

# Entwurf und Analyse von Crowd-Sensing Mechanismen für mobile Corona-Warn-Applikationen

Gabriel Bonnet

*Bsc. Science Softwaretechnik  
Universität Stuttgart  
Stuttgart, Germany*

Kai Braun

*Bsc. Science Softwaretechnik  
Universität Stuttgart  
Stuttgart, Germany*

Hannes Deichmann

*Bsc. Science Softwaretechnik  
Universität Stuttgart  
Stuttgart, Germany*

Timm Marquardt

*Bsc. Science Informatik  
Universität Stuttgart  
Stuttgart, Germany*

**Zusammenfassung**—Das Coronavirus sorgt nun seit Anfang 2020 für sowohl in Deutschland als auch weltweit für Ausnahmezustände. Um die Ausbreitung des Virus weitestmöglich einzudämmen, wurden vielerlei Maßnahmen und Vorkehrungen getroffen, wie etwa Kontaktbeschränkungen oder Lock-downs. Eine der effektivsten Methoden, die Infektionsketten zu unterbrechen und die Ausbreitung zu verlangsamen, ist die Kontaktverfolgung. Dabei werden von einem Infizierten alle Kontaktpersonen innerhalb der letzten 14 Tage ermittelt, welche sich dann in Isolation begeben sollten. Da die manuelle Kontaktverfolgung ein aufwändiges Unterfangen darstellt wurde im Auftrag des RKI die Corona Warn App entwickelt, die diese Kontaktverfolgung per Smartphone automatisieren soll. Die Corona Warn App stand seit ihrer Herausgabe jedoch häufig unter Kritik und scheint in vielerlei Hinsicht noch nicht komplett ausgereift. Unser Ziel ist es, mögliche Erweiterungen der Corona Warn App zu finden und zu analysieren. Hierbei wird ein besonderer Fokus auf die Crowd-Sensing Mechanismen gelegt, welche die Corona Warn App vorzuweisen hat. Die Erweiterungen wurden unterteilt in die Nebenfeatures, sowie die zwei Hauptfeatures der *Erweiterten Risikoberechnung* und der *Density-Map*. Jede vorgeschlagene Erweiterung wird anschließend diese Erweiterungen unter verschiedenen Aspekten beurteilt.

**Index Terms**—Corona, Corona-Warn-App, Density Map, Risikoberechnung

## 1. Einleitung

Seit nun mehr als eineinhalb Jahren breitet sich das Corona Virus auf der Erde aus und zwingt die Weltbevölkerung in bis dahin noch nicht erlebte Ausnahmezustände. Laut derzeitigem Stand wurden etwa 173 Millionen mit dem Virus infiziert und rund 3,7 Millionen sind an den Folgen des Virus gestorben [1], wobei die Dunkelziffer weitaus höher Werte vermuten lässt. In Deutschland wurden dem Robert Koch Institut insgesamt 3,7 Millionen Infizierte und etwa 90.000 Todesopfer gemeldet [2]. Insbesondere stellen dabei das extrem hohe Ansteckungsrisiko und der lange infektiöse Zeitraum von 14 Tagen, in denen ein Infizierter andere anstecken kann, ohne selber Symptome zu zeigen, eine große Herausforderung dar.

Um die Ausbreitung des Virus weitmöglichst einzudämmen, wurden im Land zahlreiche Vorkehrungen, wie etwa Kontaktbeschränkungen, Maskenpflicht oder Lock-downs, getroffen. Mittlerweile befinden sich auch verschiedene Impfstoffe gegen das Virus im Umlauf, welche eine baldige Rückkehr zur Normalität erhoffen. Vor der Erforschung und Herstellung der Impfstoffe gilt jedoch die Kontaktverfolgung und Beschränkung als effektivste Methode das Virus einzudämmen. Hierbei wird bei einer infizierten Person versucht, alle Kontakte und damit potenzielle angesteckte Personen nachzuverfolgen und zu isolieren, damit die Infektionsketten des Virus unterbrochen werden. So ist man beispielsweise verpflichtet, in Restaurants oder Bars seine Kontaktdaten zu hinterlegen, damit man im Falle einer potenziellen Begegnung mit einem Infizierten in der Einrichtung benachrichtigt werden kann. Eine solche manuelle Kontaktverfolgung kann sich jedoch als ein sehr schwieriges und aufwändiges Unterfangen herausstellen, da man sich alle Kontakte innerhalb des langen Zeitraumes von zwei Wochen merken und wieder ins Gedächtnis rufen muss. Aus diesem Grund wurde vom Robert Koch Institut eine App entwickelt, die diese Kontaktverfolgung automatisieren sollte: Die Corona Warn App [3]. Diese soll dabei helfen, die Kontaktverfolgung zu automatisieren und dabei effektiver und effizienter vorgehen als die manuelle Methode. Ein zentraler Bestandspunkt der App ist die Privatsphäre, die mittels dezentraler Kontaktverfolgung und anonymen Datenaustausches garantiert ist.

Seit ihrer Herausgabe gab es viele Diskussionen und Kritik rund um die App. So stand sie hinsichtlich des Datenschutzes öfters in der Kritik und musste noch während ihrer Entwicklung nachgebessert werden [4]. Andere Stimmen behaupteten, die App sei überhaupt nicht effektiv und würde nicht nennenswert zur Eindämmung der Pandemie beitragen, da zu wenige Leute die App nutzen [5]. Mittlerweile sind die Stimmen der Kritiker leiser geworden und viele der Bedenken, die von der Bevölkerung geäußert werden (wie etwa, dass die App uns ausspionierte) stellten sich als Mythen heraus. Trotzdem scheint es so, als sei die Corona Warn App nicht der Erfolg, den sich die Bundesregierung erhofft hatte und es gäbe noch Lücken und Erweiterungsbedarf für eine wirkungsvolle und ausgereifte App.

Das Ziel unserer Arbeit ist es, Erweite-

rungsmöglichkeiten der bisherigen Corona Warn App zu finden und diese zu beurteilen. Einen besonderen Fokus legen wir hierbei auf die Crowd-Sensing Mechanismen der Corona Warn App, da hier von einer großen Menge an Nutzern über die Smartphones anonymisierte Kontaktdaten untereinander ausgetauscht und verarbeitet werden. Das Ziel des Crowd-Sensings ist es, ein großes Sensornetz mit bereits vorhandenen Geräten (wie beispielsweise Smartphones) aufzubauen und gesammelte Daten zentral verarbeiten zu können. Dies kann in kürzester Zeit ohne teure Infrastruktur realisiert werden, da Smartphones bereits flächendeckend vorhanden sind. Die Probleme der Crowd-Sensing Methode, wie die Integrität und Privatheit der Daten, wurden in der Corona Warn App gezielt aufgegriffen und behandelt durch anonymen Datenaustausch per Bluetooth Low Energy und von Teststellen verifizierte positive Testergebnisse.

Unsere Erweiterungen sollen zum einen dazu beitragen, die Ausbreitung des Coronavirus noch effektiver einzudämmen, zum anderen aber auch weitere nützliche Features bieten, die im Hinblick auf das Virus und für die Nutzer interessant und hilfreich wären. Dabei sind wir folgendermaßen vorgegangen: Zunächst wurde eine Vielzahl aller möglichen funktionalen Erweiterungen identifiziert und eine Übersicht erstellt. Daraufhin wurden zwei vielversprechende Features herausgegriffen: Die **erweiterte Risikoberechnung** und die **Density-Map**.

Bei der Erweiterten Risikoberechnung wird die aktuelle Risikoberechnung der Corona Warn App um zwei weitere Summanden erweitert und feingranularer in der Nutzoberfläche dargestellt. Dabei soll neben dem Kontakt mit infizierten Personen, welche derzeit ausschließlich betrachtet werden, auch Begegnungen mit allen anderen Personengruppen berücksichtigt werden.

Die Density Map hingegen soll eine grafische Übersicht darüber geben, an welchen Orten sich derzeit eine Vielzahl von Menschen aufhalten, damit diese gezielt gemieden werden können. Hierbei wurden verschiedene Lösungsansätze zur Umsetzung ausgewählt, ohne dabei die bestehenden Sicherheits- und Privatsphärestandards der bestehenden Corona Warn App zu verletzen.

Diese beiden Erweiterungen werden zusammen mit einer Reihe weiterer Nebenfeatures ausgearbeitet und hinsichtlich Sicherheit, Privatsphäre, Nutzen und Aufwand bewertet.

Der Rest der Arbeit ist folgendermaßen aufgebaut: In Abschnitt 2 befinden sich bestehende Arbeiten. Insbesondere gehen wir genauer auf die Funktion der bisherigen Corona Warn App ein. In Abschnitt 3 werden dann eine Reihe von Nebenfeatures beschrieben und beurteilt, bevor in Abschnitt 4 und 5 jeweils die beiden Hauptfeatures der *erweiterten Risikoberechnung* und der *Density-Map* im Genaueren betrachtet werden. In Abschnitt 6 befindet sich unsere abschließende Zusammenfassung und ein Ausblick.

## 2. Bestehende Arbeiten

In diesem Abschnitt wird zunächst auf die Corona Lage, sowie auf die Funktionsweise der aktuellen Corona Warn App, eingegangen, auf der unsere Arbeit basiert. Des

weiteren werden bestehende Arbeiten in Richtung Crowd-Sensing, insbesondere mittels des BLE Features, untersucht.

### 2.1. Corona Lage

Bei Corona oder COVID-19 handelt es sich um eine Krankheit, die durch die Coronaviren SARS-CoV-2 ausgelöst wird. Anfang 2020 waren diese Viren verantwortlich für die darauf folgende, weltweite Corona Pandemie, die aktuell zu ca. 3 500 000 Toten weltweit geführt hat. Corona kann symptomatisch sowie asymptomatisch auftreten, wobei Husten und Atemprobleme zu typischen Symptomen zählen. Übertragen wird Corona durch Schleimhäute, die durch Tröpfcheninfektion oder Aerosole gereizt werden können. Bei der Tröpfcheninfektion handelt es sich um Infektionen auf kurze Distanz. Durch zum Beispiel Niesen werden Tröpfchen in die Luft geschleudert, die innerhalb von 1-2 Metern auf den Boden sinken und in diesem Radius auch Infektionen auslösen können. Die Aerosole dagegen sind leichter und können somit länger in der Luft schweben, was dazu führt, dass in geschlossenen Räumen eine Ansteckungsgefahr bis zu vier Meter um die infektiöse Person besteht. Der Ansteckungszeitraum beträgt 10-20 Tage und sorgt dafür, dass potenziell infektiöse Personen bis zu 14 Tage in Quarantäne kommen. Mittlerweile wurden Impfstoffe gegen das Virus entwickelt und weltweit 1,4 Milliarden Impfdosen verabreicht, um das Coronavirus einzudämmen.

Die Pandemie schränkt das Alltagsleben der gesamten Bevölkerung ein. Durch verschiedene Lockdowns war es nicht mehr möglich, nach 20 Uhr das Haus ohne driftigen Grund zu verlassen und sich im öffentlichen Raum ohne Maske aufzuhalten. Aufgrund der hohen Übertragungsrate ist es wichtig, Kontakte nachzuverfolgen und Infektionsketten zu unterbrechen. Nicht nur um die Anzahl Toter zu verringern, sondern auch um die Rückkehr in den Alltag zu ermöglichen, sollte Corona eingedämmt werden. Aus diesen Gründen wurde nach einer Corona Warn App verlangt, die ermöglichen soll, dass Kontakte automatisch nachverfolgt werden können und gefährdete Personen gewarnt werden.

### 2.2. Corona Warn App

Die deutsche *Corona Warn App* (CWA) wurde im Auftrag der Bundesrepublik Deutschland von SAP, der deutschen Telekom und weiteren Unternehmen entwickelt [3]. Ihr Hauptziel ist die automatisierte und anonyme Kontaktverfolgung mittels Smartphones, um möglichst viele Infektionsketten zu unterbrechen und die Ausbreitung des Coronavirus weitmöglichst zu verlangsamen. Zum aktuellen Zeitpunkt ist die App seit etwa einem dreiviertel Jahr aktiv und wurde knapp 30 Millionen Mal in den App Stores von Apple und Google heruntergeladen (Stand Mai 2021) [25].

Die Funktion der Corona Warn App ist aus Nutzersicht relativ simpel. Der Nutzer lädt sich die App auf sein Smartphone und aktiviert seine Bluetooth-Funktion. Falls er nun mit einer weiteren Person, die ebenfalls die CWA installiert hat, in Kontakt kommt, werden anonym Daten ausgetauscht

und die Begegnung auf den jeweiligen Endgeräten gespeichert. Sofern man sich nun mit dem Coronavirus infiziert und ein positives Testergebnis erhält, hat man die Option das positive Testergebnis in die App einzutragen. Alle Kontaktpersonen werden daraufhin informiert, ohne zu wissen, von wem diese Mitteilung letztendlich stammt.

Die Corona Warn App unterscheidet prinzipiell zwischen zwei Risikostufen, die mithilfe eines grünen bzw. roten Bildschirms auf der Startseite der App angezeigt werden. Ein grüner Bildschirm bedeutet hierbei keine/kaum Risiko-begegnungen und damit ein vermindertes Risiko, während ein roter Bildschirm ein erhöhtes Risiko anzeigt.

Der Corona Warn App zugrunde liegendes Feature ist das „Exposure Notification Framework“, welches in Kooperation von Google und Apple für Kontaktverfolgungs-Apps entwickelt wurde und das Bluetooth Low Energy (BLE) Feature nutzt, um die Begegnungsdaten der Smartphones anonym untereinander auszutauschen.

### 2.3. Exposure Notification Framework:

Das Exposure Notification Framework basiert auf dem *Decentralized Privacy-Preserving Proximity Tracing* (DP-3T) Protokoll [7]. Jeden Tag wird vom Framework ein sogenannter *Temporary Exposure Key* (TEK) generiert. Dieser hält 24 Stunden und ist ein zufällig ausgewählter Schlüssel. Aus diesem TEK werden *Rolling Proximity Identifier* (RPIs) abgeleitet. Diese RPIs haben eine Lebensdauer von 10-15 Minuten, bevor sie gelöscht werden und ein neuer RPI aus dem TEK erstellt wird. Wichtig ist hierbei, dass sich aus den RPIs nicht der TEK ableiten lässt. Die RPIs werden nun unter den Smartphones im Falle einer Begegnung mithilfe von BLE ausgetauscht. Neben den RPIs werden zudem verschlüsselt Informationen über die Begegnungsdauer, sowie dem ungefähren Abstand ausgetauscht. Falls sich nun eine Person mit dem Coronavirus infiziert, werden dessen TEKs (nun *Diagnosis Keys* genannt) der letzten 14 Tage auf einen zentralen Server hochgeladen. Dieser zentrale Server wird in diesem Dokument als *CWA Server* bezeichnet. Jedes Smartphone hat Zugriff auf den CWA Server und kann die Diagnosis Keys aller infizierten Personen herunterladen. Mithilfe der Diagnosis Keys können dann die gespeicherten RPIs aus den Begegnungen abgeglichen und, im Falle einer Übereinstimmung, die Begegnungsdauer und der Abstand entschlüsselt werden. Zusammen mit dem *Transmission Risk Level* (TRL) bestimmt die Corona Warn App aus diesen Informationen einen *Risikowert*, der das Infektionsrisiko der Person widerspiegelt.

### 2.4. Transmission Risk Level (TRL):

Das Transmission Risk Level ist ein Maß dafür, wie infektiös eine erkrankte Person ist. Je höher das TRL ist, desto wahrscheinlicher ist es, sich von dieser Person mit dem Virus anzustecken. Das TRL wird in der CWA lokal berechnet und auf dem CWA Server zu jedem Diagnosis Key abgespeichert, um diesem ein Übertragungs Risiko (TRL) zuzuord-

nen. Dieses Risiko Level beruht auf der Annahme, dass eine infizierte Person nicht über den gesamten Krankheitsverlauf das gleiche Übertragungspotenzial besitzt. Dieses ist anfangs niedrig, steigt dann auf ein Maximum an und nimmt innerhalb von 14 Tagen wieder ab. Um daraus ein Risiko Level abzuleiten, müssen verschiedene Zeitpunkte bekannt sein. Darunter zählen der Zeitpunkt der Ansteckung, der Symptombeginn, das Infektiossein und drei weitere Zeitpunkte, die beim Testen aufgezeichnet werden müssen: Zeitpunkt des Testens, Zeitpunkt des Ergebnis bekommen und Zeitpunkt des Hochladens. Aus diesen Werten ist es möglich den TRL so zu bestimmen, dass er das Übertragungspotenzial widerspiegelt [32]. Allerdings werden diese Zeitpunkte der aktuellen CWA noch nicht bereitgestellt und es wird somit auf einen Standardwert zurückgegriffen. Deshalb wird in Abschnitt 4.3 das TRL Update Feature vorgestellt, dass eine Lösung zeigt, diese fehlenden Daten zu sammeln.

### 2.5. Crowd-Sensing und Bluetooth Low Energy

Die allermeisten technischen Geräte besitzen in der heutigen Zeit die Möglichkeit, sich mit dem Internet zu verbinden und Daten untereinander auszutauschen. Seien es Smartphones, Laptops oder weitere IoT-Geräte, sie alle besitzen die Funktion die Umgebungsdaten mittels Sensoren zu sammeln und diese Daten über das Internet weiterzugeben. Mobile Crowd-Sensing beschreibt den Vorgang, ein großes System solcher mobilen Endgeräte aufzubauen, um Sensordaten im großen Maßstab zu sammeln und auswerten zu können [8]. Die Vorteile des Crowd-Sensings liegen auf der Hand: Anstatt Daten im großen Stile aufwändig über ein zentrales Gerät zu sammeln kann die Vielzahl an bereits existierenden externen Geräten genutzt werden, um diese Daten dezentral zu sammeln und zu verarbeiten. Somit muss lediglich die Auswertung der Daten auf einem zentralen Server stattfinden, zu dem die einzelnen Geräte ihre Daten senden.

Bezüglich Crowd-Sensing existieren bereits umfangreiche Arbeiten aus verschiedensten Anwendungsbereichen. So werden beispielsweise Smartphones verwendet, um die Luftverschmutzung in verschiedenen Stadtteilen von Großstädten oder in Gebäuden zu messen [9]. Die standortbasierten Daten werden dann auf einen Server geladen, damit Nutzer die aktuelle Belastung in den Regionen (oder auch Räumen in Gebäuden) abfragen können. Eine ähnliche Umsetzung gibt es hinsichtlich der Lärmbelastung [10].

