not get used for purposes other than public health.
3. Is the data temporary and destroyed after a specified period of time?
4. Does the application collect the minimum amount of data necessary?
5. Is the application's code base open-source and transparent?

In Europe, countries like Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Iceland, Italy, and Norway all met the five parameters stated above. Iceland uses the user's location as its tracking method as opposed to Bluetooth technology used by the rest of the mentioned countries. Estonia's HOIA application fulfills only two parameters, the voluntary parameter and the data minimization parameter [1].

Another analysis conducted by Kevin Watkins [24] compares the different accesses that some COVID-19 tracing applications require. Most notably in Europe, the Polish application can access the user's calendar, microphone and has data transport security exceptions. The Swiss application can access the user's contact list. The Macedonian application has SSL CA validation disabled, meaning that the gathered data is not sent securely to the server.

2.5 Information Security

Information security is the practice of protecting information by preventing the unauthorized or unlawful access, deletion, disruption, leakage, modification or corruption

of that information. It also includes the necessary steps that need to be taken to mitigate the effects of these incidents [25].

Information security focuses on balancing the protection of the CIA triad – Confidentiality, Integrity and Availability [26]:

1. Confidentiality: It refers to the property of information not being revealed to unauthorized individuals. It is a component of privacy that aims to secure our data from being viewed by users who should not have access to that data. An example of compromised confidentiality would be password theft.

2. Integrity: It refers to the property of information not being modified in an unauthorized manner. In other words, when the data is read by the receiver, it should be exactly the same as when it was written. An example of compromised integrity would be a man in the middle attack where the hacker is in between the two communicating parties, deceiving them into thinking they are communicating directly with each other when the entire conversation is being controlled by the hacker as he changes the content of a message before it reaches the receiver [27].

3. Availability: It refers to the property of information being available when it is needed. This means that all the interconnected systems that are meant to provide this information must be securely protected and maintained. An example of compromised availability would be a DoS attack on a server causing it to be irresponsive and, thus, causing the data to be unavailable for those who query it. A DoS attack involves flooding the victim's network with traffic to the point of shutting it down [28].

Another main principle of information security is Non-Repudiation. It means that the sender of a transaction cannot deny sending it, and the receiver of a transaction cannot deny receiving it. It encapsulates authenticity and accountability. Authenticity means that users are whom they claim they are, and accountability means that it is possible to trace the activities of an individual back to that individual.

Bluetooth, the technology used by a lot of Coronavirus tracing applications, provides the following information security services [29]:

1. Authentication: In this step, the identities of Bluetooth devices are verified. Authentication on the user level is not provided.

In this authentication process, one device is named the claimant and the other device is named the verifier. The role of the claimant is to prove his identity, while the role of the verifier is to validate the identity of the claimant. The claimant begins the pairing process by sending an access request to the verifier alongside a secret 128-bit random number to be used as what is referred to as the link key. After that, the claimant sends an authentication request with its BD_ADDR, which is a unique 48-bit identifier assigned to each Bluetooth device by the manufacturer. The verifier then responds to the claimant with a challenge 128-bit random number referred to as the AU_RAND. Then both devices perform calculations using the BD_ADDR, the AU_RAND and the link key. Then, the claimant sends his result to the verifier, who compares it to his result. If the calculated value matches, then authentication is successful, otherwise the authentication fails.

2. Confidentiality: This step ensures that only authorized devices can access the data transferred between the paired devices. This is done by using encryption with the E0 stream cipher in which a stream of pseudorandom characters is generated using the BD_ADDR and the link key, which are then combined with a plaintext message to produce a ciphertext.

3. Authorization: This step allows only authorized devices to access Bluetooth's functionalities. It is done through checking a database located on the device itself to determine whether the device requesting to pair has been previously authorized or not. If it has been authorized before, then access is granted. Otherwise, authentication must be performed first before authorization is established.

Security services that are not included in the Bluetooth standard include integrity, non-repudiation and audit.

Bluetooth's issues in the past have gone unnoticed because of its low importance, as it was mostly used to connect to wireless home devices. The case is not the same with contact tracing technology, as a false negative result can mean threatening the lives of many people, while a false positive result can bring a lot of unnecessary anxiety and inconvenience in terms of self-isolation and testing.

The founders of Bluetooth technology have stated that one of the main issues with it is the detection range. The signal reduction can differ a lot depending on the environment in which it is propagated, such as if the space is crowded or has a lot of metal objects [30].

Bluetooth has also been suspectable to numerous types of attacks in the past.

The security vulnerabilities in Bluetooth are mostly due to the step of pairing two devices together. Different attacks can be performed before the pairing process is completed. Even after the pairing of devices is complete, an adversary can still steal enough information and perform a Man in the Middle attack. The most important factor in the assessment of Bluetooth vulnerabilities is the version of Bluetooth that is being used, and the security of communications between devices depends on the device with the older version. Since many older smartphones are still being used today, the vulnerabilities in the older versions of Bluetooth continue to be a threat [31].

The following is a diagram illustrating these attacks:

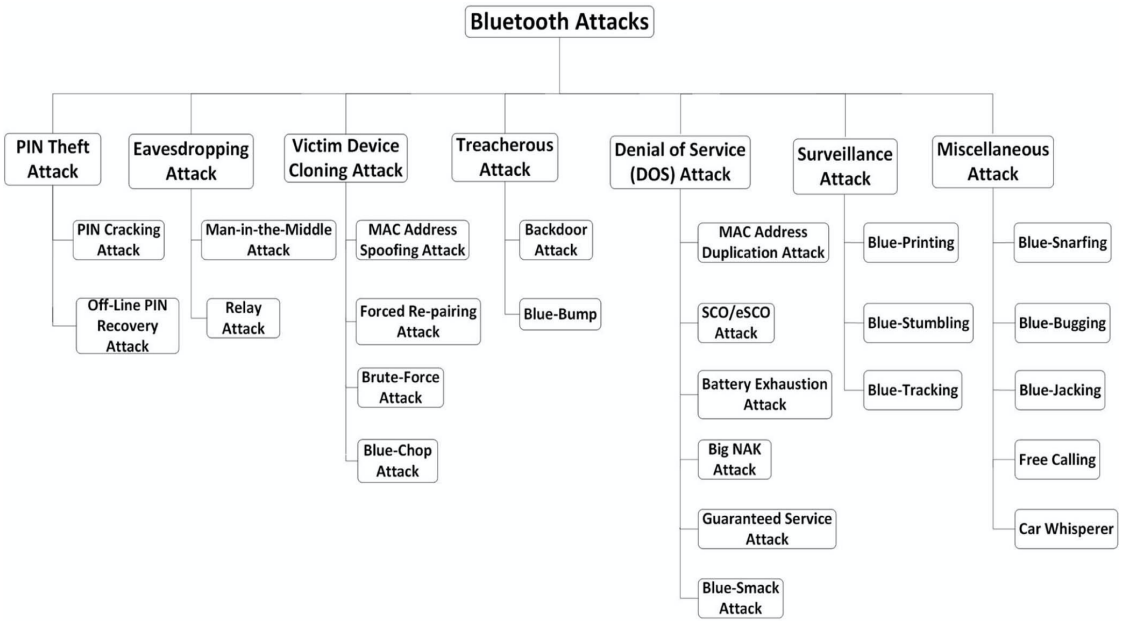

Figure 5. Bluetooth attacks diagram [31].

However, newer versions of Bluetooth are significantly more secure than the previous versions. Having said that, it is important that users take necessary steps to further secure themselves such as updating their phone systems since these updates can include security patches for Bluetooth [32].

When it comes to accuracy, Bluetooth's accuracy is debatable and is related to a lot of factors that can cause false negatives or false positives, that is why Coronavirus tracing applications relying on Bluetooth technology should be treated as supplementary tools instead of the only tools because additional measures should always be taken to ensure optimal safety [33].

Finally, there are multiple security standards that COVID-19 tracing applications should follow. The OWASP foundation has a Mobile App Security checklist from which some examples can be drawn in the case of these applications:

- Data considered sensitive in the context of the mobile app is clearly identified.
- The app should comply with privacy laws and regulations.
- The app does not hold sensitive data in memory longer than necessary, and memory is cleared explicitly after use.
- The app only requests the minimum set of permissions necessary [34].

2.6 Centralized and Decentralized approaches

When it comes to Coronavirus tracing apps, there have been two security models that were adopted. Some countries, like the UK, have originally opted for the centralized solution [35]. However, most countries have opted for the decentralized solution since it was the more privacy-preserving model implemented by Google and Apple.

With the centralized model, the anonymized IDs of the user and the people he got in contact with are gathered and uploaded to a remote server where matches are made with other users' anonymized IDs in case someone reported an infection status.

On the other hand, the decentralized model provides more privacy, because only the user's anonymized ID is uploaded to the remote server, while the IDs collected from other users remain on the phone. The Coronavirus tracing applications then download the remote server's database and do the matching process locally on the device itself, thus providing the server with very little information and maintaining a relatively better level of privacy [36].