Im Verkehrssektor existieren ebenfalls eine Vielzahl von Lösungen, um mithilfe von Crowd-Sensing Verkehrsdaten zu sammeln, analysieren und die Lage zu verbessern. Saremi et al. [11] benutzen unter anderem GPS Daten von Smartphones oder Fahrzeugen, um Informationen über die Existenz von Ampeln an Kreuzungen zu extrahieren.

Anhand dieser Beispiele lassen sich auch die Probleme des Crowd-Sensings herauskristallisieren. Wie etwa wird die Integrität der gegebenen Daten sichergestellt? Für das Beispiel der Luftverschmutzung könnte beispielsweise der Fall auftreten, dass ein Smartphone bewusst falsche Messdaten an den zentralen Server sendet, ohne dass dieser

die Inkorrektheit der Daten feststellen kann. Die Corona Warn App stellt die Datenintegrität dadurch her, dass Nutzer einen vom Gesundheitsamt bereitgestellten Code eingeben müssen, damit man als Infizierter verifiziert ist [3]. Ein weiteres Problem stellt die Privatheit der Daten dar. Werden beim Crowd-Sensing sensible Nutzerdaten, wie der Standort mitgeliefert, könnte damit unter Umständen eine Menge von Nutzerdaten auf einem zentralen Server gesammelt und missbraucht werden. Auch in diesem Fall bietet die Corona Warn App Schutz, indem nur anonyme TEKS auf den CWA Server geladen werden und sensiblere Daten, wie die Begegnungsdauer und Entfernung dezentral und verschlüsselt auf den einzelnen Smartphones vorliegen.

## 2.6. BLE Crowd-Sensing:

Auch im Bereich BLE existieren Crowd-Sensing Anwendungen zum Erfassen von Daten und Erstellen von Density-Maps. Tschuhiya et al. [12] beschäftigen sich mit der allgemeinen Nutzbarkeit des BLE Features zum Erstellen von Maps. Anstelle von Smartphones werden hier Arduinos verwendet, um verschiedene Sensordaten, wie Druck, Luftfeuchtigkeit, Lichtintensität usw. zu messen. Diese Daten werden dann per BLE an die Smartphones weitergegeben, wo sie ausgewertet und dann auf einen Server hochgeladen wurden.

In [13] werden Smartphones und vorinstallierte Raspberry Pies mit CICN (Community Information-Centric Networking) verwendet, um eine IoT Sensoren Map innerhalb geschlossener Gebäude, wie etwa Einkaufszentren, zu erstellen. Die Raspberry Pies werden in den einzelnen Räumen vorinstalliert und messen Sensordaten, wie Temperatur oder Luftfeuchtigkeit. Außerdem werden sie mit BLE Beacons versehen, damit Nutzer per Smartphone automatisch die Daten auswerten und eine Übersichtskarte des Gebäudes erstellen können.

Basalamah [14] beschäftigt sich mit Bluetooth Low Energy Tags zum Aufzeichnen von Bewegungsströmen größerer Menschenmassen. Die BLE Tags stoßen regelmäßig Signale aus, die sich mithilfe von Smartphones empfangen lassen. Dadurch kann der Standort und Status der einzelnen Tags gezielt nachverfolgt werden. Im Versuch wurden Pilger beim Haddsch (Pilgerfahrt nach Mekka) mit diesen BLE Tags ausgestattet und deren Bewegung mittels Smartphones, welche an bestimmten Orten als Aufzeichnungsmast agierten, nachverfolgt. An diesem Ansatz ist unsere Density-Map Lösung mittels zentraler Masten stark angelehnt, jedoch bestand die Idee hier mittels der Smartphones gezielt einzelne Tags zu identifizieren und daraus folgend die Bewegungsströme der Pilger nachzuvollziehen. Unser Ansatz hingegen ist es, per BLE möglichst anonym die Daten zu erfassen und damit eine Density-Map zu erstellen. Außerdem verwenden wir keine speziellen BLE Tags sondern die Smartphones selbst zum Datenaustausch.

Ein Feature, das unserer Density-Map am nächsten kommt sind vermutlich die Live-Stoßzeiten des Unternehmens Google [15]. In dessen Suchmaschine lässt sich zu fast jeder öffentlichen Einrichtung, sowie Restaurants oder

Einzelhandel Läden neben den Öffnungszeiten die aktuelle Besucherzahl anzeigen. Zudem wird einem noch die durchschnittliche Aufenthaltsdauer von Personen, sowie eine Abschätzung der Auslastung über den Tag verteilt, mitgeteilt. Die Besucherzahlen werden nicht als konkreter numerischer Wert angegeben, sondern mithilfe von Balkendiagrammen grafisch angezeigt, sodass man lediglich die Auslastung mit anderen Uhr-/Tageszeiten vergleichen kann, ohne die genauen Zahlen zu kennen. Das Live Stoßzeiten Feature wird jedoch nur als Diagramm für jede Einrichtung und nicht als Karte realisiert. Für die Umsetzung ihres Features benutzt Google voraussichtlich die personalisierten Nutzerdaten seiner User, wie beispielsweise den Standortverlauf der Google Maps Anwendung [17]. Auf deren Webseite gibt Google keine weiteren Informationen darüber preis, lediglich wie das Feature funktioniert.

Zusammenfassend lässt sich sagen, dass es eine Fülle von Arbeiten gibt, welche sich mit dem Thema Crowd-Sensing beschäftigen. Die meisten verwenden jedoch personalisierte Daten der Nutzer, wie den Standort, um entsprechende Karten zu generieren. Location-Privacy ist ein aktuelles Forschungsgebiet und es existieren eine Reihe von Lösungen, die dieses Problem behandeln [?]. Die Corona Warn App hingegen verfolgt den Ansatz der kompletten Pseudonymisierung der Daten, bei dem der Nutzer zu jeder Zeit hinter einem zufällig generierten RPI verborgen ist. Auch bei der Nutzung von BLE wurden sensible Daten gezielt mitgegeben, um konkrete Standorte nachzuverfolgen, anstatt die Anonymität des BLE Features gezielt auszunutzen. Das Erstellen einer Density-Map mittels anonymer Nutzerdaten und BLE scheint es nach unserem Wissensstand demnach nicht zu geben.

## 3. Lösungsvorschläge

In diesem Abschnitt werden die Erweiterungen der Corona Warn App präsentiert, die im Verlauf der Forschungsarbeit von uns entwickelt und analysiert wurden. Ausgenommen sind die beiden Hauptfeatures **erweiterte Risikoberechnung** und **Density-Map**, welche separat in Abschnitt 4 und 5 beschrieben sind. Da das Thema Corona derzeit hochaktuell und präsent ist, wurden einige der vorgestellten Erweiterungen im Verlauf unserer Arbeit von anderen Parteien bereits in ähnlicher Form umgesetzt. Auch die Corona Warn App befindet sich kontinuierlich in Entwicklung, wodurch die von uns hier vorgeschlagenen Features in der aktuellen Version der App bereits vorhanden sind oder in naher Zukunft erscheinen sollen.

### 3.1. Bewertungskriterien

Um einen möglichst einheitlichen Vergleich der Features zu garantieren, haben wir sechs Bereiche festgelegt, in denen die Features beschrieben werden. Diese Bereiche sind wie folgt aufgeteilt:

*Privatsphäre:* Wie stark greift das Feature in die Privatsphäre ein?

*Nutzen:* Welchen effektiven Nutzen bringt das Feature und

lohnt sich der Aufwand für den effektiven Nutzen?

*Umsetzbarkeit:* Wie realistisch ist es, dieses Feature zu implementieren in Anbetracht von Datenschutz, Programmierung oder anderen Hindernissen?

*Akzeptanz:* Da eine möglichst hohe Nutzung der App in der Bevölkerung erzielt werden soll, ist es wichtig, dass das Feature entweder einen Anreiz zur Nutzung schafft oder zumindest diesen nicht schwächt. Die Frage ist also, ob die neue Funktion Skepsis oder Missbilligung beim Nutzer auslöst.

*Sicherheit:* Wie gut lässt sich das implementierte Feature vor Missbrauch und Angriffen schützen?

*Daten:* Welche Daten werden mindestens benötigt, dass das Feature funktioniert?

## 3.2. Übersicht aller Erweiterungen

Im Folgenden werden Erweiterungen beschrieben, welche als Lösungen für bestehende Probleme dienen und als Feature in die Corona Warn App implementiert werden könnten oder dies mittlerweile bereits sind. Sie werden anhand unserer in Abschnitt 3.1 genannten Kriterien erläutert und spezifiziert.

**3.2.1. Impfpass.** In der aktuellen Situation ist es wichtig, nachweisen zu können, dass man geimpft wurde. Dies könnte man entweder dadurch regeln, einen analogen Impfnachweis immer bei sich zu tragen, oder durch einen digitalen Impfpass in der Corona-Warn-App.

Größter Vorteil eines in der App integrierten Nachweises wäre die Digitalisierung des Verwaltungswesens und dadurch eine Vereinfachung des Organisationsaufwands. Anstatt mehrere Unterlagen mit sich herumzutragen, hat man diese stattdessen gesammelt in einer App.

Digitalisierung ist zudem ein positiv konnotierter Begriff, welcher einen Anreiz für die Bevölkerung schafft, die App zu nutzen. Dies ist von großer Relevanz, da es das Ziel ist, dass die Corona-Warn-App von 100% der Bevölkerung genutzt wird.

Implementiert man zusätzlich dazu eine Art Krankenakte, in der man Merkmale wie Gewicht und Vorerkrankungen einträgt, ließe sich dadurch die Impfpriorisierung und die Reihenfolge in der Bevölkerung besser organisieren. Benötigte Daten für einen generellen Impfpass wären:

- Name
- Krankenkasse
- Durchgeführte Impfungen

Für die Einschätzung der Impfpriorisierung wäre zusätzlich erforderlich:

- Alter
- Vorerkrankungen
- Gewicht
- Beruf/Arbeitsumfeld

Bei einem unverbindlichen Impfpass wären diese Angaben nicht gesichert, die Integrität der Daten also nicht gewährleistet. Dies resultiert darin, dass der Impfpass nicht von offiziellen Einrichtungen genutzt werden könnte, sondern lediglich vom Nutzer selbst.

Stattdessen müsste man, um die Daten als integer zu garantieren, die Bearbeitung des Impfpass beschränken, sodass nur offizielle Stellen wie Krankenhäuser oder Arztpraxen diesen bearbeiten können.

Somit wird die Integrität der Daten gewährt, welche dann für das Krankwesen genutzt werden können.

Im Verlauf dieser Arbeit wurde ein Teil dieses Feature in der Corona-Warn-App implementiert [31], allerdings nur für die Impfung gegen Corona. Impfungen gegen andere Krankheiten werden noch nicht berücksichtigt.

Der gewählte Ansatz verfolgt eine beschränkte Bearbeitung des Impfnachweises, sodass der Nachweis als offizielles Dokument für z.B. Reisen in andere Länder benutzt werden kann.

**3.2.2. Gamification.** Eine Möglichkeit, das Problem einer zu niedrigen Nutzung der App zu adressieren, ist Gamification. Dadurch versuchen wir, den Anreiz zu steigern, die App herunterzuladen und auch aktiv zu benutzen.

Gamification der App kann es in verschiedenen Varianten geben. Zum Beispiel lassen sich einfache Spielmechaniken wie persönliche Highscores einbauen, welche man mit anderen Nutzern der App vergleichen kann. Dafür bieten sich Statistiken an, die sich bereits aus der Nutzung der App erstellen lassen. Dazu zählen die betroffenen Personen an einem Tag, die durchschnittliche Dauer einer Begegnung oder der ermittelte Risikofaktor. Ziel der Highscores wäre es dann, diese Werte so möglich wie gering zu halten oder im Vergleich zu anderen Nutzern bessere Werte zu haben.

Vorstellbar sind auch daraus resultierende optische Veränderungen der App durch neue Farben oder ein anderes Design.

Dieses Feature richtet sich vermutlich an eine jüngere Zielgruppe, die mit solchen Mechaniken meist mehr vertraut ist als ältere Generationen. Es könnte sogar sein, dass dieses Feature als Verharmlosung der ernstesten Thematik interpretiert wird, was in einem gegensätzlichen Effekt zur eigentlichen Intention resultieren würde.

Eine Alternative zu Highscores und anderen virtuellen Errungenschaften wäre eine Art Belohnung in der Realität, zum Beispiel durch eine in der App durchgeführte Lotterie. Dies ist vermutlich attraktiver für alle Generationen, jedoch schwieriger umzusetzen, da hier Betrug oder mehrmalige Teilnahme verhindert werden muss. Vorstellbar wäre auch eine Verbindung mit den ermittelten Statistiken, sodass Leute mit guten Werten eine höhere Gewinnchance haben.

Eine Umsetzung dieses Features wurde im Laufe dieser Forschungsarbeit bereits in den USA, als Lotterie im Rahmen der Impfkampagne eingeführt [26]. Jeder Geimpfte hat dabei eine Chance den Hauptgewinn von einer Millionen US-Dollar zu gewinnen.

**3.2.3. Ortsbezogene Informationen.** Ein Problem das im Verlauf der Pandemie auftritt ist die undurchsichtige und inkonsistente Informationsgebung hinsichtlich der beschlossenen Maßnahmen und Einschränkungen an die sich die Bevölkerung richten muss. Viele dieser Maßnahmen wurden nicht auf Bundesebene beschlossen, sondern unabhängig von den Landesministerien, je nach Ausmaß der Krankheitsfälle im jeweiligen Bundesland. Selbst bei der bundesweiten Notbremse bezogen sich die Regelungen auf die Inzidenz der einzelnen Landkreise [27]. Dies führte zu einer großen Verwirrung welche derzeitigen Gesetze und Verordnungen gelten, zumal diese teilweise im Tagesrhythmus geändert wurden.

Auf Grund dessen ist ein Feature hilfreich, das die aktuellen Verordnungen abhängig vom jeweiligen Landkreis darstellt und einen Überblick verschafft. Dabei können die Nutzer unabhängig von ihrem Standort auf die Daten eines jeweiligen Landkreises zugreifen, um auf einfachem Weg die offiziellen Informationen zu erhalten.

Die ortsbezogenen Informationen benötigen keinerlei Nutzerdaten und liefern lediglich zusätzliche Informationen über die aktuelle Corona Lage Innerhalb eines Landkreises. Aus diesem Grund ist eine Implementierung dieses Features in keinsten Weise bedenklich im Sinne der Sicherheit und Privatheit der Nutzer. Durch einen Hackerangriff könnte man im schlimmsten Fall über diese Funktion herausfinden, in welchem Landkreis sich ein Nutzer mit hoher Wahrscheinlichkeit aufhält. Da die Informationen größtenteils auf den jeweiligen Webseiten der Länder zu Verfügung steht, ist eine Umsetzung des Features auch nicht mit einem großen Aufwand verbunden.

Dieses Feature existiert seit Oktober 2020 bereits in inoffizieller Ausführung als „Darf ich das?“ -App von der InTradeSys GmbH [28].

**3.2.4. Integration von Applikationen.** Es existieren viele verschiedene Plattformen, die fast täglich neue Nachrichten zur aktuellen Corona-Situation herausgeben [18], [19], Daten zur Eindämmung der Pandemie sammeln [20], [21] oder Neuigkeiten zu Forschungsarbeiten mit Bezug zu Corona veröffentlichen [22]. In diesem Feature werden Nachrichten, Datensammlungen, Umfragen und Forschungsarbeiten als *Informationen* zusammengefasst. Die CWA ist der perfekte Ort all diese Informationen der Plattformen zu vereinen. Da sie von der Bundesregierung in Auftrag gegeben wurde, besitzt sie ein hohes Maß an Integrität und Datenschutz steht im Vordergrund.

Es ist Ziel dieses Features die CWA so zu erweitern, dass sie Anlaufstelle für jede Person wird, die aktuelle, relevante und qualitativ hochwertige Informationen zu Corona sucht. Dabei steht im Vordergrund, dass die Informationen geprüft sein sollen, um die Verbreitung von falschen Informationen zu vermeiden. Dazu dürfen nur vertrauenswürdige Quellen verwendet werden, wie zum Beispiel die Bundesregierung, das Robert-Koch-Institut oder Forschungsinstitute. Die CWA soll so erweitert werden, dass die Informationen zwar weiterhin auf den eigenen Webseiten der Herausgeber

veröffentlicht werden, jedoch zusätzlich in die CWA eingebunden werden.

Durch das Zusammenführen verschiedenster Informationen wird eine sehr große Reichweite erzeugt, die gerade für die Datensammlung und Forschung große Relevanz hat. Des Weiteren würde die CWA durch die zusätzlichen Informationen attraktiver, was einen positiven Effekt auf die Kontaktverfolgung hätte.

In Version 1.13 der CWA wurde bereits eine freiwillige Datenspende in Zusammenarbeit mit dem Robert-Koch-Institut eingefügt [30].

## 4. Erweiterte Risikoberechnung

Die Risikoberechnung der aktuellen Corona Warn App zeigt einem Nutzer zwei Risiken an: das erhöhte Risiko mit einem roten Bildschirm und das geringe Risiko mit einem grünen Bildschirm. Generell wird angenommen, dass, wegen der Inkubationszeit, Tests beziehungsweise Begegnungen innerhalb der letzten 14 Tage stattgefunden haben. Um das Risiko zu berechnen wird das *Übertragungspotenzial* aus Begegnungsdauer und Begegnungsnähe mit einer Person, die ein positives Testergebnis in die CWA eingetragen hat, mit einbezogen. Wenn das Übertragungspotenzial sich insgesamt zu einer Begegnungsdauer von 13 Minuten in weniger als 8 Metern Abstand aufsummiert, wird der Bildschirm rot und ein erhöhtes Risiko angezeigt. Diese Berechnung hätte einen höheren Nutzen, wenn mehr Personen die App herunterladen würden und mehr Tests zur Verfügung gestellt werden könnten [34]. Je mehr Tests durchgeführt werden und auch in die CWA eingetragen werden, umso genauer wird die aktuelle Risikoberechnung. Die aktuelle Risikoberechnung der CWA ist in Abbildung ?? zu sehen.

Die aktuelle CWA berechnet nur Begegnungen mit Infizierten, die ihr Testergebnis in die CWA eingetragen haben. In der erweiterten Risikoberechnungen sollten jedoch Begegnungen mit allen Personengruppen mit einberechnet werden.

Dabei wird zwischen zwei Personengruppen unterschieden, die jeweils in 3 Untergruppen unterteilt werden, da die Infektionswahrscheinlichkeit in manchen Personengruppen höher ist als in Anderen. Es wird grundsätzlich unterschieden zwischen *CWA-User*, die von der CWA somit aufgezeichnet werden und *Not-CWA-User*, die die CWA nicht aktiv nutzen und somit nicht aufgezeichnet werden. Die CWA nicht aktiv nutzen heißt, dass keine TEKs geteilt werden. Die CWA-User werden in drei Untergruppen aufgeteilt:

- *CWA-Positive*: Alle Personen, die einen positiven Test in die CWA eingetragen haben.
- *CWA-Likely-Positive*: Alle Personen, die mindestens eine Begegnung mit CWA-Positive hatten und einen roten Bildschirm auf der CWA angezeigt bekommen.
- *CWA-Likely-Negative*: Alle Personen, die meist keine Begegnung mit CWA-Positive hatten und somit einen grünen Bildschirm auf der CWA angezeigt bekommen.

Ebenso werden die NOT-CWA-User in drei Untergruppen aufgeteilt:

- *Not-CWA-Positive*: Alle Personen, die ein positives Test Ergebnis haben, dieses aber nicht in die CWA eingetragen haben (Es ist möglich, dass sie die CWA trotzdem passiv nutzen).
- *Hidden-Positive*: Alle Personen, die infiziert wurden, jedoch kein positives Testergebnis nachweisen können.
- *Hidden-Negative*: Alle Personen, die nicht infiziert sind.

Bei näherer Betrachtung der aktuellen Risikoberechnung der CWA sind uns mehrere Kritikpunkte aufgefallen. Zum einen fließen nur Begegnungen mit CWA-Positive Personen in diese mit ein. Deshalb sollen in der erweiterten Risikoberechnung alle Personengruppen miteinbezogen und unterschiedlich gewichtet werden. Zum anderen wird die Dunkelziffer nicht berücksichtigt, die im Folgenden durch die Begegnungsanzahl betrachtet werden soll. Des Weiteren soll die geringe Nutzerzahl der CWA zum Hochkorrigieren der Begegnungsanzahl genutzt werden. Als letzten Kritikpunkt wird die Granulierung und Transparenz der Risikoberechnung aufgezeigt. Aus diesem Grund soll der Risikowert, statt der bisherigen zwei, in vier unterschiedliche Risikostufen (grün, gelb, orange, rot) aufgeteilt werden, dessen Zustandekommen mit angezeigt wird.

Für die erweiterte Risikoberechnung werden die Features Familienfeature, Feature TRL Update und Feature PK-Server benötigt.

Das Familienfeature verhindert das mehrmalige Aufzeichnen von vertrauenswürdigen Personen, wie etwa Familienmitglieder, damit die Dunkelziffer besser miteinberechnet werden kann. Mit dem Feature TRL Update ist es möglich, den Krankheitsverlauf in die Risikoberechnung mit einfließen zu lassen, um zu beachten, wie infektiös eine Person ist. Dabei ist dies nicht mit dem im Abschnitt 4 definierten Begriff Übertragungspotenzial zu verwechseln. Das TRL beschreibt den Grad der Infektiosität im Krankheitsfall an, während das Übertragungspotenzial aus Begegnungsdauer und Nähe mit einer infizierten Person berechnet wird, um die Gefahr einer Infektion zu berechnen. Das Feature DK-Server wird benötigt, um Begegnungen mit CWA-Likely-Positive Personen zu erkennen und diese in der erweiterten Risikoberechnung gesondert zu gewichten.

Die oben genannten Features werden in Abschnitt 4.1 bis 4.3 vorgestellt. In Abschnitt 4.4 werden Probleme und Lösungsansätze der aktuellen Risikoberechnung diskutiert und die drei Summanden, aus denen der Risikowert zusammengesetzt ist, genauer erläutert. Zuletzt wird die erweiterte Risikoberechnung in Abschnitt 4.5 bewertet.

## 4.1. Familienfeature

Im Folgenden wird das Familienfeature vorgestellt, welches später für die erweiterte Risikoberechnung benötigt wird. Das Familienfeature ermöglicht es Gruppen von

Verwandten bzw. Freunden zu erstellen und mit diesen Temporary-Exposure-Keys zu teilen. Mithilfe der geteilten TEKS ist es möglich von Freunden bzw. Verwandten empfangene Rolling Proximity Identifier bei der Risikoberechnung gesondert zu behandeln und somit einen genaueren Risikowert zu berechnen. Dies geschieht, indem bei der Risikoberechnung die Begegnungs RPIs mit den TEKS aus den einzelnen Gruppen abgeglichen werden und nur die RPIs mit einbezogen werden, die keinen passenden TEK in irgendeiner Gruppe haben.

Das Austauschen von TEKS kann allerdings Eingriffe in die Privatsphäre zur Folge haben. Erkrankt eine Person aus einer Gruppe an Corona und gibt ihre TEKS als Diagnosis Keys frei, ist es für die anderen Gruppenangehörigen mit hoher Wahrscheinlichkeit möglich, die erkrankte Person zu bestimmen. Aus diesem Grund sollten Gruppen nur mit vertrauenswürdigen Personen, wie engen Freunden oder Verwandten, gegründet werden. In Abbildung 1 ist eine Übersicht des Familienfeatures gegeben, welches im Folgenden beschrieben wird:

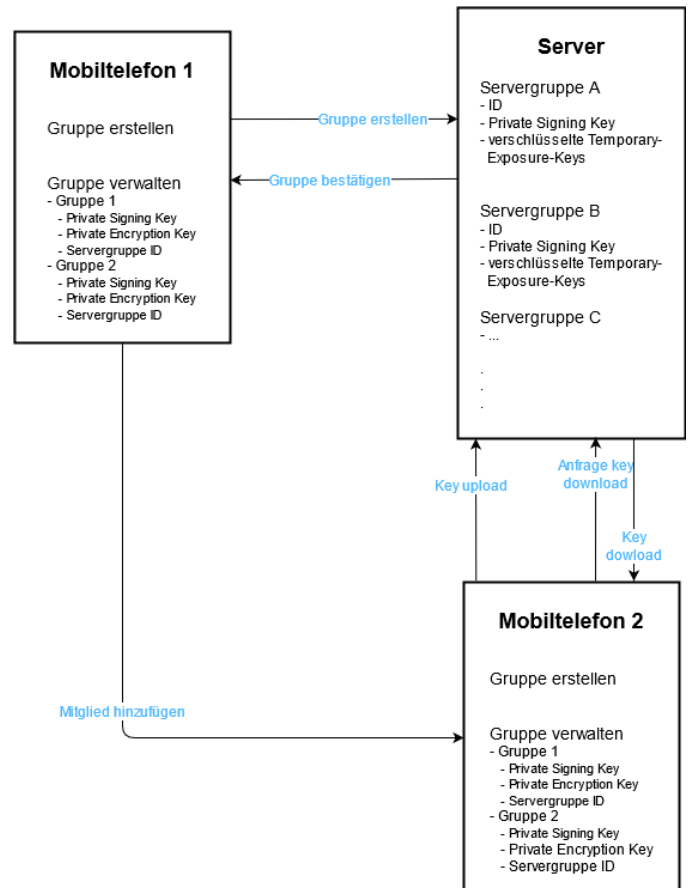


Abbildung 1. Familienfeature

Als Servergruppe wird eine Gruppe auf einem speziell für das Familienfeature angelegten Server bezeichnet. Über den Server werden später die TEKS von unterschiedlichen CWA Nutzern ausgetauscht. Eine Servergruppe besitzt eine

für den Server einmalige *ID*, einen *Private Signing Key* und eine *Liste mit verschlüsselten TEKs*.

Die Liste der verschlüsselten TEKs ist beim Anlegen einer neuen Servergruppe leer.

Jeder Nutzer der CWA kann beliebig viele lokale Gruppen in der eigenen CWA erstellen, mit denen der Nutzer TEKs austauschen möchte. Innerhalb einer lokalen Gruppe wird ein *Private Signing Key* (PSK), ein *Private Encryption Key* (PEK) und eine *Servergruppen ID* gespeichert. PSKs und PEKs werden genutzt, um TEKs zu verschlüsseln und zu signieren. Die Verwendung des PSK, PEKs und der Servergruppen ID wird in den nächsten Absätzen genauer erläutert. Der PSK und insbesondere der PEK einer Gruppe müssen geheim gehalten werden, damit die Privatsphäre keines Mitglieds verletzt wird. Es gibt zwei verschiedene Möglichkeiten den PSK, PEK und die Servergruppen ID einer lokalen Gruppe hinzuzufügen. Dafür kann man entweder eine neue Servergruppe anlegen oder einer bereits existierenden Servergruppe beitreten.

Entscheidet man sich dazu eine neue Servergruppe anzulegen, so wird lokal in der CWA ein PSK und ein PEK generiert. Der PSK wird zusammen mit einer Anfrage, eine neue Servergruppe zu erstellen, an den Server gesendet. Ein Verlust des PSKs kann dazu führen, dass ein Angreifer willkürlich TEKs hochladen und herunterladen kann, die dieser Gruppe zugeordnet werden. Allerdings reicht der PSK nicht aus, um die TEKs von Gruppenmitgliedern lesen zu können, da diese mit dem PEK verschlüsselt sind. Der Server erstellt auf die Anfrage hin eine neue Servergruppe mit einer noch nicht verwendeten ID und schickt die ID an den Anfrager zurück. Der lokal erstellte PSK, PEK, sowie die erhaltene Servergruppen ID werden der Gruppe zugehörig abgespeichert.

Im Falle der zweiten Möglichkeit, einer bereits existierenden Servergruppe beizutreten, benötigt man Kontakt zu einem Mitglied der Servergruppe. Das Mitglied der Servergruppe teilt den PSK, PEK und die Servergruppen ID durch zum Beispiel einen QR-Code. Diese müssen dann nur noch abgespeichert werden.

An dieser Stelle ist anzumerken, dass jedes Mitglied neue Personen zur Gruppe hinzufügen kann, ohne, dass andere Mitglieder etwas davon mitbekommen. Aus diesem Grund wird nochmal betont Gruppen nur mit vertrauenswürdigen Personen zu teilen, um die Privatsphäre zu schützen. Sollte eine Person unbefugten Zugriff auf eine Gruppe haben, ist es nicht möglich, dieser Person den Zugriff zu entziehen. In diesem Fall muss die lokale Gruppe gelöscht werden, damit keine weiteren TEKs geteilt werden.

Hat man eine oder mehrere Gruppen lokal erstellt und die zugehörigen PSKs, PEKs und Servergruppen IDs hinzugefügt, werden automatisch alle 24h die eigenen TEKs auf den Server hochgeladen und die TEKs der anderen Mitglieder heruntergeladen. Damit außer der Gruppenmitglieder niemand Zugriff auf die TEKs hat, wird dabei folgendes Verfahren angewandt:

Die TEKs werden mithilfe des PEKs einer Gruppe und eines sicheren, symmetrischen Verschlüsselungsverfahrens verschlüsselt und dann mit dem PSK der Gruppe und einem

sicheren, symmetrischen Signaturverfahren signiert. Dies wird zusammen mit der Servergruppen ID der Gruppe an den Server gesendet. Der Server nutzt den PSK zugehörig zur mitgeschickten ID, um die TEKs zu validieren. Sind die TEKs validiert, werden sie der Liste der verschlüsselten TEKs zusammen mit ihrem Hochladedatum hinzugefügt. Die verschlüsselten TEKs werden wegen der Inkubationszeit nach 14 Tagen aus der Liste der verschlüsselten TEKs gelöscht, da sie danach nicht mehr benötigt werden. Dies geschieht iterativ für jede lokale Gruppe.

Um die TEKs der Gruppenmitglieder einer Gruppe herunterzuladen, muss eine Anfrage an den Server gestellt werden. Diese Anfrage beinhaltet die Servergruppen ID der Gruppe und das Datum des letzten Downloads. Mithilfe des PSKs wird die Anfrage validiert und die verschlüsselten TEKs vom Server an die CWA gesendet, die nach dem Datum des letzten lokalen Downloads hochgeladen wurden. Die verschlüsselten TEKs können dann mit dem PEK der Gruppe entschlüsselt werden. Auch hier geschieht dies iterativ für jede lokale Gruppe.

## 4.2. Feature DK-Server

Das Feature DK-Server ist Voraussetzung für die erweiterte Risikoberechnung. In der aktuellen CWA werden nur Begegnungen mit CWA-Positive Personen in der Risikoberechnung betrachtet. Jedoch sind CWA-Likely-Positive Personen mit einer erhöhten Wahrscheinlichkeit infiziert und ansteckend, da sie Kontakt zu CWA-Positive Personen hatten. Diese sollen mithilfe dieses Features in den Summanden  $R_{CWA-Likely-Positive}$  in Abschnitt 4.4.3 mit einbezogen werden.

Damit diese in die Risikoberechnung mit einfließen können, benötigen wir die TEKs der CWA-Likely-Positive Personen. Ähnlich wie beim CWA Server sollen diese TEKs auch auf einem Server gesammelt und von da aus verteilt werden. Dieser Server wird in diesem Dokument als der *Direkte Kontakte Server* (DK-Server) bezeichnet. Wird in der CWA ein erhöhtes Risiko mit einem Risikowert  $\geq 13$  erkannt und durch einen roten Bildschirm angezeigt, so sollen die TEKs der letzten 14 Tage, verbunden mit dem zum roten Bildschirm führenden Risikowert, auf den DK-Server hochgeladen werden. Dies kann entweder durch das freiwillige Bestätigen einer Option oder als Integration in die freiwillige Datenspende implementiert werden. Aufgrund der Inkubationszeit werden TEKs auf dem DK-Server, die älter als 14 Tage sind, gelöscht.

Die CWA lädt regelmäßig alle neuen TEKs mit den zugehörigen Risikowerten vom DK-Server herunter und leitet daraus die RPIs ab. Die Verwendung der RPIs und Risikowerte in der Risikoberechnung wird im Abschnitt 4.4.3 beschrieben. Das Ableiten der RPIs aus den heruntergeladenen TEKs funktioniert genau so, wie das Ableiten der RPIs aus den Diagnosis Keys auf dem CWA Server. Der mit einem TEK verbundene Risikowert wird dazu genutzt, um die Wahrscheinlichkeit zu ermitteln, ob der Teilende des TEKs mit Corona infiziert ist.



Anders als beim Teilen von TEKs als Diagnosis Keys mit dem CWA Server im Falle eines positiven Testergebnisses wird kein eindeutiger, autorisierter Schlüssel benötigt, um TEKs auf den DK-Server hochladen zu können. Dadurch entfällt die Integrität der hochgeladenen TEKs. Dies wirft verschiedene Probleme auf, die im Folgenden diskutiert werden. Die Lösung dieser Probleme geht jedoch über den Umfang dieser Forschungsarbeit hinaus und wird deswegen nur angerissen.

Die vorgeschlagene Implementierung dieses Features sieht vor, dass ein TEK nicht mehrmals innerhalb von 14 Tagen auf den DK-Server hochgeladen wird. Sollten bereits auf dem DK-Server existierende TEKs erneut hochgeladen werden, so müssen diese abgelehnt werden, weil nicht bestätigt wird, ob das erneute Hochladen von der gleichen Person getätigt wird. Würden diese nicht abgelehnt werden, wäre es ein Leichtes für einen Angreifer die Risikowerte der TEKs zu verändern, da der Lesezugriff auf die TEKs und Risikowerte auf dem DK-Server nicht eingeschränkt ist. Dies führt dazu, dass nur der erste Risikowert, der zu einem roten Bildschirm führt, zusammen mit den TEKs auf den DK-Server hochgeladen werden kann. Da die Risikowerte in direkter Verbindung mit der Risikoberechnung stehen, kann dies zu einer ungenaueren Risikoberechnung führen, falls die erste Begegnung, die bei einer Person zu einem Risikowert  $\geq 13$  geführt hat, nicht die Begegnung ist, an der sich die Person angesteckt hat.

Es ist des Weiteren möglich falsche TEKs auf den DK-Server hochzuladen, da weder die Existenz eines TEKs, noch das Vorhanden sein eines Risikowertes  $\geq 13$ , überprüft wird. Dies ermöglicht Denial of Service Angriffe mit verhältnismäßig geringem Aufwand und Ressourcen.

### 4.3. Feature TRL Update

In diesem Feature wird das volle Potenzial des TRLs, welches in Abschnitt 2.4 genauer beschrieben wurde, genutzt. Dafür wird festgelegt, dass das TRL den ursprünglichen Standardwert von Vier nicht unterschreiten darf, da mit zu geringer Sicherheit gesagt werden kann, dass eine infizierte Person nicht mehr infektiös ist. Die Implementierung für das TRL existiert bereits in der CWA, jedoch existieren die benötigten Daten noch nicht. Es handelt sich hierbei konkret um den Zeitpunkt, ab dem man infektiös ist. Diesen Zeitpunkt kann man allerdings nicht so einfach bestimmen, da nicht täglich Tests an der selben Person durchgeführt werden. Der Zeitpunkt ist jedoch wichtig, denn nur mit diesem ist es möglich den Krankheitsverlauf mit in die Risikoberechnung einzubeziehen, der dafür sorgt, dass eine Person nicht immer gleich infektiös ist. Um diesen Zeitpunkt zu approximieren werden Umfragen erstellt. In diese Umfragen, die beim Eingeben eines positiven Corona Tests aktiviert werden, sollen die Infizierten ihren Infektionszeitpunkt und den Symptom-Start-Zeitpunkt angeben. Es wird angenommen, dass beim Eingeben des Testergebnisses der Zeitpunkt des Testens, der Zeitpunkt des Ergebnis Erhaltens

und der Zeitpunkt des Eingebens in der App gespeichert werden. Eine Umfrage könnte folgende Fragen beinhalten:

- Wann haben Sie in den letzten 14 Tagen Symptome verspürt?
- Wann wurden Sie vermutlich infiziert?

Im Falle, dass es nicht möglich ist, die Zeitpunkte des Testens und des Rückmeldens im Testergebnis zu codieren, kann man folgende Fragen anhängen:

- Wann haben Sie sich testen lassen?
- Wann haben Sie das positive Ergebnis bekommen?

Mit diesen Daten kann dann der Zeitpunkt des Infektionsseins kalkuliert werden und mithilfe des TRLs für jeden Diagnosis Key abgeschätzt werden, damit Nutzer besser gewarnt werden können.

### 4.4. Berechnung

In diesem Abschnitt werden die drei Summanden  $R_{CWA-Positive}$ ,  $R_{CWA-Likely-Positive}$  und  $R_{CWA-Likely-Negative}$  beschrieben, aus denen sich der neue Risikowert  $R_{new}$  zusammensetzt.