Each model has its advantages and disadvantages. In reality, both models are genuine attempts at providing a system that helps to combat the pandemic. However, when that system is going to be used by millions of people, genuine objectives are not sufficient to ensure that the system will not be exploited in the future. That is why it is vital to minimize the risk of the system being abused for purposes other than what it was originally meant for, such as surveillance of the masses. The centralized model is prone to attacks that could cause it to be potentially used for malicious purposes. The main issue with this model is that the authority controlling the server can link the different anonymized IDs broadcasted by the same person. So, technically speaking, if that authority installs Bluetooth sensors all across the country, they will be able to track the location of every user of the application for the entire period of them using it. It can be argued that the IDs obtained are pseudo-anonymous, however, research by De Montjoye et al [37] has shown that these IDs are generally easy to re-identify. This research illustrates how in 95% of the cases, only the location and time data points are enough to uniquely identify a person in a dataset containing millions of users. Through that data, the authority can deduct all the places that a user has visited while the application was in use. Another issue with the centralized model is that people who test positively upload their anonymized IDs alongside all the IDs collected from the people they came in contact with, allowing the authority to construct a social graph of the population. Since the IDs are anonymized, the social graph would be pseudo-anonymous. However, some researchers have proven that pseudo-anonymous social graphs can be re-identified in some instances [38, 39].

All of the abovementioned attacks are very complex to achieve, they only work on positively tested users who report their status to the application and for a duration of 14 days only. However, for some users who work as journalists, politicians or activists, that data poses a serious threat. Thus, from an information security point of view, it is better to use a model that is not vulnerable to the possibility of such attacks. In this model, the adversary is only the remote server.

When it comes to the decentralized model, the system makes Bluetooth broadcasting of infected users public, which could also lead to mass surveillance. In this case the adversary can be any malicious person. Further work needs to be done to find a hybrid solution combining the benefits of both systems and mitigating their dangers [36].

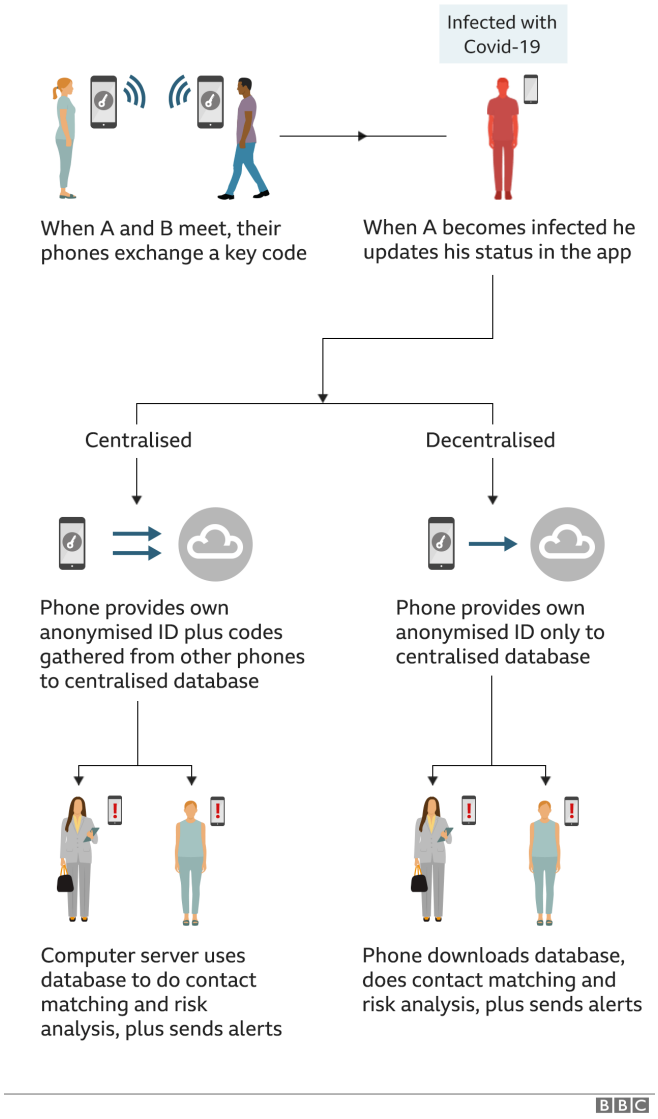

Figure 6. Centralized vs Decentralized privacy approaches [40].

2.7 GAEN API

As the pandemic situation escalated quickly, Google and Apple decided to partner up to create an API that can be implemented by countries wanting to deploy their own Coronavirus tracing application. The first version was made available on the 20th of May 2020 [41]. The API is based on the decentralized model and promises both privacy and security for the user.

Some of the key privacy features of the GAEN API are [42]:

1. You choose whether to enable receiving exposure notifications or not. It can always be turned off.

2. It does not involve any location tracking. The API utilizes Bluetooth technology to identify close contacts without tracking the location of the users.

3. The API does not ask for identity information. Apple, Google or other users cannot see your identity or have access to your data. However, in some cases, the authority developing the application may ask for information such as a phone number in order to contact you to provide safety information.

4. The API can only be accessed and implemented by health authorities. These health authorities must also comply with certain security, privacy and data usage specifications provided by Google and Apple.

5. The Bluetooth identifiers generated through the API are replaced every 10-20 minutes to prevent the unique identification of an individual.

6. Apple and Google specified that they would turn off the exposure notifications system when the pandemic is over, to prevent its use for other purposes than it was initially intended for.

2.6.1 Main Principles of the API

When the user enables the exposure notification technology and Bluetooth on their Coronavirus tracing application, their smartphone will routinely broadcast a Bluetooth beacon that contains a randomized ID that would change every 10-20 minutes. Other smartphones which have the exposure notification technology and Bluetooth enabled will be listening for these Bluetooth beacons as well as sending out their own beacons. When two devices are in close proximity from each other for long enough, they will receive each other's beacons and record that close contact interaction securely on the device. Each day, the device will be prompted to download a list from the database of the health authority containing the randomized IDs belonging to people who have reported themselves as COVID-19 positive. After that, the device will check its local list of close contact interactions and determine if there is a match between it and the list downloaded from the health authority database. In case there is a match, the person will be notified of their close contact with a COVID-19 positive person and will be instructed to self-isolate and stay at home. The user identity or infection status will not be shared with Google, Apple or other users when using this technology.

The API is designed with battery power draining in mind, as the continuous enabling of Bluetooth and broadcasting of Bluetooth beacons can drain the battery easily. Thus, Bluetooth Low Energy (BLE) protocol is used in order to filler through duplicate broadcasting beacons [42].

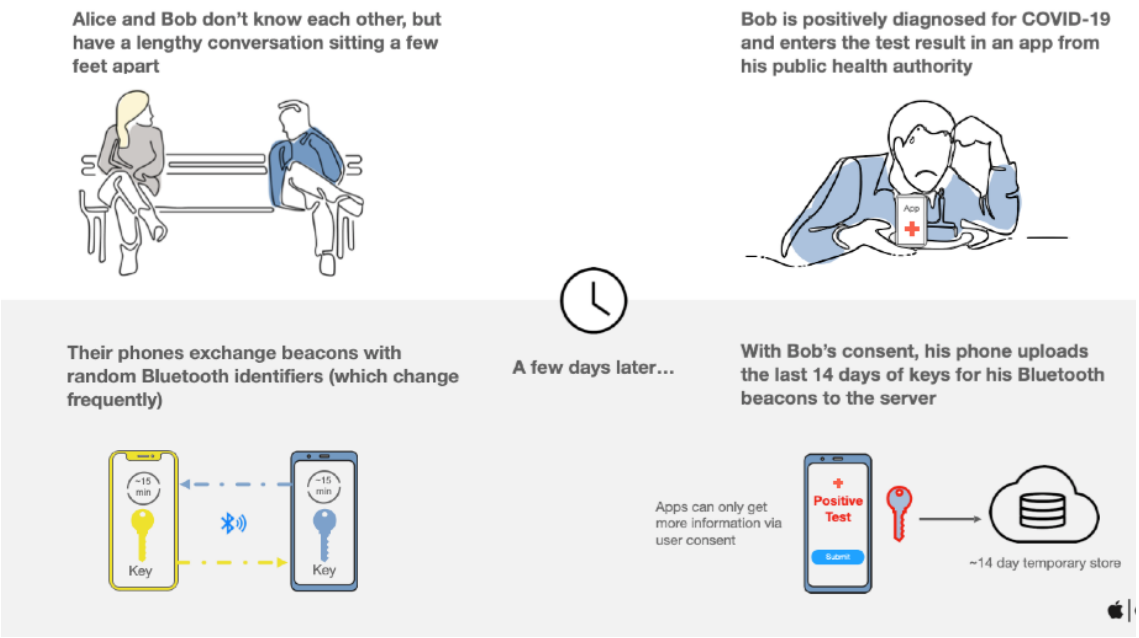

Figure 7. Registering contacts in the GAEN API [42].