In Abschnitt 4.4.1 werden Probleme und Lösungsansätze der aktuellen Risikoberechnung genauer erläutert. Danach werden in Abschnitt 4.4.2 - 4.4.4 die drei Summanden näher beschrieben, wobei die angepasste Grafik in Abbildung 2 dies veranschaulicht.

**4.4.1. Probleme und Lösungsansätze.** Das Hauptproblem der Risikoberechnung der aktuellen CWA ist, dass diese zu wenige, relevante Faktoren mit einberechnet, denn nur CWA-Positive Personen werden mit einbezogen. Alle anderen Personengruppen werden ignoriert. Der Risikowert ist somit geringer als er eigentlich sein sollte. Im Folgenden werden weitere Faktoren beschrieben, die die erweiterte Risikoberechnung beachtet:

Mit der Dunkelziffer lässt sich die Anzahl an tatsächlich Infizierten abschätzen und sie sorgt dafür, dass CWA-Likely-Negative Personen, unter Einbezug der infizierten Not-CWA-User, zu einem Risiko führen können, welches in der neuen Risikoberechnung nicht ignoriert wird. Die Downloadzahl wird genutzt, um den Risikowert nach oben zu korrigieren, da die Not-CWA-User nicht als Begegnung aufgezeichnet werden. Begegnungen mit CWA-Likely-Positive Personen, die eine erhöhte Wahrscheinlichkeit haben infiziert zu sein, da sie direkten Kontakt zu Infizierten hatten, werden in der neuen Berechnung anders als Begegnungen mit CWA-Likely-Negative Personen gewichtet. Außerdem ist die Risikoberechnungsanzeige nicht transparent genug, zwar ist die Anzahl an Begegnungen mit CWA-Positive Personen sichtbar, jedoch sieht man weder den Grad des Übertragungspotenzials, noch wie der Risikowert zustande gekommen ist.

Die neue Berechnungsvorschrift unterscheidet zwischen Begegnungen mit sicher Infizierten (CWA-Positive Gruppe mit positivem Testergebnis), wahrscheinlich positiven Personen (CWA-Likely-Positive Gruppe) und wahrscheinlich

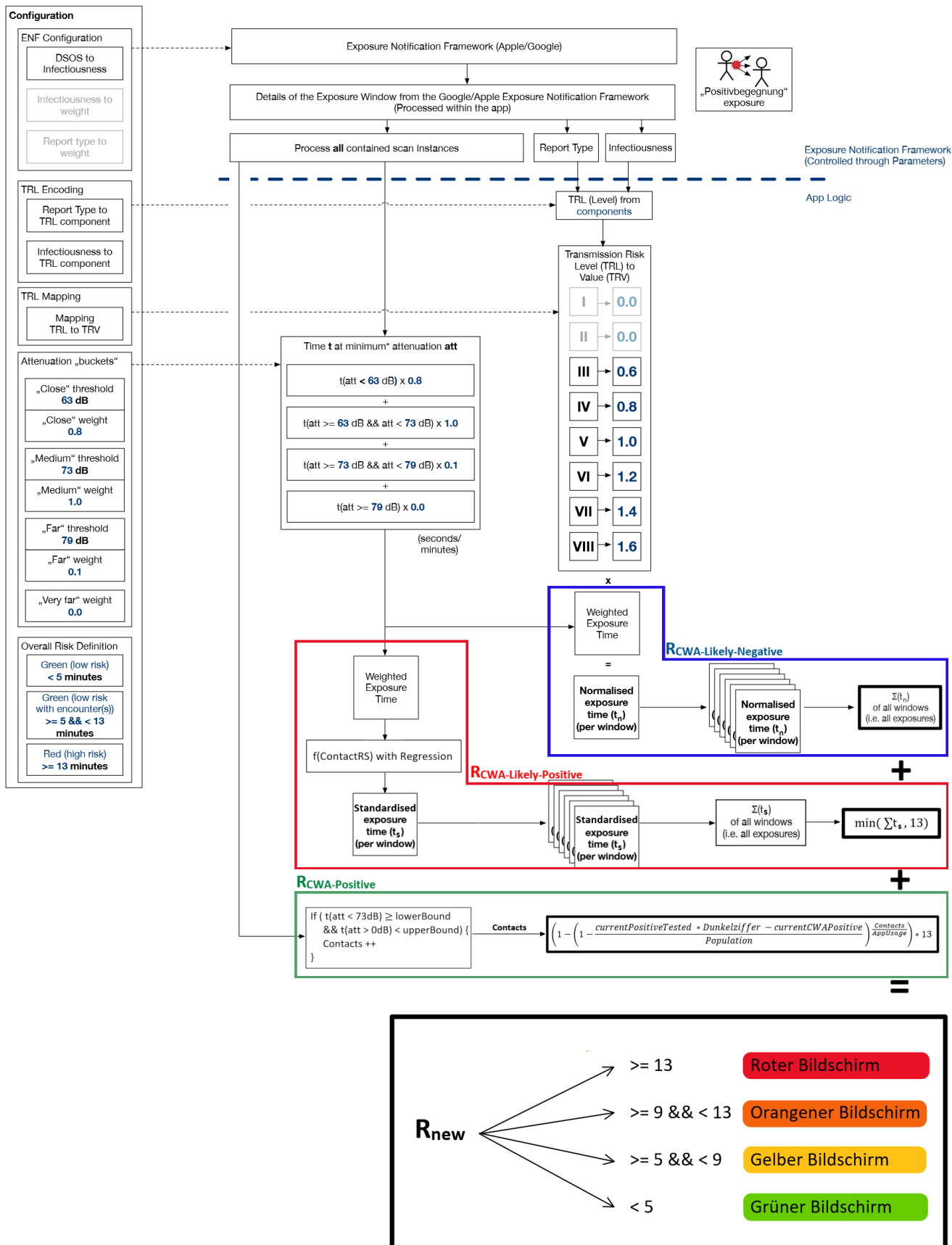


Abbildung 2. Erweiterte Risikoberechnung adaptiert von Abbildung ??

nicht Infizierten (CWA-Likely-Negative Gruppe). Dementsprechend setzt sich der neue Risikowert aus drei Summanden zusammen, damit die unterschiedlichen Personengruppen und Faktoren mit einbezogen werden können. Der *neue Risikowert*  $R_{\text{new}}$  setzt sich somit aus folgenden Summanden zusammen:

$$R_{\text{new}} := R_{\text{CWA-Positive}} + R_{\text{CWA-Likely-Positive}} + R_{\text{CWA-Likely-Negative}}$$

Der erste Summand ist  $R_{\text{CWA-Positive}}$  und berechnet das Risiko, indem nur Begegnungen mit positiv getesteten (CWA-Positive) Personen einbezogen werden. Dabei wird jede Begegnung gewichtet, je nach dem wie hoch das Übertragungspotenzial der Begegnung ist.

Der zweite Summand  $R_{\text{CWA-Likely-Positive}}$  schätzt das Risiko von Begegnungen mit Likely-Positive Personen, sprich roten Bildschirmen bzw. CWAs mit einem Risikowert von über 13, aber keinem positiven Test Ergebnis, ab.

Außerdem gibt es noch den dritten Summanden  $R_{\text{CWA-Likely-Negative}}$ , der die Begegnungen mit Personen aus der Personengruppe CWA-Likely-Negativ mit einberechnet, wobei auch die Personengruppe der Not-CWA-User mit betrachtet wird. Der neue Risikowert wird anschließend in eine Bildschirmfarbe umgewandelt, die als Höhe des Infektionsrisikos interpretiert werden kann.

Zudem sollte es eine geringere Abstufung der Risikostufen geben, denn nur grün und rot wecken den Eindruck, dass man bei einem grünen nicht gefährdet wäre. Ab einem gewissen Risiko sollte auch ein gelber sowie ein orangener Bildschirm angezeigt werden, damit die Nutzer ihr Verhalten besser regulieren können.

Dabei werden anstatt der ursprünglichen zwei Farben nun vier Farben Rot, Orange, Gelb und Grün verwendet, die die entsprechenden Risikoklassen symbolisieren.

$$\begin{aligned} 0 \leq R_{\text{new}} < 5 &\rightarrow \text{grüner Bildschirm} \\ 5 \leq R_{\text{new}} < 9 &\rightarrow \text{gelber Bildschirm} \\ 9 \leq R_{\text{new}} < 13 &\rightarrow \text{orangener Bildschirm} \\ 13 \leq R_{\text{new}} &\rightarrow \text{roter Bildschirm} \end{aligned}$$

Ein weiteres Problem der aktuellen Risikoberechnung ist die geringe Transparenz des Zustandekommens des Risikowertes. Ein transparenter Risikowert macht es für Nutzer leichter nachvollziehbar, wie sie sich bei einem erhöhten Risiko Verhalten sollten und fördert die aktive Nutzerzahl. Damit die Risikoberechnung transparenter für die Nutzer wird, werden die Anteile der einzelnen Summanden zum farbigen Bildschirm in der App angegeben, sowie die Begegnungsdauer und Begegnungsnähe zu CWA-Positive Nutzern.

**4.4.2.  $R_{\text{CWA-Positive}}$ .** Das Risiko  $R_{\text{CWA-Positive}}$  entspricht der aktuellen Risikoberechnung der CWA (Stand 26.04.2021). In dieses Risiko fließen alle Begegnungen mit der Personengruppe CWA-Positive ein. Da CWA-Positive Personen

ein positives Testergebnis in der CWA eingetragen haben, ist die Positiv-Begegnungs-Wahrscheinlichkeit hier 1.

Es wird also nur das Übertragungspotential betrachtet. Dafür wird zunächst zu einer Begegnung die *Weighted Exposure Time* ausgewertet. Die Weighted Exposure Time bestimmt die Gewichtung der Begegnungszeit  $t$  anhand der geringsten Dämpfung (attenuation)  $att$  wie folgt:

$$\text{WeightedExposureTime} := t \cdot \text{attFactor}$$

Wobei der  $\text{attFactor}$  durch 22 dB folgende Intervalle aus der Distanz  $att$ , des BLE-Signals bestimmt wird:

$$\begin{aligned} att < 63 \text{ dB} &: \text{attFactor} = 0.8 \\ 63 \text{ dB} \leq att < 73 \text{ dB} &: \text{attFactor} = 1.0 \\ 73 \text{ dB} \leq att < 79 \text{ dB} &: \text{attFactor} = 0.1 \\ 79 \text{ dB} \leq att &: \text{attFactor} = 0 \end{aligned}$$

Dabei kann man die Werte 63 dB bzw. 73 dB, als Signalstärke von einem Handy interpretieren, das 3 Meter bzw 8 Meter weit entfernt ist. Dabei schwächen Kleidung, Körper oder andere Gegenstände das Signal ab und sorgen dafür, dass die meisten Begegnungen erst auf kleinerer Distanz aufgezeichnet werden [33].

Mithilfe des Transmission Risk Levels TRL und des daraus resultierenden *Transmission Risk Values* TRV wird danach aus der Weighted Exposure Time die *Normalised Exposure Time*  $t_n$  berechnet.

$$t_n := \text{Weighted Exposure Time} \cdot \text{TRV}$$

Das Transmission Risk Level wird in Abschnitt 2.4 genauer beschrieben und bezieht die Wahrscheinlichkeit mit ein, dass eine Person die Krankheit auch wirklich überträgt. Der TRV ergibt sich aus dem TRL nach der folgenden Vorschrift:

$$\begin{aligned} \text{TRL} = I &\rightarrow \text{TRV} = 0.0 \\ \text{TRL} = II &\rightarrow \text{TRV} = 0.0 \\ \text{TRL} = III &\rightarrow \text{TRV} = 0.6 \\ \text{TRL} = IV &\rightarrow \text{TRV} = 0.8 \\ \text{TRL} = V &\rightarrow \text{TRV} = 1.0 \\ \text{TRL} = VI &\rightarrow \text{TRV} = 1.2 \\ \text{TRL} = VII &\rightarrow \text{TRV} = 1.4 \\ \text{TRL} = VIII &\rightarrow \text{TRV} = 1.6 \end{aligned}$$

Mit dem TRV lässt sich dann  $t_n$  bestimmen.

Zuletzt wird aus den Normalised Exposure Times der einzelnen Begegnungen die Summe gebildet und man erhält damit das Risiko  $R_{\text{CWA-Positive}}$ :

$$R_{\text{CWA-Positive}} := \sum t_n$$

**4.4.3.  $R_{\text{CWA-Likely-Positive}}$ .** Das Risiko  $R_{\text{CWA-Likely-Positive}}$  ergibt sich aus den Begegnungen mit CWA-Likely-Positive Personen, die eine erhöhte Wahrscheinlichkeit aufweisen,

infiziert zu sein und deswegen in der Berechnung gesondert behandelt werden.

Hierbei werden Begegnungen mit Gruppenangehörigen aus dem Familienfeature, Abschnitt 4.1, nicht mit einbezogen, da diese sonst aufgrund der potenziell langen Begegnungsdauer und kurzen Distanz ein hohes  $R_{CWA-Likely-Positive}$  Risiko hervorrufen könnten und man im Regelfall diesen Personen ohnehin sofort mitteilt, wenn man ansteckend sein könnte.

Zudem werden Begegnungen mit Diagnosis Keys nicht mit einberechnet, da diese bereits im Summand  $R_{CWA-Positive}$  betrachtet werden. Dies geschieht, indem bei der Risikoberechnung die Begegnungs RPIs mit den Diagnosis Keys auf dem CWA Server abgeglichen werden und nur die RPIs mit einbezogen werden, für die keine passenden Diagnosis Keys auf dem CWA Server existieren.

Zuerst wird, wie beim Summand  $R_{CWA-Positive}$ , das Übertragungspotenzial einer Begegnung bestimmt, indem die Weighted Exposure Time ausgewertet wird.

Aus der Weighted Exposure Time wird danach allerdings nicht die Normalised Exposure Time bestimmt, wie dies beim Summand  $R_{CWA-Positive}$  der Fall ist, da dafür ein positives Testergebnis vorausgesetzt wird, welches CWA-Likely-Positive Personen nicht vorweisen können.

Anstelle dessen wird als nächstes aus der Weighted Exposure Time jeder Begegnung die *Standardized Exposure Time*  $t_s$  berechnet.

Da die Begegnung der CWA-Likely-Positive Personen-Gruppe angehört, ist es nicht sicher, dass diese auch Corona hat. Aus diesem Grund wird die *Positiv-Begegnungswahrscheinlichkeit* benötigt, welche die Wahrscheinlichkeit angibt, dass die Begegnung infiziert ist. Hierzu wird die Funktion  $f$  eingeführt, die die Positiv-Begegnungswahrscheinlichkeit der Begegnung angibt.

$$f : \text{Risikowert} \rightarrow \text{Wahrscheinlichkeit}$$

Die Funktion  $f$  nimmt als Input den Risikowert der Begegnung und gibt die Wahrscheinlichkeit zurück, dass diese infiziert ist. Der Risikowert wurde zuvor mit dem Temporary Exposure Key zusammen vom DK-Server heruntergeladen, wie in Abschnitt 4.2 beschrieben.

Um  $f$  zu bestimmen wird folgendes Verfahren angewandt: Zunächst werden, Intervalle gebildet, da die Menge der Risikowerte nicht diskret, sondern stetig ist. Anhand der auf dem DK-Server und dem CWA Server gesammelten Daten wird für jedes Risikowert-Intervall  $x_i$  mit  $i \in N$  und einer Intervallgröße von 1 die Wahrscheinlichkeit bestimmt, dass eine Begegnung, mit einem Risikowert innerhalb  $x_i$ , infiziert ist. Dabei gilt für jedes Intervall:

$$x_i := [13 + i, 14 + i[$$

Es sind somit nur Risikowerte über 13 relevant, da die Risikowerte auf dem DK-Server alle  $\geq 13$  sind. Damit die Wahrscheinlichkeit bestimmt werden kann, dass eine Person mit einem Risikowert innerhalb eines bestimmten Intervalls  $x_i$  infiziert ist, wird die Anzahl an Personen benötigt, die

einen Risikowert innerhalb des Intervalls  $x_i$  haben und die Anzahl an Personen, die sowohl einen Risikowert innerhalb des Intervalls  $x_i$  haben, als auch später an Corona erkrankt sind.

Dafür werden die TEKs auf dem DK-Server und die Diagnosis Keys auf dem CWA Server als historische Daten verwendet. Aus den TEKs auf dem DK-Server kann die Anzahl an Personen bestimmt werden, die einen Risikowert innerhalb des Intervalls  $x_i$  haben. Werden zusätzlich zu den TEKs auf dem DK-Server die Diagnosis Keys auf dem CWA Server dazu genommen, kann damit die Anzahl an Personen bestimmt werden, die sowohl einen Risikowert innerhalb des Intervalls  $x_i$  haben, als auch später an Corona erkrankt sind, indem die TEKs auf dem DK-Server und den zugehörigen Risikowerten mit den Diagnosis Keys auf dem CWA Server abgeglichen werden. Daraus lässt sich für jedes Intervall  $x_i$  die Wahrscheinlichkeit  $P_{x_i}$ , infiziert zu sein, wie folgt bestimmen:

$$P_{x_i} := \frac{\text{Anzahl TEKs auf dem DK-Server und CWA Server mit Risikowert} \in x_i}{\text{Anzahl TEKs auf dem DK-Server mit Risikowert} \in x_i}$$

Zur Anzahl TEKs auf dem DK-Server und CWA Server mit Risikowert  $\in x_i$  werden nur die TEKs gezählt, die auf DK-Server und den CWA Server hochgeladen wurden und einen Risikowert innerhalb  $x_i$  aufweisen.

Zur Anzahl TEKs auf dem DK-Server mit Risikowert  $\in x_i$  werden alle TEKs auf dem DK-Server gezählt, die einen Risikowert innerhalb  $x_i$  aufweisen.

Da in der Regel zwischen dem Hochladen eines TEKs auf den TEK Server und dem Hochladen des gleichen TEKs als Diagnosis Key auf den CWA Server einige Tage vergehen und Keys auf beiden Servern nur 14 Tage lang gespeichert werden, muss für die Bestimmung der Funktion  $f$  die richtige Datenmenge genommen werden.

Die Anzahl TEKs auf dem DK-Server mit Risikowert  $\in x_i$  werden ab dem Startzeitpunkt  $s_0$ , der 14 Tage in der Vergangenheit liegt, bis zu einem Endzeitpunkt  $s_1$ , der 7 Tage nach  $s_0$  liegt, gezählt. Die Anzahl TEKs auf dem DK-Server und CWA Server mit Risikowert  $\in x_i$  werden ab dem gleichen Startzeitpunkt  $s_0$  bis zum Tag der Bestimmung der Funktion  $f$  gezählt. Dabei muss allerdings der TEK im Zeitraum  $s_0$  bis  $s_1$  auf den DK-Server hochgeladen worden sein.