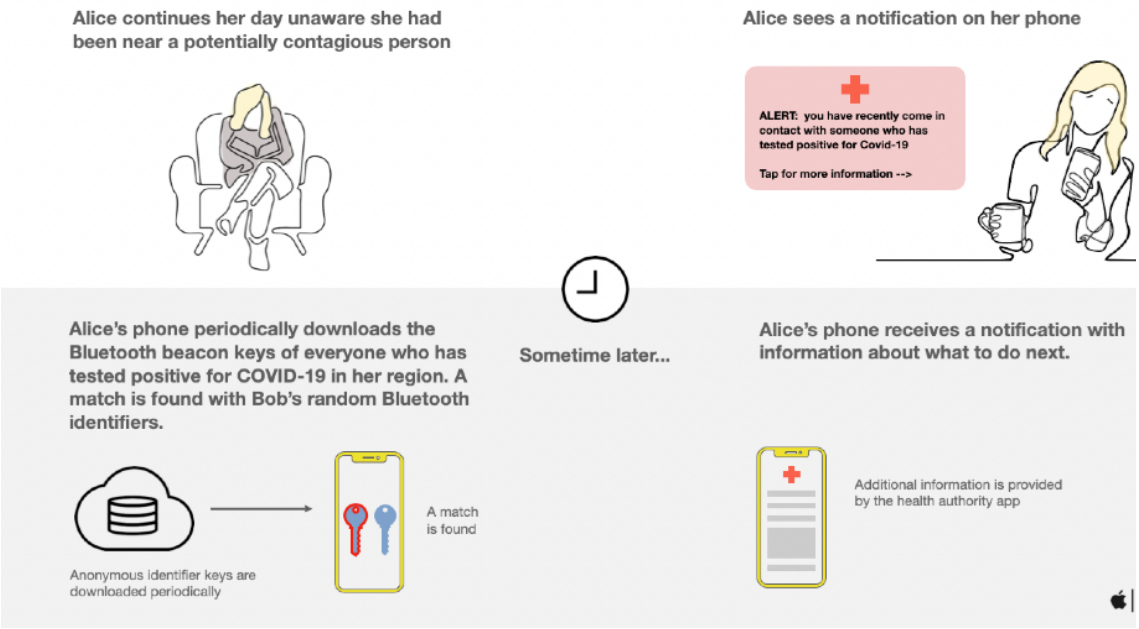

Figure 8. Exposure notification in the GAEN API [42].

2.6.2 Broadcasting and scanning flow in the GAEN API

The data used and exchanged during broadcasting and scanning is [43]:

1. Temporary Exposure Key: This is a randomized key that is generated once a day to increase the anonymity of users.

2. Diagnosis Key: The part of a Temporary Exposure Key that is uploaded to the health authority's database when a person reports themselves as COVID-19 positive.

3. Rolling Proximity Identifier: A unique random identifier that is updated approximately every 15 minutes to avoid wireless tracking of the person's device. It is derived from the Temporary Exposure Key and sent within the Bluetooth beacon during broadcasting.

4. Associated Encrypted Metadata (AEM): Encrypted metadata that is used to keep track of the versioning of the protocol as well as transmission of power for enhanced distance approximation. It is updated approximately every 15 minutes to avoid wireless tracking as well.

Broadcasting flow:

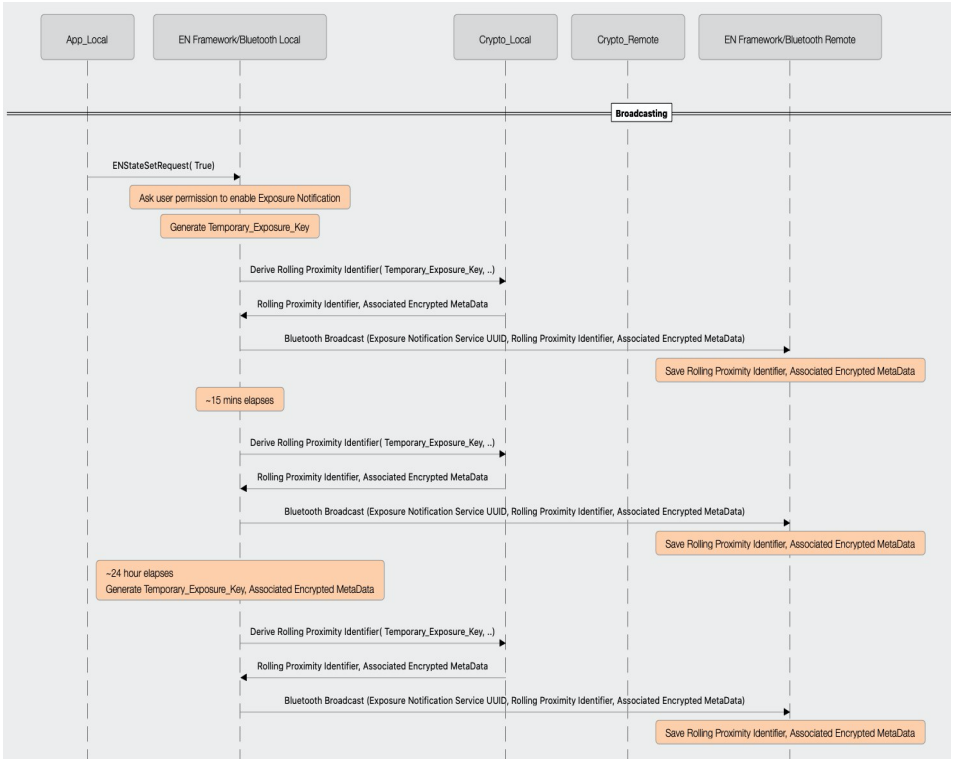

Figure 9. Broadcasting flow [43].

Scanning Flow:

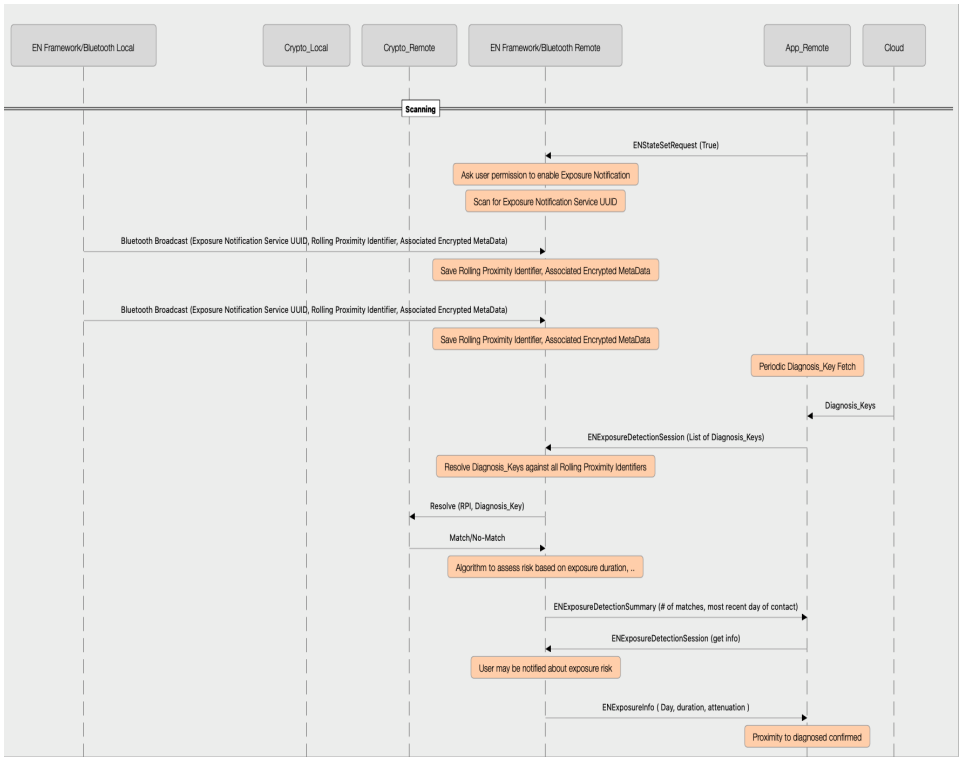

Figure 10. Scanning flow [43].

2.6.3 Measurement-based evaluation of the GAEN API

A study was conducted on a bus in Ireland using the GAEN API. It involved gathering data from 60 smartphones that were placed in different locations within the bus. Generally speaking, a radio signal gets weaker as it travels further from its transmitter since the transmission power is spread over a larger area. With the GAEN API, it uses BLE technology to broadcast beacons routinely and then, if the strength of the signal of the received beacon is high enough, a close-contact interaction is recorded. However, this study revealed that the transmission of radio signals, particularly indoors, is different and complex in practice as it involves a lot of factors such as floors, furniture and ceilings that can absorb or reflect these signals, affecting their strength. One other important factor is the human body since it is capable of absorbing BLE radio signals as well, meaning that the strength of a signal can be affected in case there was a person between the transmitter and the receiver. Therefore, the study concludes that the attenuation level calculated by the GAEN API does not necessarily increase with the increase in distance between smartphones, as there are a lot of other factors that can both reduce or increase a signal's strength [33].

2.6.4 Compliance of the GAEN API with the GDPR

A study conducted by Laura Bradford et al [44] investigates the compliance of the GAEN API with the GDPR in Europe as well as HIPPA and CCPA in the US. The study identifies four types of data within the GAEN API: Bluetooth identifier codes and associated contact event information, positive diagnosis information, associated information, as well as notifications to exposed users.

The following points are taken into consideration when it comes to the compliance of COVID-19 tracing applications with the GDPR:

1. Is This Data Relating to an Identifiable Natural Person?

The information gathered by the application should be personally identifiable information. The GDPR describes personal data as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or

more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person". The randomized broadcasted Bluetooth signals can be in some cases uniquely linked to a particular individual; thus, it falls under the GDPR's definition of personal data.

2. Is This a Special Category of Data Concerning Health?

The GDPR description of 'data concerning health' includes "data that reveal information about an individual's health status". This type of data requires extended protection. The data collected by COVID-19 tracing applications is to inform that the individual has been exposed to COVID-19, thus, this data should be treated as special category data under Article 9 of the GDPR.

3. Is the Data Anonymized or Pseudonymized?

The European Data Protection Board states that true anonymization is an extremely high bar and data gatherers typically do not truly anonymize their data. There are complex techniques that can re-identify users from data that is described as being anonymous. With the constant development of technologies, both data gatherers and users cannot be fully certain that their current anonymized data will remain anonymized in the future. This can be observed in the GAEN API because, even though Apple and Google claim that the data is anonymous, they still have implemented measures to prevent re-identification in their design, following the GDPR's data minimization and security of processing principles. These measures ensure that the data is at least pseudonymized. Pseudonymization is described by the GDPR as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."

The study states that "All of the GDPR Article 5 principles, (i) lawfulness, fairness, and transparency, (ii) purpose limitation, (iii) data minimization, (iv) accuracy (v) storage limitation, (v) integrity and confidentiality, and (vi) accountability, must be observed in the design and implementation of these systems. Articles 12–23 of the GDPR mandate clear 'pre-defined' rights for the data subject, including rights of access, rectification, and erasure."

When comparing GDPR to HIPPA and CCPA, the Privacy Rule of HIPPA encompasses data gathered by health providers or third-party commercial businesses hired by the health providers to analyze their data. This is problematic in the case of COVID-19 tracing applications, as an infection report using the GAEN API does not fall under this category, since the disclosure of the information is happening through the user and not directly through the health provider.

The limitation also applies to the CCPA since it is a consumer protection law. It excludes information that is covered by laws protecting the privacy of medical data, such as the data covered by HIPPA. The CCPA also does not distinctly state whether its legislations apply to pseudonymized data. Therefore, the data gathered through the GAEN API may be considered anonymized under the CCPA as opposed to pseudonymized under the GDPR.

The study concludes that the GDPR has a wider scope than HIPPA and CCPA, giving it the advantage to offer a blueprint for a system that is well suited with essential rights [44].

3 Methodology

In this thesis, the author used a quantitative research methodology by conducting a survey with the residents of Estonia as the sample size. The sampling method used was non-probability voluntary response sampling. The survey was distributed between social media groups for people residing in Estonia, as well as some of Tallinn's university of technology students and lecturers, during March 2021. The author relied only on the primary data set received from the survey and no other secondary data set was used. The reason for choosing that sampling method is due to the pandemic and social interaction restrictions that were enforced at the time of writing this thesis. Therefore, the author was forced to opt for distributing the survey online. The author's rationale for choosing a quantitative research method was to make a short yet concise survey to get the most number of people to fill it. A total of 180 responses were received and the initial minimum goal was to get 60 participants. The author believed that having a large number of opinions through a quantitative survey can give a more representative overview of the situation than conducting focused qualitative interviews with a small number of people.

The survey was made using Google Forms because it is sufficient for the author's needs, simple for the users and available for free. The data analysis was done using Pivot Tables feature in Google Sheets, as that feature helped to find the correlation between different answers within the data set. The author also relied on the Pie Charts generated by Google Forms. The data was prepared and corrected by removing blanks and searching for outliers.

The survey was based on nine Dichotomous Questions, five 4-Point Likert Scale System Questions and 3 Multiple Choice Questions. Dichotomous Questions are good for getting a clear distinction when analyzing the data. The author is aware of the debate that a 5-point Likert scale is more accurate than a 4-point Likert scale [45], but in the case of this thesis, it is a lot more fitting to not give a 'neutral' option for people to choose. The reason for this is that people's current status of using the application is already non-interactive

and the application does not have any features that incentivize the user to open it regularly. That is why choosing ‘neutral’ on a suggested feature means that it makes no difference to them, which means that it is going to maintain their current status of not interacting with the application, thus, it is more accurate to say that the user, in this case, disagrees with the suggested feature.

The author analyzed the possible missing features within the HOIA application, as well as researched the most common issues people have with Coronavirus tracing applications in general, and the questions were formulated based on that. The author also utilized questions provided by the Computational Privacy Group from the Imperial College London as a benchmark for how to develop private Coronavirus tracing applications and, thus, formulating the right questions about them. The suggestions were [46]:

1- How do you limit the personal data gathered by the authority?

2- How do you protect the anonymity of every user?

3- Does your system reveal to the authority the identity of users who are at risk?

4- Could your system be used by users to learn who is infected or at risk, even in their social circle?

5- Does your system allow users to learn any personal information about other users?

6- Could external parties exploit your system to track users or infer whether they are infected?

7- Do you put in place additional measures to protect the personal data of infected and at risk users?

8- How can we verify that the system does what it says?

4 Analysis

The aim of this chapter is to analyze the survey data in order to formulate the guidelines for developing exposure notification applications. The author used the charts provided by Google Forms as well as the Pivot Tables feature in Google Sheets to find correlations between various survey questions that would help deduct what users would find most useful and practical.

4.1 Application Effectiveness

- 73.6% of the participants disagreed that the HOIA application has been effective in battling the pandemic. Thus, proving the hypothesis of this thesis that the HOIA application requires further improvement to reach its maximum potential.

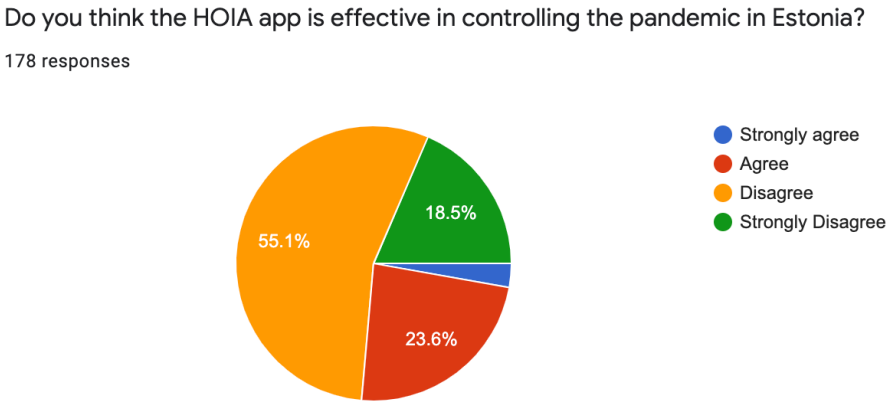

Figure 11. HOIA application effectiveness.

- 31.28% of the people who said they are not using the application answered that they would self-isolate if they had the application installed and it notified them to quarantine. This shows that there is potential to convince those people to switch to using the application if it was improved. This is extremely important as the more people use the application, the more likely we are able to control the infection rate of the virus. This further solidifies the rationale behind this thesis, as it shows that the people who are not using these applications are not using them because they think

they lack some necessary features or improvements, and they would be willing to use them and improve their efficiency in battling the pandemic if those missing features were addressed.

Do you use the HOIA app? (You have downloaded it and enabled bluetooth)	*Would you quarantine if the you received a notification that you have been in close contact with a COVID positive person?*	
	No	Yes
No	5.59%	31.28%
Yes	1.68%	61.45%

Figure 12. Users' feedback on self-isolation through the HOIA application.

- 35.96% of the participants who are using the HOIA application agreed that it should be mandatory. How to make it mandatory is a question outside of the scope of this thesis, but the author wanted to point out what people think about making the usage of the application mandatory to increase its efficiency.

Do you use the HOIA app? (You have downloaded it and enabled bluetooth)	*Do you think using the HOIA app should be mandatory?*	
	No	Yes
No	29.78%	6.74%
Yes	27.53%	35.96%

Figure 13. Users' opinion on making the HOIA application mandatory.

4.2 Reasons for Not Using the Application

The top reasons the survey participants selected for not using the HOIA application were:

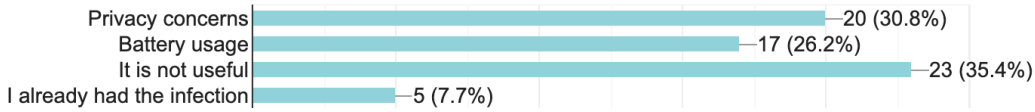

Figure 14. Reasons for not using the HOIA application.

This indicates that the top two reasons, privacy concerns and battery usage, need to be addressed seriously to get more people to use the application. Currently the application uses BLE to consume less battery power, but more research should be done to find optimizations that can be done to reduce the battery drainage when the person is, for instance, at home or not around a lot of people.

Some of the other reasons participants have given for not using the HOIA application were:

- They use a mask and avoid public spaces, so they do not need it.
- They have older devices that the application does not support.
- They have not heard of it or heard negative opinions about it.

4.3 Age Groups and Citizenship

- The survey data shows that most of the application users are in the range of 25-34 and there are fewer people using it as the age increases. This can be due to lesser technical literacy but it is also a limitation of this study as the survey was distributed online on social media so the target sample would be mostly between the ages of 25-34 [47].