Verwendet man zur Bestimmung der Funktion  $f$  zum Beispiel die TEKs der letzten 7 Tage auf dem DK-Server, so werden viele dieser TEKs nicht als Diagnosis Keys auf den CWA Server hochgeladen, obwohl diese infiziert sind, da noch kein positives Testergebnis aufgrund der geringen zeitlichen Distanz vorliegt. Die davon betroffenen TEKs würden die Funktion  $f$  negativ beeinflussen.

Nachdem für jedes Intervall  $x_i$  die Wahrscheinlichkeit, infiziert zu sein, bestimmt wurde, wird durch ein Regressionsverfahren die Funktion  $f$  bestimmt, die jedem Risikowert die Positiv-Begegnungswahrscheinlichkeit zuordnet. Hierbei ist besonders zu beachten, dass die Funktion  $f$  periodisch mit aktuelleren Daten neu berechnet und für die CWA aktualisiert werden muss. Eine Aktualisierungsperiode sollte ca. 7 Tage andauern, da der Zeitraum  $s_0$  bis  $s_1$  7 Tage lang

ist. Beim Herunterladen von Keys (vom CWA Server oder DK-Server) innerhalb der CWA muss dann gleichzeitig die aktualisierte Funktion  $f$  heruntergeladen werden. Mithilfe der Funktion  $f$  kann dann  $t_s$  bestimmt werden:

$$t_s := \text{Weighted Exposure Time} \cdot f(\text{ContactRS})$$

Hierbei ist ContactRS der Risikowert des Begegneten. Als nächstes wird aus den Standardized Exposure Times der einzelnen Begegnungen mit CWA-Likely-Positive Personen die Summe gebildet und das Ergebnis auf 13 begrenzt, um daraus das Risiko  $R_{\text{CWA-Likely-Positive}}$  zu erhalten:

$$R_{\text{CWA-Likely-Positive}} := \min(\sum t_s, 13)$$

Die Begrenzung des Summanden  $R_{\text{CWA-Likely-Positive}}$  auf 13 hat zur Folge, dass CWA-Likely-Positive Personen nur einen Risikowert  $\leq 13$  bei anderen Personen hervorrufen können. Dies verhindert, dass sich eine Kaskade von CWA-Likely-Positive Personen bilden kann.

Nimmt man diese Begrenzung nicht vor, ist es möglich, dass eine einzelne CWA-Likely-Positive Person bei einer großen Anzahl an Personen einen hohen Risikowert  $\geq 13$  hervorrufen kann, die wiederum bei weiteren Personen hohe Risikowerte  $\geq 13$  hervorrufen können und so weiter. Durch die Begrenzung kann zwar eine CWA-Likely-Positive Person mit einem hohen Risikowert immer noch erhöhte Risikowerte  $\leq 13$  bei anderen hervorrufen, jedoch können die Personen, die dadurch einen erhöhten Risikowert erhalten haben, diesen nur durch sehr lange Begegnungen, z.B. einen ganzen Tag, bei weiteren Personen hervorrufen.

**4.4.4.  $R_{\text{CWA-Likely-Negative}}$ .** Das Risiko  $R_{\text{CWA-Likely-Negative}}$  (Abbildung 2, grüner Kasten) bezieht alle Begegnungen mit Personen aus den Gruppen CWA-Likely-Negative und Not-CWA-User mit ein. Neben dem Unterschied der gezählten Personengruppen, gibt es bei den Risiken  $R_{\text{CWA-Positive}}$  und  $R_{\text{CWA-Likely-Positive}}$  einen Indikator aus historischen Daten, der beschreibt, dass diese Personengruppen bereits an Corona erkrankt sind oder mit einer gewissen hohen Wahrscheinlichkeit später an Corona erkranken werden. Hierbei handelt es sich um Verifizierte Daten in Form von Testergebnissen der CWA-Positive Personengruppe, die in die CWA eingetragen wurden. Bei dem Risiko  $R_{\text{CWA-Likely-Negative}}$  fehlen diese Indikatoren jedoch, obwohl auch von ihnen eine Gefahr aus geht, denn es befinden sich in den beschriebenen Personengruppen Infizierte, deren Erkrankung weder festgestellt noch hochgeladen wurde.

Somit muss das Risiko mit den folgenden Hauptpunkten angepasst werden:

- **Dunkelziffer:** Mithilfe der Dunkelziffer ist eine genauere Abschätzung der Anzahl der mit Corona infizierten deutschen Bürger möglich. Denn Covid-19 hat die Eigenschaft nicht immer symptomatisch aufzutreten, es gibt somit viele Infizierte, die nicht aufgezeichnet werden. Dazu kommt noch, dass nicht

alle Personen ihr Testergebnis in die App eintragen. Die Dunkelziffer ist somit ein Faktor, der beschreibt wie viele Personen mehr erkrankt sind, als positiv getestet wurden. Mithilfe der Dunkelziffer kann somit die Größe der Gruppe Hidden-Positive geschätzt werden. Hidden Positive wurden in Abschnitt 4 definiert und sind alle Personen, die infiziert wurden, jedoch kein positives Testergebnis nachweisen können.

Die Dunkelziffer wurde erstmals durch eine Studie in Kupferzell [23] nachgewiesen und bestimmt. Da die Datenerhebung dieser Studie vom 20. Mai bis 9. Juni 2020 durchgeführt wurde und sich die Testmöglichkeiten seitdem stark entwickelt haben, ist die damals bestimmte Dunkelziffer von Sechs möglicherweise ungenau. Eine Dunkelziffer von Sechs würde allerdings bedeuten, dass die Anzahl der Infizierten um den Faktor Sechs größer ist, als die Anzahl durch Tests aufgezeichneten Infizierten.

- **AppUsage:** Eine Begegnung mit anderen Personen kann nur dann aufgezeichnet werden, wenn diese die Corona Warn App auch installiert haben. Begegnungen mit der Personengruppen Not-CWA-User werden somit nicht aufgezeichnet, sowie auch der Anteil an Infizierten aus dieser Gruppe. Da die derzeitige Nutzung deutlich unter 50% liegt, muss dieser Faktor in die Risikoberechnung mit einfließen, indem die Anzahl der aufgezeichneten Begegnungen nach oben korrigiert wird. Dazu wird die Downloadzahl der CWA (*DownloadCount*) [25] und die Einwohnerzahl Deutschlands (*Population*) [24] genutzt. Aktuell würde die AppNutzung somit ca.  $\frac{1}{3}$  betragen.

$$\text{AppUsage} := \frac{\text{DownloadCount}}{\text{Population}}$$

Wie auch beim Risiko  $R_{\text{CWA-Positive}}$  werden beim Risiko  $R_{\text{CWA-Likely-Negative}}$  die Begegnungen der letzten 14 Tage betrachtet. Obwohl sich jemand nur mit Personen aus den Gruppen CWA-Likely-Negative und Not-CWA-User getroffen hat ist es bei einer hohen Anzahl an Begegnungen wahrscheinlicher infiziert zu werden, als bei einer niedrigen, denn in der Not-CWA-User Gruppe gibt es viele Personen, die infektiös sind. Deswegen wird hierfür die Personengruppe Likely-Negative benötigt.

Hierbei werden, wie beim Risiko  $R_{\text{CWA-Likely-Positive}}$ , Begegnungen mit Gruppenangehörigen aus dem Familienfeature, Abschnitt 4.1 und Begegnungen mit Diagnosis Keys, sowie zusätzlich Begegnungen mit Likely-Positive Personen nicht mit einberechnet, da diese bereits in die Risiken  $R_{\text{CWA-Positive}}$  und  $R_{\text{CWA-Likely-Positive}}$  mit einfließen.

Für dieses Risiko wird hauptsächlich mit einer Positiv-Begegnungs-Wahrscheinlichkeit gerechnet, da man mit dieser die Gruppen Not-CWA-Positive und Hidden-Positive mit einfließen lassen kann. Dabei wird das Übertragungspotenzial nicht wie bei den anderen Risiken  $R_{\text{CWA-Positive}}$  und  $R_{\text{CWA-Likely-Positive}}$  berechnet, da die

Positiv-Begegnungs-Wahrscheinlichkeit einer CWA-Likely-Negative Person zu gering ist. Die Positiv-Begegnungs-Wahrscheinlichkeit mit Begegnungen aus der CWA-Likely-Negative Gruppe wird genutzt, um den Einfluss der Gruppen Not-CWA-Positive und Hidden-Positive zu berechnen und wird unter anderem durch die Dunkelziffer und die Appnutzung bestimmt.

Um das Übertragungspotenzial trotzdem miteinfließen zu lassen werden alle Begegnungen gezählt bei denen die Signalstärke  $att$  unter 73 Dezibel ist, der Abstand somit kleiner als 8 Meter ist und die Begegnungsdauer zwischen einem *LowerBound* und einem *UpperBound* ist.

Da nicht jede noch so kurze Begegnungen aufgezeichnet werden soll, wird ein *LowerBound* auf 2 Minuten gesetzt.

Bei einer einzigen langen Begegnung werden viele Begegnungen mit einer Begegnungsdauer von ca. 15 Minuten aufgezeichnet, da der RPI alle 15 Minuten gewechselt wird. Um dies zu verhindern gibt es zum einen das Familienfeature, welches meist lange Begegnungen mit Freunden bzw. Verwandten von vornherein ausschließt und zum anderen gibt es den *UpperBound*. Für den *UpperBound* soll jede Signalstärke mit einberechnet werden, damit zeitlich kurze Distanzierungen keinen großen Einfluss haben. Für diesen Wert wird 14 Minuten geschätzt, damit Verbindungsfehler keinen großen Einfluss auf die Begegnungsanzahl haben. Dabei werden zwar Begegnungen zwischen 14 und 15 Minuten auch herausgefiltert, jedoch werden Begegnungen mit einer Begegnungsdauer  $\geq 30$  Minuten meist als zwei Begegnungen aufgezeichnet, da nur eine der drei aufgezeichneten Begegnungen eine Dauer von genau 14 bis 15 Minuten hat. Die Anzahl der Begegnungen gleicht sich damit aus. Somit wird jede Begegnung mit Begegnungsdauer zwischen *Lower-* und *UpperBound* gezählt, deren Distanz kleiner als 73 dB ist. Dafür wird die Funktion  $t(x)$  verwendet, die für geltendes  $x$  die Begegnungsdauer einer Begegnung bestimmt.

```
if(t(att < 73 dB)) ≥ lowerBound
&&(att > 0 dB) < upperBound){
Contacts + +
}
```

Um die Wahrscheinlichkeit zu berechnen, bei welcher Begegnungsanzahl, man zu einer hohen Wahrscheinlichkeit infiziert wird, wird zuerst die aktuelle Anzahl Infizierter ausgerechnet, um damit die Positiv-Begegnungs-Wahrscheinlichkeit zu berechnen. Bei der Positiv-Begegnungs-Wahrscheinlichkeit handelt es sich um die Wahrscheinlichkeit, wie wahrscheinlich eine Begegnung infiziert ist.

Im Folgenden wird beschrieben wie *PersonIsInfected* zusammengesetzt ist. *PersonIsInfected*, beschreibt die Wahrscheinlichkeit, dass eine Person infiziert ist. Dafür wird die Anzahl Infizierter durch die gesamte Bevölkerung geteilt. Die aktuelle Anzahl positiv Getesteter wird mit der Dunkelziffer multipliziert, um die Anzahl aktuell Infizierter zu erhalten. Davon wird *currentCWAPositive* subtrahiert, da das Risiko der CWA-Positive Gruppe bereits in  $R_{CWA-Positive}$  mit ein-

fließt. Es werden somit nur noch die positiv Getesteten mit einberechnet, die kein Testergebnis in die CWA eingetragen haben, denn *CurrentCWAPositive* ist die Anzahl an Personen, die ein positives Testergebnis in den letzten 14 Tagen in die CWA eingetragen hat.

$$\text{PersonIsInfected} := \frac{((\text{currentPositiveTested} \cdot \text{Dunkelziffer}) - \text{currentCWAPositive})}{\text{Population}}$$

Aktuell beträgt die *AppUsage* ca.  $\frac{1}{3}$  [25], was dazu führt, dass nur  $\frac{1}{3}$  der Begegnungen aufgezeichnet werden. Die Anzahl aufgezeichneter Begegnungen muss somit um den Faktor drei nach oben korrigiert werden, um die Anzahl wirklicher Begegnungen *CorrectContacts* besser abzuschätzen.

$$\text{CorrectContacts} := \frac{\text{Contacts}}{\text{AppUsage}}$$

Anschließend wird mit einem Beschränkten Wachstum gearbeitet, da von den Personengruppen die in das Risiko  $R_{CWA-Likely-Positive}$  einfließen, ein geringeres Risiko ausgeht.

$$\text{BeschränktesWachstum}(x) := S - S \cdot p^x = (1 - p^x) \cdot S$$

Für die Schranke  $S$  wird der Wert 13 verwendet, da dieser Wert in der aktuellen CWA zu einem erhöhten Risiko führt. Außerdem soll eine Person mit vielen Begegnungen zwar gewarnt werden, allerdings nicht als hoch infektiös gelten.

Für  $p$  wird die Gegenwahrscheinlichkeit von *PersonIsInfected* verwendet, denn umso mehr Personen man trifft, desto näher soll  $R_{CWA-Likely-Negative}$  an der Schranke 13 sein.

Dabei wird für  $x$  *CorrectContacts* eingesetzt. Damit ergibt sich aus  $p^x$  die Wahrscheinlichkeit, nur nicht infizierte Personen getroffen zu haben. Wenn die Wahrscheinlichkeit  $p^x$  gegen Null geht und es somit sehr wahrscheinlich wird mindestens eine infizierte Person getroffen zu haben, geht der Wert für  $R_{CWA-Likely-Negative}$  gegen die Schranke 13.

Im Anschluss wird die zusammengesetzte Funktion, um  $R_{CWA-Likely-Negative}$  zu berechnen, dargestellt.

$$R_{CWA-Likely-Negative} = (1 - (1 - \text{PersonIsInfected})^{\text{CorrectContacts}}) \cdot 13$$

## 4.5. Diskussion

Mithilfe der erweiterten Risikoberechnung kann das Infektionsrisiko besser abgeschätzt werden, da mehr Faktoren betrachtet werden. Weiterhin gilt jedoch: Je mehr Personen die CWA nutzen, desto effektiver ist diese. Dabei wird auf die Privatsphäre der Nutzer geachtet, indem freiwillige, anonymisierte Datensammlungen verwendet werden. Da die Datenspende des Features DK-Server und des Features TRL Update freiwillig ist, benötigt man eine hohe Beteiligung an Nutzern, die Daten teilen. Ohne die Datenspende ist es kaum möglich eine genaue Funktion für den Risikowert  $R_{CWA-Likely-Positive}$  zu bestimmen. Zudem ist die Berechnung

des Risikos  $R_{CWA-Positive}$ , welches aus der aktuellen CWA übernommen wurde, in der erweiterten Risikoberechnung mit eingebunden, wodurch nur ein höherer Risikowert zusammen kommen kann und somit Personen höchstens früher gewarnt werden.

Die erweiterte Risikoberechnung bietet einen hohen Grad an Transparenz bezüglich des Zustandekommens und der Bedeutung des Risikowertes. Außerdem werden durch die feinere Granulierung der Risikoberechnung und Risikoanzeige auch kleinere Risikostufen wahrgenommen. Somit werden schneller Veränderungen des Infektionsrisikos dargestellt, wodurch dem in der CWA angezeigten Risiko mehr vertraut werden kann.

Für die Verwendung des Familienfeatures und des DK-Servers müssen Sicherheitsmechanismen gegen Denial of Service Angriffe implementiert werden, auf die im Rahmen dieser Forschungsarbeit nicht weiter eingegangen wird. Im Vergleich dazu wird bei den im Abschnitt 4.3 vorgestellten Umfragen der Besitz eines eindeutigen Schlüssels vorausgesetzt, der die Teilnahme an einer Umfrage ermöglicht. Der dafür benötigte Schlüssel ist nur im Falle eines positiven Testergebnisses zu erhalten und sorgt dafür, dass kein Spam möglich ist.

Obwohl es zwei weitere Server gibt (DK-Server und Familienfeature Server), auf denen Daten zentral gespeichert werden, bleiben die Risikoberechnung und Aufzeichnungen der RPIs dezentral. Auf den zusätzlichen Servern werden ausgewählte TEKs gespeichert, wie auf dem CWA Server. Zudem werden die Daten auf dem DK-Server und dem CWA Server dazu genutzt, die in Abschnitt 4.4.3 beschriebene Funktion  $f$  zentral zu bestimmen.

Da auch hier auf freiwilliger Basis nur pseudonymisierte Identifier geteilt und auf die Server hochgeladen werden, wird die Privatsphäre der Nutzer nicht weiter, als in der aktuellen CWA, eingeschränkt. Außerdem werden Privatsphäre kritische Daten durch Kryptographie im Familienfeature vor äußeren Eingriffen geschützt und werden somit nur mit der eigenen Gruppe geteilt.

Durch die Installation von zwei weiteren Servern, werden zusätzliche Serverkosten anfallen. Da diese Server einen ähnlichen Aufbau wie der CWA Server haben, werden die Kosten für diese einen ähnlichen Umfang haben.

## 5. Density-Map

Eine Density-Map zeigt an, wo sich bestimmte Datenpunkte auf einer Karte häufen. In der Regel findet sie Anwendung bei Bevölkerungsaufzeichnungen. Im Sinne der erweiterten Corona Warn App möchten wir das Crowd-Sensing Feature benutzen, um in nahezu Echtzeit die Besucherzahl an öffentlichen Orten oder Geschäften aufzuzeichnen und grafisch auf einer Karte darzustellen. Damit kann eine Nutzer eine Density-Map innerhalb der App aufrufen und stark besuchte Gegenden identifizieren und meiden. Ein ähnliches Feature wird bereits vom Unternehmen Google angeboten. Wer in dessen Suchmaschine nach Geschäften in der Umgebung sucht kann neben den Öffnungszeiten meist auch die

live Stoßzeiten und durchschnittliche Aufenthaltsdauer ansehen. Google verwendet hierfür jedoch personalisierte Daten seiner Nutzer, wie z.B. den Standort um diese Aufzeichnung vorzunehmen. Um die Sicherheit und Privatsphäre der Nutzer zu schützen schlagen wir deshalb eine alternative Lösung mithilfe des BLE-Features der Corona Warn App vor.