Which age group are you in?	*Do you use the HOIA app? (You have downloaded it and enabled bluetooth)* No	Yes	Grand Total
18-24	10.80%	14.20%	25.00%
25-34	11.93%	26.70%	38.64%
35-44	7.95%	11.93%	19.89%
45-57	4.55%	6.82%	11.36%
58-70	0.57%	3.41%	3.98%
Above 70	0.57%		0.57%
Under 18	0.57%		0.57%
Grand Total	**36.93%**	**63.07%**	**100.00%**

Figure 15. Age group distribution of the survey participants.

- There was no significant difference in the data between people who are citizens of Estonia and expatriates. Both groups had similar percentages when it came to agreeing and disagreeing about allowing location tracking in the application.

Would you be okay with your location data being collected if it helped reduce the spread of the pandemic, knowing that the security of that data cannot be 100% guaranteed?	*Are you a citizen of Estonia?* No	Yes	Grand Total
No	17.98%	19.10%	37.08%
Yes	34.83%	28.09%	62.92%
Grand Total	**52.81%**	**47.19%**	**100.00%**

Figure 16. Correlation between nationality and acceptance of location tracking.

4.4 Application Accuracy

10.1% of participants mentioned that they have been in contact with a COVID-19 positive person who reported their infection status in the HOIA application, but they did not receive an exposure notification. Moreover, one user reported that the application did not work correctly for their friend, as it kept her infectious status and did not change it when she was healthy again, so the application was recording her interactions as COVID-19 positive close interactions when she was nearby people after recovering from the virus. This shows the importance of having a feedback form or a feedback button for the HOIA application, as this feature is not available currently. Having that feature would help the application developers identify these isolated cases more clearly and investigate them properly.

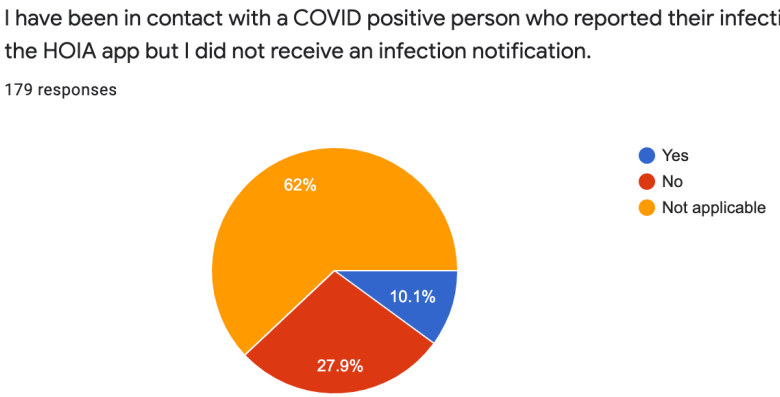

Figure 17. Accuracy of the HOIA application.

4.5 Data Collection

- There were more people aware of what data the HOIA application collects as opposed to other applications on their phones. This may indicate that people are more reserved when it comes to state-owned applications as opposed to private companies. Or it can also mean that people are not willing to give up social media regardless of their privacy policies, so they are more nitpicky when it comes to other applications.

Do you use the HOIA app? (You have downloaded it and enabled bluetooth)	*Do you know what data the HOIA app collects?* No	Yes	Grand Total
No	18.89%	17.78%	36.67%
Yes	19.44%	43.89%	63.33%
Grand Total	**38.33%**	**61.67%**	**100.00%**

Figure 18. Knowledge about HOIA application data collection.

- The survey data revealed that people were informed enough about the data the HOIA application collects as well as generally other applications on their phones. However, it would still be more intuitive if the FAQ section on the HOIA application would be included within the application itself, so the user does not have to visit a separate website to get that information. There was not a significant difference in knowledge about data collection between people working in the tech sector and people working in other sectors.

Do you know what data the HOIA app collects?	*Do you know what data is collected by other apps you already have on your phone?*	
	No	Yes
No	26.82%	11.73%
Yes	21.79%	39.66%

Figure 19. Knowledge about data collection by other applications.

4.6 Evaluation of The Suggested Guidelines

- The survey data shows that 72.3% participants agree that the HOIA application should provide official statistics such as the number of cases, number of recoveries, number of vaccinated people, etc. It is better for users to have a centralized repository for this data rather than relying on different resources for this information. Koroonakaart.ee website already implements great visuals that can be implemented within the application. 9.5% of the participants strongly disagreed, most likely due to this feature increasing the application size, thus, amplifying the storage problem that some people had with the current minimal state of the application.

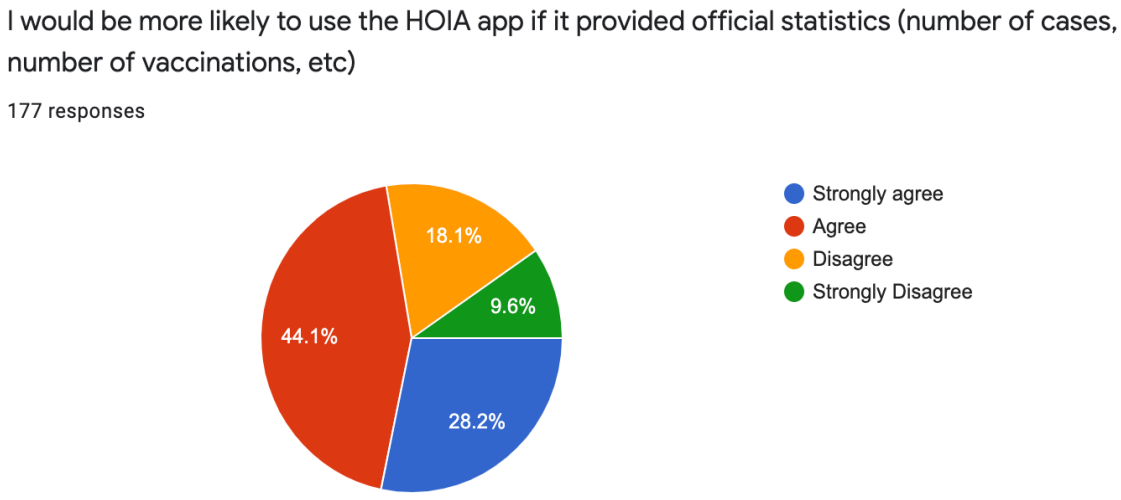

Figure 20. Users' opinion adding statistics feature to the HOIA application.

- The survey data shows that 78.1% of the participants agree with the suggestion that the HOIA application should provide official news regarding the virus such as lockdown extensions, travel restrictions, new rules, etc. This also shows how people want to use it more as an official news source rather than just statistics. This would reduce the effect of fake news and social engineering attacks that can spread on social media, as this application can be used as the single source of truth for governmental announcements regarding the pandemic.

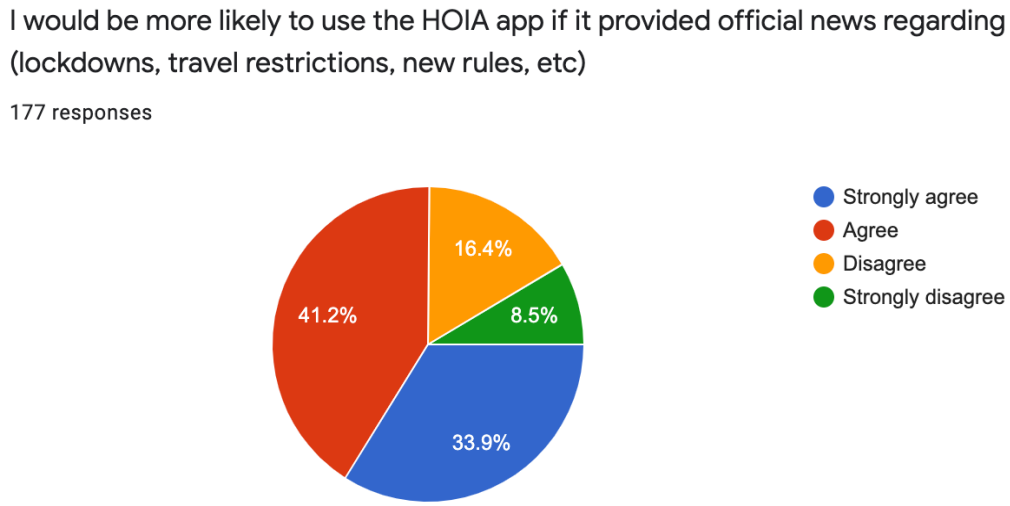

Figure 21. Users' opinion on adding the news feature to the HOIA application.

- 69.9% of the participants agreed on implementing the feature of receiving a weekly report about infection trends, the state of positive cases and vaccinations. This feature would aid in making the application more interactive and for people to find more value in it to keep track of the situation and be fully aware of it. 30.3% of the participants disagreed with this feature probably because of the discomfort that is associated with receiving notifications from this particular application as it can mean that they have been in contact with a COVID-19 positive person. One solution to that is having that report be scheduled at a particular time every weekend so that people are aware that this notification is regarding the report and not a close contact event.