### 5.1. Lösungsansatz

Das zentrale Problem, mit dem wir uns beim Erstellen einer Density Map beschäftigen ist, wie die anonymen Daten gebündelt und auf einen zentralen Server geladen werden können, ohne die Sicherheits-/Privacy Aspekte signifikant einzuschränken/verletzen. Ein naiver Ansatz wäre beispielsweise, dass jedes Smartphone seine gesammelte Begegnungsanzahl auf einen zentralen Server hochlädt und dieser diese Daten auswertet. Jedoch müssten dafür neben der Standorte der einzelnen Smartphones auch die RPIs aller aufgezeichneten Geräte mit hochgeladen werden, um etwa eine doppelte Zählung bestimmter Endgeräte zu verhindern. Eine derartige Sammlung aller Kontaktdaten (auch wenn es sich bei diesen Daten um anonyme RPIs handelt) auf einem zentralen Server würde jedoch der grundsätzlichen Philosophie der Corona Warn App widersprechen. Diese verfolgte in ihrer Kontaktverfolgung einen möglichst dezentralen Ansatz, um die Privatsphäre der Nutzer zu garantieren. Dementsprechend suchen wir ebenfalls gezielt nach Verfahren, welche weitmöglichst eine dezentrale Lösung verfolgen. Hierfür wurden für die Umsetzung zwei Verfahren erarbeitet, mit welchen das Problem angegangen wird: Eine infrastrukturbasierte Lösung mittels **zentraler Masten**, welche in Echtzeit die BLE-Geräte in ihrem festen Empfangsgebiet aufzeichnen, sowie eine infrastrukturlose Lösung **ohne Masten**.

#### *Infrastrukturbasierte Lösung mit Masten:*

Die Idee zentraler Masten besteht darin, dass diese an stark besuchten Orten/Plätzen, wie z.B. Supermärkten oder Haltestellen, ihre Umgebung scannen und für ein gewisses Zeitintervall alle erfassten Smartphones aufzeichnen. Nach Ablauf des Zeitintervalls werden die erfasste Anzahl an Smartphones zunächst ausgewertet und zu einem Zahlenwert aggregiert. Diese Anzahl an Begegnungen wird dann auf einen zentralen Server hochgeladen. Anders als beim naiven Ansatz, bei dem der zentrale Server die Daten aller Smartphones erfassen würde, ist hier die Aufzeichnung auf tausende bis zehntausende einzelne Masten verteilt. Dadurch gäbe es keinen „Single point of failure“, bei dem beispielsweise die Standorte aller Smartphones preisgegeben werden würde. Da außerdem die Daten dezentral bei den Masten ausgewertet werden und dann erst auf einen Server hochgeladen, reduziert dies den Datenstrom und die Last des zentralen Servers erheblich.

Die Begegnungsaufzeichnung erfolgt analog zur Aufzeichnung mobiler Endgeräte untereinander mittels anony-

mer RPIs per BLE. Somit sind in diesem Aspekt Sicherheit und Privatsphäre gleichermaßen gesichert, wie bei der bisherigen Corona Warn App. Außerdem sollen erst ab einem gewissen Threshold von Aufzeichnungen (z.B. 5 Personen) die Daten auf den Server hochgeladen werden, sowie lediglich die Anzahl der Begegnungen und nicht die einzelnen RPI's der Geräte. Diese Aggregation der Daten stellt sicher, dass keine einzelnen RPIs oder Personen über die Masten hinweg nachverfolgt und damit identifiziert werden können. Abb. 3 zeigt den groben Ablauf eines zentralen Masten.

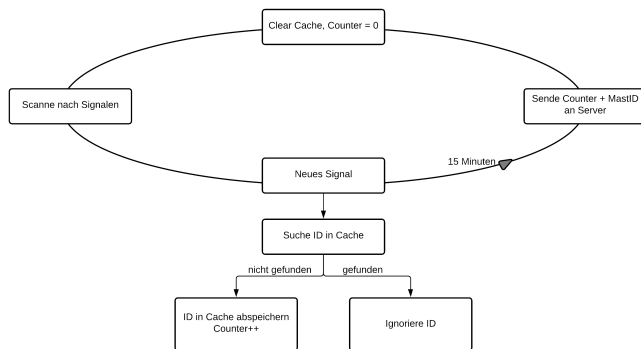


Abbildung 3. Funktionsweise eines zentralen Masten

Wie bereits erwähnt scannt ein Mast seine Umgebung nach Endgeräten und zeichnet alle Begegnungen innerhalb eines Zeitintervalls von etwa 15 Minuten (Lebensdauer eines RPIs) auf. Die Aufzeichnung erfolgt analog zur bisherigen Begegnungsaufzeichnung der Corona Warn App über BLE. Es werden demnach nur die RPIs ausgetauscht und keinerlei sensible Nutzerdaten. Nach der Aufzeichnung überprüft der Mast, ob es sich bei der Begegnung nicht um ein bereits bekanntes Gerät handelt und erhöht dementsprechend den Zähler für die Begegnungsanzahl um eins. Nach Ablauf des 15 Minuten Intervalls lädt der Mast die gesammelte Anzahl an Begegnungen auf einen zentralen Server hoch, setzt seine Aufzeichnungsdaten zurück und fängt wieder von vorne mit seiner Begegnungsaufzeichnung an.

Da ein Mast unter Umständen einen größeren Bereich abdecken muss als ein normaler Nutzer, sollte das Signal (z.B. mit Repeatern) unter Umständen verstärkt werden. Andernfalls müsste ein Ort in mehrere Bereiche unterteilt werden, die dann von einzelnen Masten aufgezeichnet werden können.

Hier wird auch deutlich, dass sich die Aufzeichnungsradien der einzelnen Masten möglichst nicht überschneiden sollte, da sonst ein Smartphone von mehreren Masten gleichzeitig aufgenommen wird. Eine mögliche Lösung für dieses Problem wäre, die BLE-Signale der Masten mit bestimmten Tags zu versehen, sodass ein Smartphone erkennt, falls es sich im Radius von mehreren Masten aufhält und dementsprechend nur einem Mast seine Daten sendet.

Da die Installation zentraler Masten an jedem belebteren Platz ein aufwändiges und kostspieliges Unterfangen sein kann schlagen wir vor, dass registrierte/zertifizierte mobile

Endgeräte ebenfalls als zentraler Mast fungieren können. Hierfür müssen jedoch Einschränkungen gelten, sodass die Funktion als zentraler Mast auf beispielsweise einem Smartphone nicht missbraucht wird. Folgende Regeln und Restriktionen sollten für alle Masten gelten:

- Masten können selbständig von Dritten erworben oder im Falle eines Smartphones, welches als zentraler Mast fungieren soll, die zertifizierte App installiert werden. Jedoch muss vor der Nutzung des Features das Gerät von einer offiziellen Behörde autorisiert werden. Damit darf dann etwa ein Restaurantbetreiber, über eine zentrale Behörde registriert, seine Einrichtung mit seinem Smartphone aufzeichnen.
- Zentrale Masten müssen während ihrer Nutzung dauerhaft ihren Standort aktivieren und können nur im Umfeld ihres registrierten Bereiches (z.B. Ort des Restaurants) als Mast fungieren. Verlässt ein Mast den für ihn zugewiesenen Aufzeichnungs-bereich sollt dieser seine Aufzeichnung automatisch stoppen. Dies verhindert, dass das Feature für unerlaubte Zwecke benutzt wird.
- Verletzung der Eigenschaften sollte von anderen Nutzern gemeldet werden können. Beispiele für Verletzung der Eigenschaften wäre etwa eine Person, die in einem unerlaubten Gebiet (z.B. Wohngebiet) einen Masten betreibt oder ein Mast, der viele Begegnungen aufzeichnet, obwohl sich keine Person im Umkreis befindet.

Masten, die ihre Begegnungen aufgezeichnet haben, laden die aggregierten auf einen zentralen Server hoch. Dieser verarbeitet die gesammelten Daten aller Masten und erstellt die Density Map, welche dann vom Nutzer auf seinem Endgerät abgerufen werden kann. Abb. 4 zeigt einen grafischen Ablauf zu Erstellung der Density Map mittels zentraler Masten.

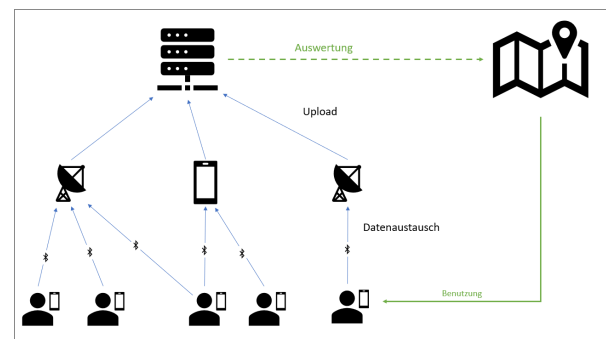


Abbildung 4. Übersicht: Erstellen der Density Map

Wie in Abschnitt 2.3 beschrieben, besitzen die RPIs eine Lebensdauer von etwa 15 Minuten. Dementsprechend muss beachtet werden, dass Smartphones auch während des Aufenthaltes im Umkreis eines Masten ihren RPI regelmäßig ändern, wodurch ein Endgerät vom Mast als zwei



unterschiedliche Begegnungen erkannt werden kann. Der Mast zeichnet dementsprechend zu viele Begegnungen auf und muss seine gemessene Anzahl an Begegnungen nach unten korrigieren, um die Tatsächliche Begegnungsanzahl herauszufinden. Wie stark der Wert nach unten korrigiert werden muss, hängt hierbei von der Durchschnittlichen Verweildauer eines Smartphones am Masten ab. Beträgt die durchschnittliche Verweildauer beispielsweise 15 Minuten (was dem Maximalwert entspricht, da sich nach 15min. der Mast zurücksetzt) wird innerhalb der Aufenthaltsdauer am Mast jedes Endgerät auf jeden Fall einmal seinen RPI ändern und als neue Begegnung vom Mast aufgezeichnet werden. In diesem Fall würde der Mast doppelt so viele Begegnungen aufzeichnen als es in der Realität sind. Ist die durchschnittliche Verweildauer hingegen kurz im Verhältnis zum Änderungsintervall der RPIs (z.B. 3 Minuten) ist die Wahrscheinlichkeit, dass sich in diesem Zeitintervall der RPI eines Smartphones ändert, sehr klein, was weniger mehrfache Aufzeichnungen zur Folge hat. Wie stark ein aufgezeichneter Wert nach unten korrigiert werden sollte wird mit folgender *ID-basierter Korrekturformel* errechnet:

$$IDs = \frac{RPIs}{(1 + p)} \quad (1)$$

Dabei ist „RPIs“ die gemessene Anzahl an RPIs des Masten und „IDs“ die voraussichtlich tatsächliche Anzahl an Endgeräten. Die Wahrscheinlichkeit  $p$  einer doppelten Registrierung ist definiert durch:

$$p = t_V / t_Z$$

wobei  $t_Z$  die Dauer eines 15-Minuten Zyklus ist und  $t_V$  die durchschnittliche Verweildauer während dieses Zyklus. Vorausgesetzt wir kennen die „IDs“ können wir mittels  $p$  ausrechnen, wie viele der ID's doppelt gezählt werden mit:  $IDs \cdot p$ . Daraus ergibt sich für die gemessenen „RPIs“ unter Annahme, dass die ID's bekannt sind, folgende Formel:

$$RPIs = IDs + IDs \cdot p$$

Diese Formel nach „IDs“ umgeformt ergibt die oben genannte Korrekturformel zur Berechnung der tatsächlichen Anzahl an Endgeräten und wird im Folgenden als ID-basierte Formel bezeichnet.

Eine Übersicht, inwiefern der gemessene Wert mithilfe der ID-basierten Korrekturformel korrigiert wird, wird in Abb. 5 aufgezeichnet. Auf der x-Achse befindet sich die durchschnittliche Verweildauer der Smartphones in Mastreichweite in Minuten, während die y-Achse die Anzahl an Smartphones (ID's) angibt. Wir gehen davon aus, dass der Mast in jedem Fall genau 100 ID's aufgezeichnet hat (blauer Graph). Je nach Verweildauer muss der gemessene Wert nun mittels Formel nach unten korrigiert werden (orangener Graph).

Eine Alternative zur ID-basierten Korrekturformel wurde zu Beginn der Forschungsarbeit erarbeitet und basiert darauf, den Anteil an doppelten Begegnungen von der gemessenen Anzahl abzuziehen. Der Anteil wird dabei an Hand der Wahrscheinlichkeit einer doppelten Registrierung berechnet, welche wiederum von der durchschnittlichen Verweildauer

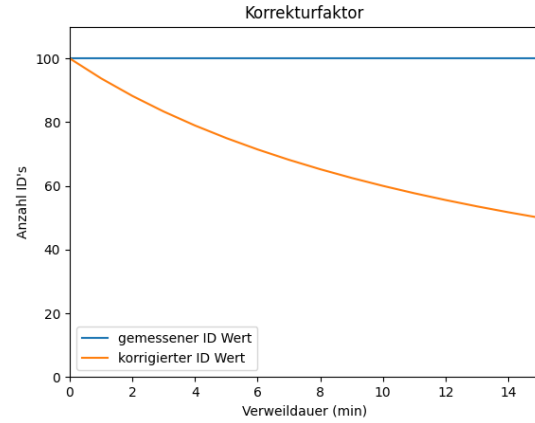


Abbildung 5. Korrekturfaktor in Abhängigkeit der durchschnittlichen Verweildauer

abhängt.

Bei einer Wahrscheinlichkeit  $p$  einer doppelten Registrierung müssen  $p/2$  Begegnungen von den RPIs abgezogen werden, um so den realen Wert zu erhalten.

Diese *alternative Korrekturformel* ist somit wie folgt definiert:

$$IDs = RPIs * (1 - p/2) \quad (2)$$

Auf Grund des Ansatzes, dass die Anzahl der IDs unbekannt ist und die Anzahl der RPIs bekannt, wird diese Formel im Folgenden als RPI-basierte Formel bezeichnet.

Beide Formeln werden im Abschnitt *Simulation* evaluiert und miteinander verglichen.

#### Ohne Masten:

Beim Ansatz ohne zentrale Masten besteht die Idee darin, dass jedes Smartphone als „mobiler Mast“ arbeitet und die Funktion eines Masten übernehmen kann. Auch bei diesem Ansatz möchten wir zunächst die Daten dezentral bündeln und aggregieren, um den bereits erwähnten Privatsphäre- und Sicherheitsrisiken des naiven Ansatzes entgegenzuwirken.

Begegnen sich zwei Smartphones tauschen diese, wie bisher in der aktuellen Corona Warn App, ihre RPIs und Daten untereinander aus. Zusätzlich wird nun eines der beiden Geräte als Mast ausgewählt (Abb. 6). Dieses zeichnet die Begegnung der beiden untereinander auf und übernimmt alle bisherigen Aufzeichnungen des anderen Gerätes in diesem Zeitintervall. Die Aufzeichnungsdaten werden also transitiv bis Ende des Intervalls weitergegeben. Nach Ablauf des Zeitintervalls lädt das Smartphone, welches als mobiler Mast ausgewählt wurde, die aggregierte Begegnungsanzahl mit seinem aktuellen Standort auf den zentralen Server.

Bei dieser Methode können natürlich nur Smartphones, welche ihren Standort aktiv haben, als mobiler Mast arbeiten. Im besten Fall hat jedes Handy diese Funktion aktiv. Smartphones, die ihren Standort nicht aktiv haben, werden können von anderen Smartphones mit aktiven Standort auf-

gezeichnet, jedoch selbst nie als mobiler Mast ausgewählt werden.

Anders als bei zentralen Masten kann nun an jedem Ort, an dem sich ein Smartphone mit der Corona Warn App befindet, eine Aufzeichnung für die Density-Map erfolgen. Die Datenerfassung der Density-Map ist somit nicht mehr von vorinstallierten zentralen Masten abhängig.

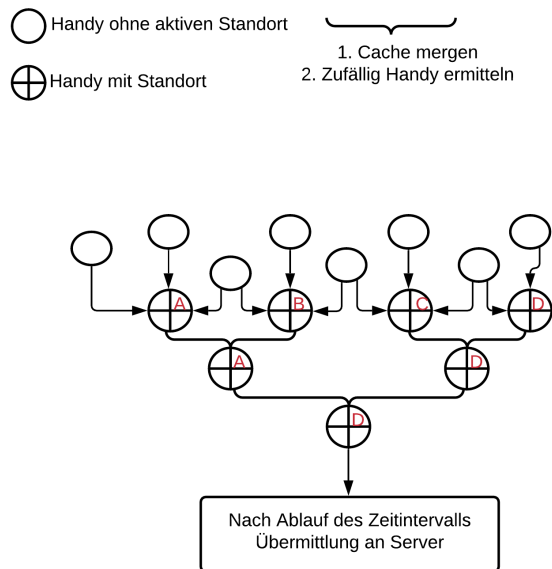


Abbildung 6. Smartphones als Mobile Masten

Beim mobilen Ansatz können diverse Probleme auftreten, welche gelöst werden müssen: Zum Beispiel können zwei Smartphones A und B ihre gesamten Daten immer an das jeweilige andere Handy überreichen, bevor ihr 15 Minuten Intervall abgelaufen ist. Somit bleiben die gesammelten Daten für unendlich lange Zeit lokal auf A oder B gespeichert, ohne dass sie jemals hochgeladen werden. Eine Lösung für dieses Problem wäre immer das Smartphone als mobilen Mast auszuwählen, das näher am „Reset“, also am Ablauf des 15-Minuten Intervalls ist.

Ein weiteres Problem bei mobilen Masten besteht darin, dass sie sich innerhalb des Zeitintervalls über längere Strecken bewegen können während sie ihre Begegnungen aufzeichnen. In diesem Fall entsprechen Ort der Aufzeichnung eines anderen Gerätes und Ort des Hochladens der Daten nach Ablauf des Zeitintervalls nicht überein und das Bild der Density-Map wird verzerrt. Dieses Problem wird teilweise bereits durch die Corona App selbst gelöst, welche Begegnungen erst dann aufzeichnet, sobald sich Geräte über etwas längere Zeit in Kontakt stehen. Bei sich schnell bewegenden Smartphones müssten sich demnach anderer Smartphones mit gleicher Geschwindigkeit in dieselbe Richtung bewegen, damit eine Begegnung aufgezeichnet wird. Das Verkürzen des Zeitintervalls für Uploads auf beispielsweise 5 Minuten würde diesem Problem ebenfalls entgegenwirken,

da in kürzeren Intervallen die Smartphones sich im Schnitt weniger weit bewegen können.

Eine weitere Lösung, die wir diesbezüglich vorschlagen ist ein sogenannter „Discount Faktor“. Dieser gewichtet indirekte Verbindungen eines mobilen Masten schwächer in der Auswertung der Density Map.