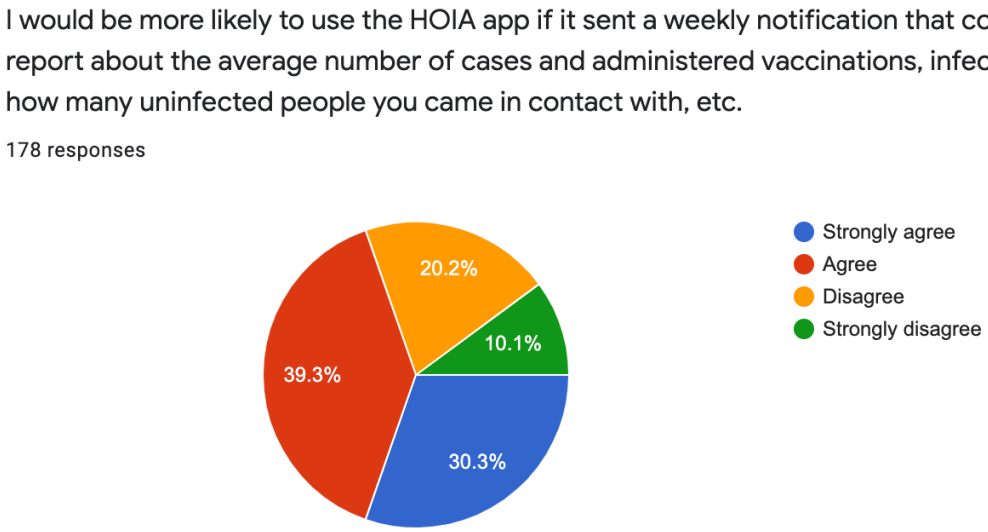

Figure 22. Users' opinion on adding the weekly report feature to the HOIA application.

- 53.1% disagreed on the feature of adding advice on safety measures and emergency numbers to call. This indicates that, by now, people are well-aware of the measures that need to be taken in order to stay safe such as wearing masks, washing hands, etc.

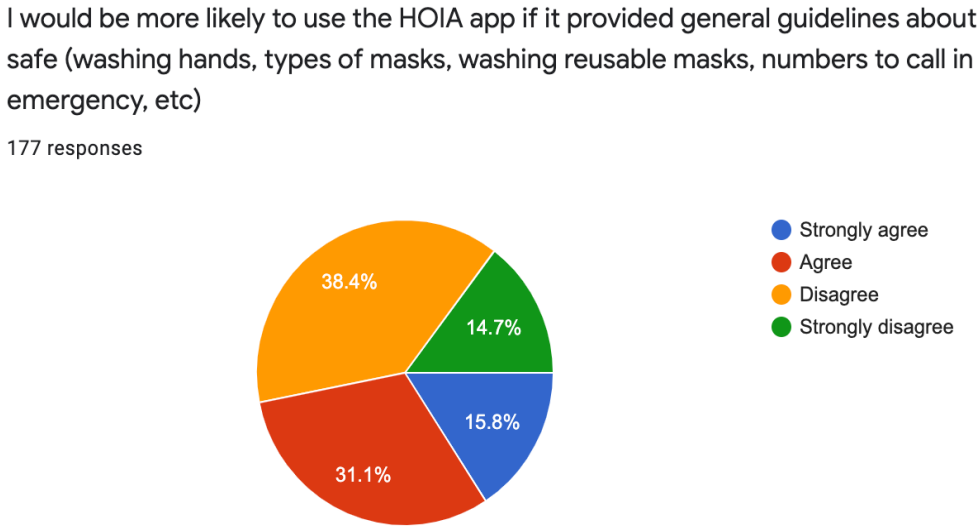

Figure 23. Users' opinion on adding safety guidelines to the HOIA application.

4.7 Location Tracking

- A good example of using location tracking to contain COVID-19 is the South Korean response to the pandemic. They were able to flatten the infection curve without

hindering businesses or people's lifestyles, and that approach received widespread acceptance from their population [48]. The location data would be collected and passed to the National Health Insurance's IT System and then further assessment and review process takes place to identify more contacts and contain the outbreak [49]. However, South Korea has raised some privacy controversies in disclosing personal data within the public sector. One solution to this problem would be to utilize the same technology but provide less comprehensive data. For example, instead of disclosing the places a person has visited to the public, the data can be used only to disinfect the places that the person has been to and contact the people the person got in contact with [50, 51]. Therefore, when implementing location tracking, the developers should still aim for the highest level of privacy possible for the users.

- About 30% of people who said they are using the HOIA application agreed to provide their location data if that helped with identifying cases better, so there is a high potential already with the people who are already okay with using the application.

	Would you be okay with your location data being collected if it helped reduce the spread of the pandemic, knowing that the security of that data cannot be 100% guaranteed?	
Do you use the HOIA app? (You have downloaded it and enabled bluetooth)	No	Yes
No	20.56%	16.11%
Yes	16.67%	46.67%

Figure 24. Users' opinion on location tracking.

- A surprising finding was that 70% of the people who said that they are not using the application because it is not useful have agreed to provide their location data given that it will contribute to controlling the pandemic better in Estonia. This is very significant as it shows there are a lot more people who would use the application if that functionality was provided and used to its maximum potential. More people joining will definitely result in the network effect, where the value of a product (HOIA in this case) would increase as more people in a person's social circle start using it, thus, driving even more people to use it and achieving a much higher population penetration for the application. Further research needs to be done on how to best utilize this effect. For instance, influencers on social media can possibly have a large impact on young people in convincing them to use the application.

If you are not using the app, select your reason	Would you be okay with your location data being collected if it helped reduce the spread of the pandemic, knowing that the security of that data cannot be 100% guaranteed?	
	No	Yes
It is not useful	30.00%	70.00%

Figure 25. Users' opinion on the practicality of the HOIA application.

- One of the most important features that can be implemented if location tracking would be enabled is adding a heatmap that would show the most infectious areas based on user reports and location so that people can avoid those places and reduce the number of infections. A user should receive a notification from the application warning them that they are in a highly infectious area if they are within such locations.

- It is important to keep in mind the difference between privacy and security. Security refers to how your data is protected, while privacy refers to the rights you have over your personal data and how it is handled by others, these rights are specified in privacy policies. People in Estonia, both citizens and international residents, have already shown willingness to provide a lot of data that can be accessed with their permission to a centralized authority through the utilization of their ID card, since this card is connected to your work, residency, medical records, etc. Therefore, it is somewhat possible that implementing location tracking within the HOIA application would not be as outrageous for people in Estonia as it would be for people elsewhere. At the very least, that feature should be provided as an option that can be turned off if users wanted to opt out of it. Enabling location tracking can seem like the harsher measure to take. However, there is the possibility that this feature might have increased the accuracy of the application and the number of people using it, and all the backlash at how privacy-invasive it is might be reversed once it shows that this measure would prevent us from going into a lockdown after lockdown, causing damage to a lot of businesses and our daily lives. Maybe that option, given that it does cause a significant shift in dealing with the pandemic, would have been the lesser of the two evils, between going into lockdowns and tracking people's locations temporarily. This prompts further research in the future about the efficiency and acceptability of this feature, as it can aid us in dealing with future global outbreaks more effectively.

Do you use the HOIA app? (You have downloaded it and enabled bluetooth)	*Would you be okay with your location data being collected if it helped reduce the spread of the pandemic, knowing that the security of that data cannot be 100% guaranteed?*	
	No	Yes
No	20.56%	16.11%
Yes	16.67%	46.67%

Figure 26. Correlation between using the HOIA application and acceptance of location tracking.

5 Proposed Solution

The author proposes the following guidelines based on the analysis of the collected data:

1. The application should provide daily statistics about the status of the pandemic. These statistics can provide official numbers such as new cases, recoveries, deaths, vaccinations. Furthermore, these numbers can be used to formulate different graphs and charts for easy visualization of the data. This feature would give users a reason to check the application on a daily basis and remain vigilant in terms of taking precautions from the virus.

2. The application should provide official news about COVID-19 such as travel limitations, lockdown extensions, new restrictions, governmental announcements, etc. This feature would help in curbing the panic that comes with fake news and turn the application into an official source of truth rather than just a reporting tool. This feature would also help in reducing the effect of social engineering attacks since people would rely on this application as the source of truth instead of fake websites and other types of malicious attacks.

3. The application should provide a weekly report that includes information about COVID-19 infection trends, the weekly average number of infected, recovered and vaccinated people as well as the most notable news from that week. It is important to have this report delivered at a predefined time every weekend to avoid causing users distress about receiving a notification from the application and thinking it might be a close-contact exposure notification.

4. The application should provide an optional location tracking feature. By default, the feature should be turned on. This feature would aid in registering more accurate close-contact information as well as aiding epidemiologists in their research. Most importantly, this feature would help in implementing a heat map functionality that would display the most infectious areas and notify users to take

precautions when they are in proximity to these areas. If the user chooses to disable location tracking, the application should use the Bluetooth decentralized model provided through the GAEN API.

5. The application should provide a feedback page where users can submit reports of malfunctioning. This would greatly help developers in isolating edge cases and debugging the issues more efficiently. Moreover, this page enables users to provide their opinions on how to improve the application and their level of satisfaction with it, which would help developers in identifying areas of improvement more concretely.