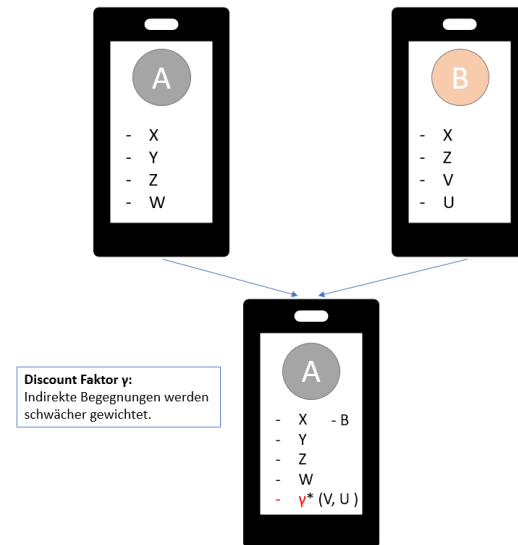


Abbildung 7. Discount Faktor für Mobile Masten

In Abbildung 7 begegnen sich die beiden Smartphones A und B. A hat bereits die RPIs X, Y, Z und W aufgezeichnet und B die RPIs X, Z, V, U. Nun tauschen die beiden Smartphones ihre RPIs gemäß Funktionsweise der Corona Warn App aus. Außerdem wird A als mobiler Mast ausgewählt und übernimmt die Aufzeichnungen von B für die Density-Map. Die RPIs V und U wurden nur von B aufgezeichnet. Dies bedeutet, dass sie sich nicht im direkten Umkreis von A befinden können. B muss sich vor der Begegnung mit A also an einem anderen Ort befunden haben und sich zu A bewegt haben (aus Sicht von A). Dementsprechend werden die Begegnungen RPIs U und V mit dem Discount Faktor  $\gamma$  versehen. Der Discount Faktor kann auf verschiedene Arten realisiert werden. Er könnte beispielsweise einen Zahlenwert im Intervall  $(0, 1)$  annehmen. Somit würden indirekte Begegnungen mit dem Faktor multipliziert und nicht voll gewertet werden. Im Beispiel wären die zwei indirekten Begegnungen U und V mit gewähltem  $\gamma = 0.6$  versehen und nur als  $2 \cdot 0.6 = 1.2$  Begegnungen aufgezeichnet werden.

Eine andere Lösung wäre beim Auswerten der Density-Map indirekten Begegnungen mit anderen Standorten zu versehen. Während alle direkten Begegnungen von A im unmittelbaren Umkreis seines Standortes auf der Map angezeigt werden U und V (beim Hochladen aggregiert in 2 Begegnungen) in einem größeren Radius um A angezeigt. In diesem Fall wäre der Discount Faktor nur eine Markierung und müsste keinen numerischen Wert erhalten.

Der Discount Faktor kann ebenfalls transitiv weitergegeben werden. Werden etwa die beiden indirekten

Begegnungen  $U$  und  $V$  an ein weiteres Smartphone  $C$  übergeben, so werden sie bei  $C$  doppelt mit dem Discount Faktor versehen.

#### Hybridlösung:

Um die Effekte beider Ansätze nutzen zu können und die Effizienz weitmöglichst zu erhöhen, schlagen wir eine Hybridlösung aus zentralen und mobilen Masten vor. Die Umsetzung hierfür ist relativ simpel. Wie oben beschrieben, werden an häufig besuchten Orten und Plätzen zentrale Masten aufgestellt, um Daten für die Density-Map zu sammeln. Überall dort, wo jedoch keine solchen zentralen Masten existieren, greift der dezentrale Ansatz und die Mobiltelefone fungieren als mobile Masten. Der im infrastrukturenbasierten Ansatz beschriebene spezielle Tag für RPIs der Masten kann hier verwendet werden, damit ein Smartphone erkennt, ob er sich zurzeit im Umkreis eines zentralen Masten befindet oder nicht. Mittels Hybridlösung könnte man die Vorteile beider Ansätze vereinen und die Nachteile ausgleichen. Zentrale Masten können zuverlässiger einen festen Ort aufzeichnen, da für diese Funktion optimiert sind und sich dauerhaft an diesem Ort befinden. Mobile Masten hingegen können an jedem beliebigen Ort eine Ansammlung vieler Smartphones und damit vieler Menschen erkennen und auch eine Bewegung großer Menschenmassen aufzeichnen. Sie sind demnach flexibler, was den Standort angeht.

Das letzte zentrale Problem, welches nicht nur unsere Lösungsansätze sondern auch die Corona Warn App im allgemeinen stark betrifft ist, dass nicht jede Person ein Smartphone besitzt und die Corona Warn App auf seinem Smartphone installiert hat. Dementsprechend liegen die aufgezeichneten Werte durch die Masten stets unter der tatsächlichen Anzahl der Personen, die sich am entsprechenden Ort befinden. Laut aktuellem Stand (Mai 2021) benutzen etwa 27,8mio (33%). [25] Unter der Annahme, dass die Nutzer der App auf die Standorte bezogen gleichverteilt sind bedeutet dies, dass man die gemessene Anzahl etwa verdreifachen muss, um die tatsächliche Anzahl an Personen zu erhalten.

## 5.2. Simulation

Um die im Lösungsansatz erarbeiteten theoretischen Überlegungen zu verifizieren und deren Umsetzbarkeit nachzuweisen, haben wir eine Simulation des mastenbasierten Ansatzes implementiert und evaluiert.

Eine reale Umsetzung des Features mit echten Masten und Smartphones erwies sich als zu umfangreich und im Rahmen unseres Forschungsprojektes zeitlich nicht umsetzbar.

Die Simulation beschränkt sich auf die Umsetzung mittels zentraler Masten. Als Grundlage diente uns hierfür die Simulation „COVID-19 Spread Simulator“ von Miguel Ángel Durán [6].

### 5.2.1. Simulation der RPI-basierten Formel.

Wie bereits in Abschnitt 5.1 beschrieben, existiert eine bestimmte Wahrscheinlichkeit einer doppelten Registrierung innerhalb eines Mastes, wodurch der gemessene Wert stets angeglichen werden muss, um den realen Wert zu approximieren.

Im ersten Teil unserer Simulationen testen wir zunächst die *RPI-basierte Korrekturformel*.

$$\text{IDs} = \text{RPIs} * (1 - p/2)$$

und im zweiten Teil dieses Abschnitts die *ID-basierte Korrekturformel*

$$\text{IDs} = \frac{\text{RPIs}}{(1+p)}$$

Der Grund, weshalb zuerst die RPI-basierte Korrekturformel getestet wird ist, dass diese Formel als erstes innerhalb der Forschungsarbeit entstanden ist und erst nach der Auswertung unserer initialen Simulationsergebnisse die Ungenauigkeit dieser Formel aufgefallen ist. Daraufhin wurde die ID-basierte Korrekturformel erstellt.

Das Ziel der Simulation ist es festzustellen, wie nah der durch beide Formeln berichtigte Wert an der tatsächlichen Anzahl an Handys liegt.

#### Aufbau:

Wir simulieren 1000 Handys, in Abbildung 8 als Punkte dargestellt, mit zufälliger Bewegung und Position über ein festgelegtes Gebiet wandern. Die Geschwindigkeit und Richtung ihrer Bewegung wird durch ein Random-Direction-Model bestimmt.

Auf diesem Gebiet sind vier Masten, in Abbildung 8 dargestellt durch größere Kreise, welche persistent speichern, welche Handys in ihrer Reichweite liegen und wie lange dies der Fall ist.

Jeder Kreis hat eine sich periodisch verändernde, temporäre ID (RPI) und eine individuelle, permanente ID. Dies ermöglicht es uns, das Zählen von RPI's zu simulieren und gleichzeitig mit der korrekten Anzahl zu vergleichen.

Die durchschnittliche Verweildauer wird immer aktuell berechnet nach Ablauf des Zyklus' eines Mastes. Dazu wird während des Zyklus zu jeder ID eine Zeitdauer gemessen und aus diesen der Durchschnitt ermittelt.

Zur Laufzeit der Simulation speichern die Masten eintretende mobile Objekte ab und zählen ab dem Zeitpunkt des Eintritts einen Zähler hoch, welcher die Dauer des Aufenthalts im Radius des Mastes misst. Am Ende eines Zyklus werden diese Zähler addiert und durch die gemessene Anzahl der Handys geteilt.

Zwischenzeitliches Verlassen des Radius des Mastes setzt den Zähler nicht zurück. Stattdessen wird nach Wiedereintritt an Hand der ID der alte Zähler weitergeführt. Die Zähler messen die Zeit in Ticks, welche die generelle Zeiteinheit der Simulation ist. Ein Durchlauf einer Simulation dauert 16000 Ticks, während der die Masten in einem Zyklus von 200 Ticks ihre Aufzeichnungen durchführen.

Je höher die durchschnittliche Verweildauer  $t_V$  des Zyklus ist, desto höher ist die Wahrscheinlichkeit  $p$ , dass ein einziges Handy doppelt registriert wird.

Zu Verifizierungszwecken haben wir die durchschnittliche Verweildauer mit Hilfe der permanenten IDs berechnet, was den tatsächlichen Wert liefert.

Benutzt man stattdessen die RPIs, verringert sich die durchschnittliche Verweildauer, da Handys doppelt gezählt werden können und dadurch fälschlicherweise mit mehreren Zählern gerechnet wird, die dabei geringe Zeitdauern liefern. Die durchschnittliche Verweildauer nimmt also ab.

Durch eine geringere Verweildauer nimmt auch die Genauigkeit des korrigierten Wertes ab. Da das Ziel unserer Simulation die Abschätzung der Güte unseres angeglichenen Wertes ist, rechnen wir mit der tatsächlichen Verweildauer, um Nebenfaktoren zu minimieren.

Die Berechnung der durchschnittlichen Verweildauer anhand von RPIs würde jedoch bei einer Umsetzung in der Praxis verwendet werden, da aus Datenschutzgründen nicht die IDs und ihre Verweildauer gemessen werden.

Die Simulation bietet zudem die Möglichkeit, Social Distancing durchzuführen. Um den Effekt dieser Maßnahme darzustellen, wurde die Geschwindigkeit von Handys zeitweise auf Null reduziert. Dies führt dazu, dass nur einige wenige Kreise sich überhaupt bewegen, die Anzahl an Ein- und Austritten bei Masten also stark reduziert ist.

#### Ergebnisse:

Abbildung 8 stellt die gesammelten Daten in einem Schaubild dar, ohne dass Social Distancing benutzt wird. Hierbei ist der rote Graph die durchschnittliche Verweildauer, der grüne Graph der gemessene Wert, der orangene Graph der korrigierte Wert und der blaue Graph der reale Wert.

Auf der x-Achse befindet sich die Zeit, die den gesamten Verlauf der Simulation darstellt.

Mit Social Distancing (Abb. 9) ist die durchschnittliche Verweildauer wesentlich höher und der angegliche Wert liegt näher an dem gemessenen Wert als in Abb. 8.

#### Schlussfolgerung:

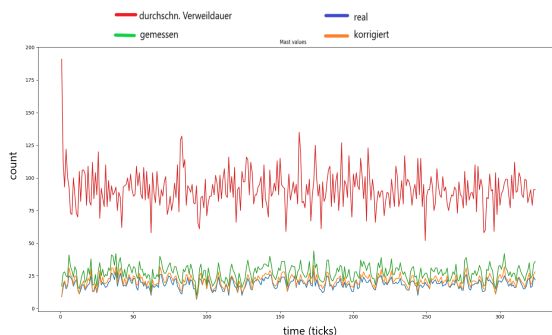


Abbildung 8. Ergebnisse ohne Social Distancing

Die gesammelten Daten belegen, dass die durch die RPI-basierte Korrekturformel berechnete Anzahl (orangener Graph) wesentlich näher an der tatsächlichen Zahl (blauer Graph) liegt als die gemessene Menge (grüner Graph). Es fällt jedoch auf, dass in fast allen Fällen die berechnete

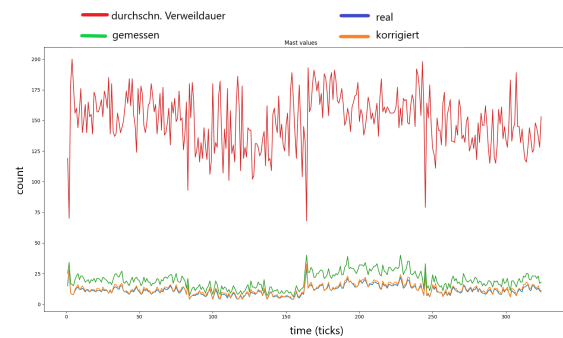


Abbildung 9. Ergebnisse mit Social Distancing

Zahl immer noch größer ist als die tatsächliche. Nur in einzelnen Fällen wird die Zahl zu niedrig abgeschätzt. Hinzu kommt die Beobachtung, dass je höher die durchschnittliche Verweildauer, desto geringer die Differenz zwischen korrigierten und realen Wert. Die Genauigkeit unserer Abschätzung wird demnach besser, je höher die Verweildauer ist.

Aus dem Ergebnis lässt sich die Erkenntnis gewinnen, dass die RPI-basierte Korrekturformel nicht optimal ist, da der berechnete Wert stets zu groß ist, sich grundlegend jedoch im Vergleich zum gemessenen Wert ein Wert näher an der realen Anzahl berechnen lässt.

#### 5.2.2. Simulation der ID-basierten Korrekturformel.

Um die ID-basierte Korrekturformel zu testen, benutzen wir dieselbe Simulation wie zuvor, mit der Berechnung des korrigierten Wertes als einzigen Unterschied.

Mit Hilfe dieser Formel ergeben sich die in Abbildung 10 und Abbildung 11 gezeigten Werte.

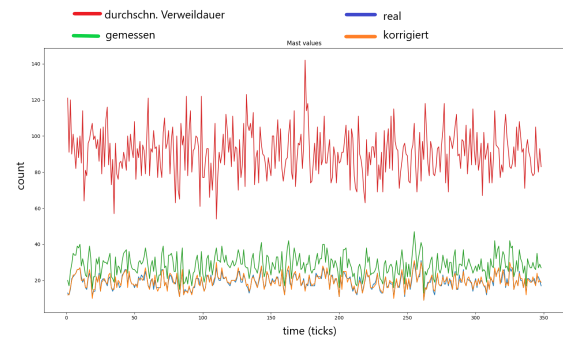


Abbildung 10. Neue Ergebnisse ohne Social Distancing

#### Ergebnisse:

Wie auch im ersten Abschnitt lässt sich feststellen, dass der durch die ID-basierte Formel korrigierte Wert immer näher am realen Wert liegt als der gemessene Wert. Die Approximation ist also auch hier generell erfolgreich. Zusätzlich dazu sind die korrigierten Werte in dieser Simulation in gleichen Anteilen etwas zu hoch oder zu niedrig und nicht standardmäßig zu hoch abgeschätzt, wie

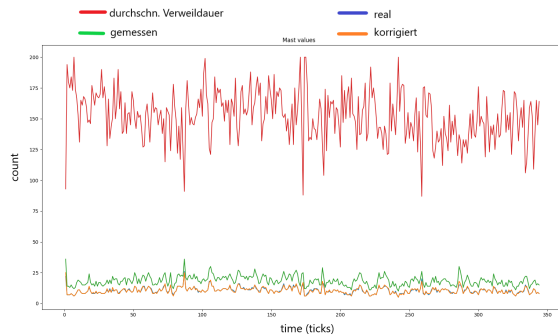


Abbildung 11. Neue Ergebnisse mit Social Distancing

dies in der vorherigen Simulation mit der RPI-basierten Formel der Fall war.

Der durchschnittliche Unterschied zwischen korrigiertem und realen Wert ist zudem niedriger mit dieser Formel als bei der Verwendung der RPI-basierten Korrekturformel.

#### *Schlussfolgerung:*

Durch den geringeren Unterschied zwischen korrigiertem und realen Wert ist die ID-basierte Korrekturformel besser als die RPI-basierte Korrekturformel für die Erfüllung unseres Ziels, die Anzahl von individuellen Handys innerhalb eines Mastes zu berechnen.

### 5.3. Bewertung

Das Feature der Density-Map erzielt im Hinblick auf die Eindämmung des Coronavirus keinen besonders hohen Nutzen im Vergleich zu anderen vorgestellten Erweiterungen.

Die Hauptidee hinter der Implementierung dieser Erweiterung innerhalb der Corona Warn App ist die, Orte großer Menschenmassen bewusst zu meiden und damit die Ausbreitung des Virus einzudämmen.

Durch einen vorausschauenderen Lebensalltag oder Kontaktbeschränkungen seitens der Regierung kann dieser Effekt auch erreicht werden.

Jedoch ist die Density-Map ein Feature, welches auch außerhalb der Pandemiezeit effektiv genutzt werden kann und einen Nutzen bringt.

Gerade in Zeiten, in denen keine größeren Einschränkungen gelten, ist es vermutlich umso wichtiger, weiterhin bewusst Abstand zu größeren Ansammlungen zu halten, um der Ausbreitung einer weiteren Infektionswelle vorzubeugen. Deshalb trägt die Density-Map vor allem in Zeiten der Lockerung der Beschränkungsmaßnahmen einen guten Teil dazu bei, diesen Zustand zu erhalten.

Unabhängig vom Kontext der Pandemie hat die Erstellung einer Density-Map, welche auf der anonymisierten Verfolgung von Kontakten basiert, ebenfalls viele Verwendungen.

So lässt sich zum Beispiel messen, wie viele Plätze in einem Zug bereits belegt sind, oder ob ein öffentlicher

Platz wie ein Spielplatz überfüllt ist. Die erarbeiteten Formeln lassen sich generell in Echtzeit-Karten verwenden, wo durch anonyme Datensammlung nach dem Prinzip des Crowd-Sensing fehlerhafte Werte entstehen können.

Die Auswertung der gesammelten Daten und Korrigierung mittels der in dieser Arbeit ermittelten Formel wäre also auch in diesen Bereichen anwendbar und hat somit Relevanz über den Zeitraum der Pandemie hinaus.

Auch die Tatsache, dass bereits ein vergleichbares Feature von Google existiert, stellt den Nutzen des Features natürlich in Frage.

Jedoch wird das Feature dem Ziel, eine im Hinblick auf Privatsphäre fortgeschrittenere Alternative zu entwickeln, gerecht. Die Density-Map sammelt in allen vorgestellten Lösungsansätzen keine sensiblen Nutzerdaten und verfolgt einen weitestgehend dezentralen Ansatz.

Trotzdem ist logisch, dass man bei einem Feature, welches die Ansammlung größerer Menschenmassen tracken möchte, nicht zu 100% anonym ist.

Vielmehr greift hier der Effekt der K-Anonymity, dass man nur eine von vielen Personen ist, sowie der Fakt, dass die eigene Identität hinter einem temporären RPI verborgen ist. Trotzdem ist es gerade im infrastrukturlosen Ansatz ohne zentrale Masten möglich, Ansammlungen von Menschen im privaten Raum (wie etwa Partys) zu tracken.

Auch im Hinblick auf Akzeptabilität könnte das Feature auf Probleme stoßen.

Unwissende Nutzer, die sich nicht mit der Umsetzung des Features auskennen, könnten davon ausgehen, dass man mithilfe der Density-Map dauerhaft getrackt werden kann. Ein ähnliches Phänomen gab es bereits, als die Corona Warn App das erste Mal vorgestellt wurde.

Dagegen hilft eine klare Aufklärung darüber, wie das Feature im Konkreten funktioniert, sowie eine freiwillige Teilnahme am Benutzen der Density-Map.

Im Aspekt der Sicherheit stellt die Density-Map kein höheres Risiko als die bisherige Corona Warn App dar.

Es existieren keine sensiblen Nutzerdaten, die durch einen Angriff auf das System preisgegeben werden könnten. Durch die dezentrale Auswertung und Aggregation der Daten existiert außerdem kein Single Point of Failure.

Es könnte natürlich passieren, dass einzelne Masten gehackt werden, wodurch beispielsweise die Begegnungsaufzeichnung für diesen Ort verfälscht werden könnte.

Dies stellt jedoch weder ein Risiko für alle anderen Masten dar, noch für einen Nutzer, der sich in Nähe eines gehackten Masten befindet, da wie bereits genannt nur anonyme und verschlüsselte Daten mit dem Masten ausgetauscht werden.

Für die Umsetzung des Density-Map Features ergibt sich ein größerer Aufwand zur Installation der zentralen Masten. Um eine gute deutschlandweite Abdeckung zu erreichen, müssten hierfür vermutlich zehn- bis hunderttausende Geräte zu Verfügung gestellt werden, die als zentrale Masten fungieren könnten.

Beim infrastrukturlosen Ansatz würde dieser Punkt jedoch



wegfallen.

Alles in Allem stellt das Density-Map Feature eine gute Erweiterung der Corona Warn App, sowohl im Hinblick auf die Eindämmung des Virus, als auch für eine unabhängige Nutzung außerhalb der Pandemie, dar. Dabei schafft sie es, das BLE-Feature der App erweitert zu nutzen und dabei den bisher gesetzten Privatsphärestandards gerecht zu werden.

## 6. Zusammenfassung und Ausblick

Die Corona Warn App ist ein essenzieller Bestandteil der Strategie zur Eindämmung der Corona-Pandemie. Im Rahmen dieser Forschungsarbeit wurden verschiedene Erweiterungen zur aktuellen Corona Warn App erläutert und auf unterschiedliche Kriterien hin beurteilt. Dabei wurden zunächst einige kleinere Nebenfeatures betrachtet, bevor auf die zwei Hauptfeatures der *Erweiterten Risikoberechnung* und der *Density-Map* näher eingegangen wurde. Jedes der Features wurde im Anschluss nach den Kriterien: Privatsphäre, Nutzen, Umsetzbarkeit, Akzeptabilität, Sicherheit und benötigte Daten bewertet.

Einige der von uns vorgestellten Nebenfeatures wurden im Verlauf der Arbeit von anderen Parteien in ähnlicher Form bereits umgesetzt oder sind in naher Zukunft zu erwarten. So existieren bereits eine Lotteriefür Geimpfte in den USA als eine Umsetzung des Gamification-Features, die „Darf ich das?“-App für ortsbegleitende Maßnahmen und Beschränkungen, sowie ein demnächst erscheinender digitaler Impfpass für die CWA. Dies bestätigt, dass Bedarf nach den Features bestand und diese eine sinnvolle Erweiterung der CWA darstellen.

Im Folgenden wurde die erweiterte Risikoberechnung und darin einfließende Features im genaueren betrachtet. Hierbei handelt es sich um das Familienfeature, Feature DK-Server und Feature TRL Update. Im Rahmen des Familienfeatures können Gruppen in der CWA angelegt werden, um längere Begegnungen mit der eigenen Familie nicht als mehrere unterschiedliche Begegnungen aufzuzeichnen. Das Feature DK-Server sammelt die TEKs der Personen, die als direkte Kontaktpersonen von Infizierten in der CWA identifiziert werden, um diese in der Risikoberechnung stärker zu gewichten. Das dritte Teilfeature TRL Update behandelt eine Umfrage, damit positiv getestete Auskunft über ihren Krankheitsverlauf geben können. Das TRL gibt dabei an, wie infektiös eine Person im derzeitigen Krankheitsverlauf ist, weshalb genauere Informationen über den Verlauf sehr wichtig sind. Mit diesen Daten lässt sich das aktuelle TRL genauer berechnen.

Die drei Features werden genutzt, um die aktuelle Risikoberechnung mit zwei weiteren Summanden zu erweitern. Die Summanden lösen aktuelle Probleme der CWA, indem sie alle Personengruppen mit einbezieht, die in Abschnitt 4 definiert werden. Außerdem wird aufgrund der geringen Nutzerzahl der CWA von 33% die Anzahl der aufgezeichneten Begegnungen nach oben korrigiert, um die tatsächliche Begegnungszahl abzuschätzen. Zudem wird

die Dunkelziffer genutzt, um auch Begegnungen mit nicht positiv getesteten ein Risiko zuzuschreiben. Des Weiteren wird die Risikoberechnung transparent und durch zwei neue Farben feingranularer dargestellt.

Mithilfe der erweiterten Risikoberechnung kann das Infektionsrisiko besser abgeschätzt werden. Außerdem steigert dieses Feature die Effizienz und Nutzen der CWA, da die transparente und feingranularere Risikoberechnung schneller und genaueres Feedback zum aktuellen Verhalten bietet. Trotz niedriger Beteiligung an Nutzern wird der aktuelle Risikowert nicht unterschritten, er würde sich allerdings nicht drastisch verändern. Obwohl zwei weitere Server zentral Daten sammeln, wird durch Anonymisierung die Privatheit und Sicherheit garantiert. Die Risikoberechnung findet weiterhin dezentral auf den Smartphones der Nutzer statt.

Das zweite Hauptfeature beschäftigt sich mit der Density-Map. Sie gibt eine grafische Übersicht darüber, an welchen Orten sich derzeit eine Vielzahl von Menschen aufhalten, damit diese gezielt gemieden werden können. Hierfür wurden unterschiedliche Ansätze überlegt, die das Problem der dezentralen und anonymen Datenerfassung angehen: der infrastrukturbasierte Ansatz mittels zentraler Masten, sowie der infrastrukturlose Ansatz ohne Masten.

Im infrastrukturbasierten Ansatz werden zentrale Masten an stark besuchten Orten und Plätzen aufgestellt, die mithilfe der BLE-Funktion die Anzahl aller erfassten Smartphones aufzeichnen. Die aufgezeichnete Anzahl wird im Intervall von 15 Minuten zunächst beim Mast aggregiert und dann auf einen zentralen Server geladen, auf dem die Daten aller Masten verarbeitet und eine Density-Map erstellt werden kann. Dabei muss beachtet werden, dass die tatsächliche Anzahl an Personen durch mehrere Faktoren von der vom Mast gemessenen Anzahl abweicht und dementsprechend korrigiert werden muss.

Zum einen werden durch die sich regelmäßig ändernden RPIs zu viele Begegnungen vom Mast aufgezeichnet, welche mittels der Korrekturformel nach unten korrigiert werden.

Des Weiteren besitzt nur ein Bruchteil der Bevölkerung die CWA. Für diesen Fall wurde von einer Gleichverteilung der Nutzer unter der Bevölkerung ausgegangen und der gemessene Wert der Masten auf die Gesamtbevölkerung abgeschätzt.

Beim infrastrukturlosen Ansatz fällt die Implementierung der zentralen Masten weg. Nun werden einzelne Smartphones untereinander als mobiler Mast festgelegt, die ihre Umgebung aufzeichnen. Auch hier werden die Daten zunächst aggregiert, bevor sie auf den zentralen Server geladen werden.

Für den infrastrukturbasierten Ansatz wurde eine Simulation erstellt, die die Umsetzbarkeit des Features und die Korrektheit der Formel zur Korrektur nachweisen soll. Dabei werden 100 Smartphones mittels Punkte simuliert, die sich in einem Random-Direction-Modell über ein festes Gebiet bewegen. Zudem gibt es noch den „Social Distancing“ Modus, bei dem etwa 80% der Handys keine Bewegung aufweisen und somit ein Lockdown simuliert wird. Mithilfe von vier zentralen Masten werden die Smartphones im Umkreis

der jeweiligen Masten aufgezeichnet. Zu Verifizierungszwecken besitzen die Handys sowohl eine feste ID, als auch die temporären RPIs, die der Mast im Falle einer Begegnung aufzeichnen kann. Die Ergebnisse der Aufzeichnung zeigen, dass die Korrekturformel sowohl im Falle mit, als auch ohne Social Distancing den gemessenen Wert gut nach unten korrigieren kann.

Die Density-Map stellt eine effektive Methode dar, Orte mit vielen Besuchern bewusst meiden zu können. Dabei bietet sie eine Privatsphäre-konforme Alternative zum Stoßzeiten Feature von Google, da hier die anonymen RPIs zur Aufzeichnung verwendet werden. Außerdem werden die Daten dezentral gebündelt und aggregiert, wodurch einzelne Personen nicht nachverfolgt werden können.

Die gewonnenen Erkenntnisse stellen im Hinblick auf eine ausgereifere CWA vielversprechende Erweiterungsmöglichkeiten dar, welche besonders durch die mit Hilfe von Crowd-Sensing gewonnenen Daten neue Aspekte bieten.

In diesem Zusammenhang wäre es lohnenswert, die in den Absätzen größtenteils theoretischen Ausarbeitungen in der Zukunft in einer realen Umgebung zu testen und zu untersuchen. Damit könnten die theoretischen Ausarbeitungen im Feld validiert und im nächsten Schritt in die aktuelle CWA eingebunden werden. Des Weiteren könnte durch Benutzerstudien untersucht werden, inwiefern sich diese Erweiterungen auf die Nutzerzahlen, die Risikoberechnung und auf die Zufriedenheit der Nutzer auswirken.

## Literatur

- [1] Center for Systems Science and Engineering, Johns Hopkins University, „COVID-19 Dashboard“, <https://coronavirus.jhu.edu/map.html>, 07.06.2021
- [2] Robert Koch Institut, „COVID-19: Fallzahlen in Deutschland und weltweit“, [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Fallzahlen.html](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html), 07.06.2021
- [3] Robert Koch Institut, „Infektionsketten digital unterbrechen mit der Corona-Warn-App“, [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/WarnApp/Warn\\_App.html](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Warn_App.html), 09.06.2021
- [4] Christina Felschen, ZEIT online, „Geplante Corona-App soll Daten doch dezentral speichern“, <https://www.zeit.de/digital/datenschutz/2020-04/datenschutz-corona-tracing-app-dezentrale-speicherung>, 03.04.2021
- [5] Juliane Frisse, ZEIT online, „Zu wenige Menschen nutzen die Corona-Warn-App“, <https://www.zeit.de/wissen/gesundheit/2020-08/corona-warn-app-contact-tracing-downloads-nutzung-niedrig>, 03.04.2021
- [6] Miguel Ángel Durán, „COVID-19 (Coronavirus) spread simulator“, <https://github.com/midudev/covid-19-spread-simulator>, 26.04.2021
- [7] C. Troncoso et al., „Decentralized Privacy-Preserving Proximity Tracing“, <https://github.com/DP-3T/documents/blob/master/DP3T%20White%20Paper.pdf>, 10.06.2021
- [8] R. K. Ganti, F. Ye and H. Lei, „Mobile crowdsensing: current state and future challenges“, in *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32-39, November 2011, doi: 10.1109/MCOM.2011.6069707
- [9] J. Dutta, F. Gazi, S. Roy and C. Chowdhury, „AirSense: Opportunistic crowd-sensing based air quality monitoring system for smart city“, *2016 IEEE SENSORS*, 2016, pp. 1-3, doi: 10.1109/ICSENS.2016.7808730.
- [10] Z. Qin and Y. Zhu, „NoiseSense: A Crowd Sensing System for Urban Noise Mapping Service“, *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)*, 2016, pp. 80-87, doi: 10.1109/ICPADS.2016.0020
- [11] F. Saremi and T. Abdelzaher, „Combining Map-Based Inference and Crowd-Sensing for Detecting Traffic Regulators“, *2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems*, 2015, pp. 145-153, doi: 10.1109/MASS.2015.18
- [12] Y. Tsuchiya and N. Ishikawa, „Mobile Crowdsensing from Sensor Devices Using Bluetooth Low Energy (BLE)“, *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2018, pp. 800-805, doi: 10.1109/CSCI46756.2018.00160
- [13] T. You, M. Umair and Y. Hong, „Mobile crowd sensing based on CICN“, *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, 2017, pp. 1272-1275, doi: 10.1109/ICTC.2017.8190919
- [14] A. Basalamah, „Sensing the Crowds Using Bluetooth Low Energy Tags“, in *IEEE Access*, vol. 4, pp. 4225-4233, 2016, doi: 10.1109/ACCESS.2016.2594210.
- [15] Google, „Stoßzeiten, Wartezeiten und Besuchsdauer“, <https://support.google.com/business/answer/6263531/sto%C3%9Fzeiten-wartezeiten-und-besuchsdauer?hl=de>, 04.06.2021
- [16] Wernke, M., Skvortsov, P., Dürr, F. et al., „A classification of location privacy attacks and approaches.“, *Pers Ubiquit Comput* 18, 163–175 (2014). <https://doi.org/10.1007/s00779-012-0633-z>, November 2012
- [17] Google, „Google Maps-Zeitachse“, <https://support.google.com/maps/answer/6258979/google-maps%E2%80%91zeitachse-computer?co=GENIE.Platform%3DDesktop&hl=de>, 04.06.2021
- [18] Robert-Koch-Institut, „COVID-19 in Deutschland“, [https://www.rki.de/DE/Home/homepage\\_node.html](https://www.rki.de/DE/Home/homepage_node.html), 09.06.2021
- [19] Bundesministerium für Gesundheit (BMG), <https://www.bundesgesundheitsministerium.de>, 09.06.2021
- [20] Robert Koch Institut, „Mit Daten von Fitnessarmbändern und Smartwatches mehr über die Verbreitung des Coronavirus erfahren“, [https://www.rki.de/DE/Content/Service/Presse/Pressemitteilungen/2020/04\\_2020.html](https://www.rki.de/DE/Content/Service/Presse/Pressemitteilungen/2020/04_2020.html), 09.06.2021
- [21] Deutsches Zentrum für Neurodegenerative Erkrankungen DZNE), „Aufruf zum bundesweiten Gedächtnistest: Was macht „Corona“ mit unserem Kopf?“, <https://www.dzne.de/aktuelles/pressemitteilungen/presse/aufruf-zum-bundesweiten-gedaechtnistest-was-macht-corona-mit-unserem-kopf>, 09.06.2021
- [22] Bundesministerium für Bildung und Forschung, „Corona: Aktuelles aus der Forschung“, <https://www.gesundheitsforschung-bmbf.de/de/corona-aktuelles-aus-der-forschung-10857.php>, 09.06.2021
- [23] Robert Koch Institut, „Corona-Monitoring lokal“, [https://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/cml-studie/Factsheet\\_Kupferzell.html](https://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/cml-studie/Factsheet_Kupferzell.html), 09.06.2021
- [24] Statistisches Bundesamt (Destatis), „Bevölkerungsstand“, [https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/\\_inhalt.html](https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/_inhalt.html), 09.06.2021
- [25] Robert Koch Institut, „Kennzahlen zur Corona-Warn-App“, [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/WarnApp/Archiv\\_Kennzahlen/Kennzahlen\\_21052021.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Archiv_Kennzahlen/Kennzahlen_21052021.pdf?__blob=publicationFile), 09.06.2021
- [26] Jule Käppel, tagesschau, „Geimpfte können Millionen gewinnen“, <https://www.tagesschau.de/ausland/amerika/usa-ohio-impfung-lotterie-101.html>, 09.06.2021
- [27] Bundesgesundheitsministerium, „Viertes Gesetz zum Schutz der Bevölkerung bei einer epidemischen Lage von nationaler Tragweite“, [https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3\\_Downloads/Gesetze\\_und\\_Verordnungen/GuV/B/4\\_BevSchG\\_BGBL.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/B/4_BevSchG_BGBL.pdf), 09.06.2021

- [28] InTradeSys GmbH, „Darf ich das?-App“, <https://www.darfichdas.info/>, 09.06.2021
- [29] Deutsche Telekom AG und SAP SE, „Corona-Warn-App Documentation RiskCalculation“ [https://raw.githubusercontent.com/corona-warn-app/cwa-documentation/master/images/risk\\_calculation/risk\\_calculation\\_enf\\_v2\\_overview.pdf](https://raw.githubusercontent.com/corona-warn-app/cwa-documentation/master/images/risk_calculation/risk_calculation_enf_v2_overview.pdf), 09.06.2021
- [30] Bundesregierung, „Freiwillige Datenspende für verbesserte Wirksamkeit“, <https://www.bundesregierung.de/breg-de/themen/coronavirus/corona-warn-app-update-1-13-1873182>, 09.06.2021
- [31] Pharmazeutische Zeitung (PZ), „Corona-Warn-App kann Impfnachweis anzeigen“, <https://www.pharmazeutische-zeitung.de/corona-warn-app-kann-impfnachweis-anzeigen-126188/>, 10.06.2021
- [32] Deutsche Telekom AG und SAP SE, „Corona-Warn-App Documentation TRL“ [https://github.com/corona-warn-app/cwa-documentation/blob/master/transmission\\_risk.pdf](https://github.com/corona-warn-app/cwa-documentation/blob/master/transmission_risk.pdf)
- [33] Deutsche Telekom AG und SAP SE, „Corona-Warn-App Documentation RiskAssessment“ <https://github.com/corona-warn-app/cwa-documentation/blob/master/cwa-risk-assessment.md>
- [34] Almagor J., Picascia S., „Exploring the effectiveness of a COVID-19 contact tracing app using an agent-based model.“, *Sci Rep* 10, 22235, 2020, doi: 10.1038/s41598-020-79000-y