6. The application should include the Frequently Asked Questions section that is available on the official HOIA website. This would help inform users about the application when they install it instead of them having to go to a separate website and read the information there.

7. The application should provide the option to update the vaccination status. This would help future researchers in evaluating vaccination efficiency as well as provide more detailed statistics about how many cases, recoveries and deaths have occurred among vaccinated individuals.

8. The application should be more backward compatible with older smartphones. This would ensure that more people use it and thus, leading to larger population coverage and better control of the pandemic.

9. The application should provide the feature of scanning QR codes. These QR codes would be placed at the entrance of social places such as restaurants and bars and must be scanned at the entry of these places. This way, if a person reports their infection after the scan, every person who has been to that place on that day and time would be notified.

6 Limitations and Future Steps

1. Due to time constraints, the author was not able to conduct a qualitative research through focused interviews which would potentially provide more insightful data about the usage of the application and detailed opinions about the suggested guidelines.

2. The author attempted to contact the official HOIA email to discuss the results of the data analysis as well as getting more insight about the future of the application and their opinion about the proposed features. However, the communication was challenging because they took a very long time and replied at a point where the thesis was finalized and submitted. There should be a more active conversation in the future between researchers of this area and the parties who are directly involved in the development of these applications.

3. The survey was conducted on a sample only from the Estonian population. More research should be done in the future to verify that the conclusions reached through this research hold for Coronavirus tracing applications developed in other countries. There is a possibility that more scalability issues would arise in countries with a larger population as that would require a large infrastructure for monitoring and reporting in order for the applications to have a significant effect.

4. The survey was distributed online, which can influence the age groups that the survey reaches. More research should be done to gather data offline from age groups that do not use social media as heavily as others.

7 Conclusion

The goal of this thesis was to propose a set of guidelines and best practices that software engineers and cyber security specialists should follow when developing Coronavirus tracing applications in terms of privacy, security and usability. The author focused on people's perception of privacy and security with regard to these applications. The author used a quantitative survey to collect data about users' opinions on how to improve these applications. The author focused on the case of the HOIA application, the official application for tracking COVID-19 in Estonia. Analysis of the gathered data revealed that around 30% of the survey participants who use the application are willing to provide their location if that improved the accuracy of the application and therefore help contain the virus and save more lives, contrary to the belief that people would value their privacy over everything. The data also reveals that 70% of the people who answered that the HOIA application is not useful are willing to provide their location data if the application required it. Another outcome of the survey was that most participants agree that the HOIA application, similar to the rest of Coronavirus tracing applications, has very minimal content and that they would use it more if it provided more data. We have the necessary technology available at hand, and it only requires the public's willingness to use that technology constructively in outbreak management.

The sampling method used was non-probability voluntary response sampling in Estonia. More research would be required in the future to verify whether these conclusions would hold in other countries, considering that different communities have different privacy standards [52].

This research shows that using the guidelines proposed by the author, such as adding statistics and optional location tracking feature to the application, software engineers and cyber security specialists can ensure that the maximum number of people would use the exposure notification technology and thus, helping with tackling the pandemic more effectively.

References

1. O'Neill P. H., Ryan-Mosley T. A flood of coronavirus apps are tracking us. Now it’s time to keep track of them. [Internet]. 2020 May 7 [cited 2021 March 1]. Available from: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/
2. Symptoms of COVID-19. [Internet]. 2020 May 13 [cited 2021 March 3]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html
3. How tracing and warning apps can help during the pandemic. [Internet]. 2020 October 19 [cited 2021 March 3]. Available from: https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/travel-during-coronavirus-pandemic/how-tracing-and-warning-apps-can-help-during-pandemic_en
4. Zhang B., Kreps S., McMurry N, McCain RM. Americans’ perceptions of privacy and surveillance in the COVID-19 pandemic. [Internet]. 2020 December 23 [cited 2021 March 1] from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0242652
5. Exposure Notifications: Using technology to help public health authorities fight COVID-19. [Internet]. [cited 2021 March 6]. Available from: https://www.google.com/covid19/exposurenotifications/
6. Boyd C. How coronavirus compares to history's deadliest pandemics: Visual timeline pits COVID-19 against Black Death, smallpox and AIDS - as experts warn current crisis could rival Spanish flu 'in its lethality and scale'. [Internet]. 2020 March 17 [cited 2021 March 10]. Available from: https://www.dailymail.co.uk/news/article-8120631/Visual-timeline-coronavirus-compares-historys-deadly-pandemics.html
7. Mouton F., De Coning A. COVID-19: Impact on the cyber security threat landscape. [Internet]. 2020 March [cited 2021 March 10]. Available from: https://www.researchgate.net/publication/340066124_COVID-19_Impact_on_the_Cyber_Security_Threat_Landscape
8. Miles T. West Africa's Ebola outbreak cost $53 billion – study. [Internet]. 2018 October 24 [cited 2021 March 10]. Available from: https://www.reuters.com/article/us-health-ebola-cost-idUSKCN1MY2F8

9. Bezuidenhout M, Mouton F, Venter HS. Social engineering attack detection model: SEADM. [Internet]. 2010 September [cited 2021 March 10]. Available from: https://www.researchgate.net/publication/224178652_Social_engineering_attack_detection_model_SEADM
10. Hadnagy C. Social Engineering: The Science of Human Hacking. Wiley; 2018 June 25 [cited 2021 March 10]
11. Leonhardt M. Coronavirus $1,000 relief check plan not even final yet and experts say fraudsters are already looking to cash in. [Internet]. 2020 March 19 [cited 2021 March 10]. Available from: https://www.cnbc.com/2020/03/19/what-to-know-about-scams-looking-to-cash-in-on-1000-dollar-covid-19-check.html
12. Nelson D. Thieves Swindle $2M From Coronavirus Preppers With Hand Sanitizer, Face Mask Scams. [Internet]. 2020 March 18 [cited 2021 March 10]. Available from: https://www.coindesk.com/thieves-swindle-2m-from-coronavirus-preppers-with-hand-sanitizer-face-mask-scams
13. Grossman J. 259 .COM domain names with the keyword “corona" have been newly registered or updated in the last 24-48 hours. [Internet]. 2020 March 3 [cited 2021 March 10]. Available from: https://twitter.com/jeremiahg/status/1234612630880321537
14. Fowler H., Duncan C. Hackers made their own coronavirus map to spread malware, feds warn. [Internet]. 2020 March 13 [cited 2021 Mar 10]. Available from: https://www.miamiherald.com/news/nation-world/national/article241171546.html
15. Christie M. Online scammers target vulnerable Internet users during coronavirus outbreak. [Internet]. 2020 March 20 [cited 2021 March 10]. Available from: https://abcnews.go.com/US/online-scammers-target-vulnerable-internet-users-coronavirus-outbreak/story?id=69675134
16. Spring M. Coronavirus: The viral rumours that were completely wrong [Internet]. 2020 August 6 [cited 2021 March 10]. Available from: https://www.bbc.com/news/blogs-trending-53640964
17. Department for International Development. UK aid to tackle global spread of coronavirus ‘fake news’. [Internet]. 2020 March 12 [cited 2021 March 10]. Available from: https://www.gov.uk/government/news/uk-aid-to-tackle-global-spread-of-coronavirus-fake-news
18. Government announces further strict measures to curb spread of coronavirus [Internet]. 2020 March 18 [cited 2021 March 11]. Available from:

https://www.power987.co.za/news/government-announces-further-strict-measures-to-curb-spread-of-coronavirus/
19. Edney A. Trump touts drug that FDA says isn't yet approved for COVID-19 [Internet]. 2020 March 19 [cited 2021 March 12]. Available from: https://www.bnnbloomberg.ca/trump-touts-drug-that-fda-says-isn-t-yet-approved-for-covid-19-1.1408984
20. Chin E., Felt A., Sekar V., Wagner D. Measuring user confidence in smartphone security and privacy. [Internet] 2012 July [cited 2021 March 7] Available from: https://www.researchgate.net/publication/254463451_Measuring_user_confidence_in_smartphone_security_and_privacy
21. Pielot M., Church K., De Oliveira R. An in-situ study of mobile phone notifications. [Internet]. 2014 September [cited 2021 March 7]. Available from: https://dl.acm.org/doi/abs/10.1145/2628363.2628364
22. Kallaste K., editor. State not yet willing to give up on coronavirus notification app HOIA. [Internet]. 2021 March 22 [cited 2021 April 1]. Available from: https://news.err.ee/1608151003/state-not-yet-willing-to-give-up-on-coronavirus-notification-app-hoia
23. Allik H-L. Series of flops or how HOIA failed. [Internet]. 2021 February 22 [cited 2021 April 1]. Available from: https://news.postimees.ee/7185647/series-of-flops-or-how-hoia-failed
24. Watkins K. Security and Privacy of COVID-19 Contact-Tracing Apps [Internet]. 2021 March 12 [cited 2021 May 13]. Available from: https://symantec-enterprise-blogs.security.com/blogs/threat-intelligence/security-privacy-covid-tracing-apps
25. Whitman ME., Mattord HJ. Principles of information security. Cengage Learning; 2011 [cited 2021 March 15]
26. Samonas S, Coss D. The CIA Strikes Back: Redefining Confidentiality, Integrity and Availability in Security. Journal of Information System Security; 2014 July 1 [cited 2021 March 15]
27. Patange T. How to defend yourself against MITM or Man-in-the-middle attack. [Internet]. 2013 October 11 [cited 2021 March 15]. Available from: https://web.archive.org/web/20131124235452/http://hackerspace.lifehacker.com/how-to-defend-yourself-against-mitm-or-man-in-the-middl-1461796382

28. What is a denial of service attack (DoS)?. [Internet]. [cited 2021 March 15]. Available from: https://www.paloaltonetworks.com/cyberpedia/what-is-a-denial-of-service-attack-dos
29. Vainio JT. Bluetooth security. [Internet] 2000 May 25 [cited 2021 March 17]. Available from: http://www.yuuhaw.com/bluesec.pdf
30. Biddle S. The Inventors of Bluetooth Say There Could Be Problems Using Their Tech for Coronavirus Contact Tracing. [Internet]. 2020 May 5 [cited 2021 March 24]. Available from: https://theintercept.com/2020/05/05/coronavirus-bluetooth-contact-tracing/
31. Hassan SS., Bibon SD., Hossain MS., Atiquzzaman M. Security threats in Bluetooth technology. [Internet]. 2018 May 1 [cited 2021 March 26]. Available from: https://www.sciencedirect.com/science/article/pii/S0167404817300615
32. Bluetooth Tracing and Covid-19 App: is this privacy safe?. [Internet]. 2020 December 10 [cited 2021 March 28]. Available from: https://www.privacyfoundation.nz/bluetooth-tracing-and-covid-19-app-is-this-privacy-safe/
33. Leith DJ., Farrell S. Measurement-based evaluation of Google/Apple Exposure Notification API for proximity detection in a light-rail tram. [Internet]. Plos One; 2020 September [cited 2021 April 6]. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0239943
34. OWASP Mobile App Security Checklist – English. [Internet]. 2020 May 14 [cited 2021 May 13]. Available from: https://github.com/OWASP/owasp-mstg/blob/master/Checklists/Mobile_App_Security_Checklist-English_1.2.xlsx
35. Kelion L. NHS rejects Apple-Google coronavirus app plan. [Internet]. 2020 April 27 [cited 2021 March 29]. Available from: https://www.bbc.com/news/technology-52441428
36. Vaudenay S. Centralized or decentralized? The contact tracing dilemma. [Internet]. 2020 May 6 [cited 2021 March 29]. Available from: https://infoscience.epfl.ch/record/277809
37. De Montjoye YA., Hidalgo CA., Verleysen M., Blondel VD. Unique in the crowd: The privacy bounds of human mobility. [Internet]. Scientific reports; 2013 March 25 [cited 2021 March 30]. Available from: https://www.nature.com/articles/srep01376
38. Ji S., Li W., Mittal P., Hu X., Beyah R. Secgraph: A uniform and open-source evaluation system for graph data anonymization and de-anonymization. [Internet]. 2015

[cited 2021 March 30]. Available from: https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/ji
39. Sharad K., Danezis G. An automated social graph de-anonymization technique. [Inernet]. 2014 November 3 [cited 2021 March 30]. Available from: https://dl.acm.org/doi/10.1145/2665943.2665960
40. Criddle C., Kelion L. Coronavirus contact-tracing: World split between two types of app. [Internet]. 2020 May 7 [cited 2021 March 30]. Available from: https://www.bbc.com/news/technology-52355028
41. Exposure Notification API launches to support public health agencies. [Internet]. 2020 May 20 [cited 2021 April 4]. Available from: https://blog.google/inside-google/company-announcements/apple-google-exposure-notification-api-launches/
42. Exposure Notifications: Frequently Asked Questions. [Internet]. 2020 September [cited 2021 April 4]. Available from: https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ExposureNotification-FAQv1.2.pdf
43. Exposure Notification: Bluetooth Specification. [Internet]. 2020 April [cited 2021 April 4]. Available from: https://blog.google/documents/70/Exposure_Notification_-_Bluetooth_Specification_v1.2.2.pdf
44. Bradford L, Aboy M, Liddell K. COVID-19 contact tracing apps: a stress test for privacy, the GDPR, and data protection regimes. [Internet]. Journal of Law and the Biosciences; 2020 January 7 [cited 2021 May 13]. Available from: https://academic.oup.com/jlb/article-abstract/7/1/lsaa034/5848138
45. The 4,5, and 7 Point Likert Scale + [Questionnaire Examples]. [Internet]. [cited 2021 March 1]. Available from: https://www.formpl.us/blog/point-likert-scale
46. De Montjoye YA., Houssiau F., Gadotti A., Guepin F. Evaluating COVID-19 contact tracing apps? Here are 8 privacy questions we think you should ask. [Internet]. 2020 April 2 [cited 2021 April 1]. Available from: https://cpg.doc.ic.ac.uk/blog/evaluating-contact-tracing-apps-here-are-8-privacy-questions-we-think-you-should-ask/
47. Tankovska H. Distribution of Facebook users in the United States as of January 2021, by age group [Internet]. 2021 February 1 [cited 2021April 7]. Available from: https://www.statista.com/statistics/187549/facebook-distribution-of-users-age-group-usa/

48. The Blue House Blueprint: How South Korea Contained Its First Coronavirus Outbreak. [Internet]. 2020 March 17 [cited 2021 March 20]. Available from: https://nationalinterest.org/blog/korea-watch/blue-house-blueprint-how-south-korea-contained-its-first-coronavirus-outbreak
49. Lee HK. South Korea's contact tracing sheds light on extensive efforts to slow spread of COVID-19. [Internet]. 2020 December 9 [cited 2021 March 20]. Available from: https://abcnews.go.com/International/south-koreas-contact-tracers-struggle-slow-spread-covid/story?id=74621480
50. Park S., Choi GJ., Ko H. Information Technology–Based Tracing Strategy in Response to COVID-19 in South Korea—Privacy Controversies. [Internet]. 2020 April 23 [cited 2021 March 20]. Available from: https://jamanetwork.com/journals/jama/article-abstract/2765252
51. Ahn NY., Park JE., Lee DH., Hong PC. Balancing personal privacy and public safety during COVID-19: The case of South Korea. IEEE Access; 2020 Sep 22 [cited 2021 March 20]. Available from: https://ieeexplore.ieee.org/abstract/document/9203800
52. Stranieri S. Global Data Privacy Laws: US, EU, China And More. [Internet]. 2019 August 20 [cited 2021 April 11]. Available from: https://blog.ipswitch.com/global-data-privacy-laws-us-eu-china-and-more

Appendix 1 – Thesis Survey Questions

The following are the questions and answer options formulated by the author in the survey he conducted to gather the necessary data for the thesis:

1. Do you use the HOIA app? (You have downloaded it and enabled bluetooth)
    a. Yes
    b. No
2. If you are not using the app, select your reason:
    a. Privacy Concerns
    b. Battery usage
    c. It is not useful
    d. I already had the infection
    e. Other
3. Do you know what data the HOIA app collects?
    a. Yes
    b. No
4. Do you know what data is collected by other apps you already have on your phone?
    a. Yes
    b. No
5. Do you think the HOIA app is effective in controlling the pandemic in Estonia?

a. Yes

b. No

6. Do you think the HOIA app is secure?

    a. Yes

    b. No

7. I would be more likely to use the HOIA app if it provided official statistics (number of cases, number of vaccinations, etc)

    a. Strongly agree

    b. Agree

    c. Disagree

    d. Strongly disagree

8. I would be more likely to use the HOIA app if it provided official news regarding the virus (lockdowns, travel restrictions, new rules, etc)

    a. Strongly agree

    b. Agree

    c. Disagree

    d. Strongly disagree

9. I would be more likely to use the HOIA app if it sent a weekly notification that contained a report about the average number of cases and administered vaccinations, infection trends, how many uninfected people you came in contact with, etc.

    a. Strongly agree

    b. Agree

c. Disagree

d. Strongly disagree

10. I would be more likely to use the HOIA app if it provided general guidelines about how to stay safe (washing hands, types of masks, washing reusable masks, numbers to call in case of emergency, etc)

    a. Strongly agree

    b. Agree

    c. Disagree

    d. Strongly disagree

11. Would you quarantine if the you received a notification that you have been in close contact with a COVID positive person?

    a. Yes

    b. No

12. Would you be okay with your location data being collected if it helped reduce the spread of the pandemic, knowing that the security of that data cannot be 100% guaranteed?

    a. Yes

    b. No

13. I have been in contact with a COVID positive person who reported their infection status in the HOIA app but I did not receive an infection notification.

    a. Yes

    b. No

    c. Not applicable

14. Do you think using the HOIA app should be mandatory?

    a. Yes

    b. No

15. Which age group are you in?

    a. Under 18

    b. 18-24

    c. 25-34

    d. 35-44

    e. 45-57

    f. 58-70

    g. Above 70

16. Do you study or work in the tech industry?

    a. Yes

    b. No

17. Are you a citizen of Estonia?

    a. Yes

    b. No

Appendix 2 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I, Hayder Hasan Ali Al-Sabti

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis “Guidelines for Developing Coronavirus Tracing Applications: the Case of HOIA”, supervised by Kaido Kikkas.
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

17.05.2021

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.