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

*Note: Sub-titles are not captured in Xplore and should not be used

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

Betreuer: Dr. rer. Nat. Frank Dürr
Institut für Parallele und Verteilte Systeme
Universität Stuttgart
Stuttgart, Germany
frank.duerr@ipvs.uni-stuttgart.de

Prüfer: Prof. Dr. rer. Nat. Kurt Rothermel
Institut für Parallele und Verteilte Systeme
Universität Stuttgart
Stuttgart, Germany
kurt.rothermel@ipvs.uni-stuttgart.de

***Zusammenfassung*—TODO**

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

I. Einleitung

II. Bestehende Arbeiten

A. Corona Lage

Bei Corona oder COVID-19 handelt es sich um eine Krankheit, die durch die Coronaviren SARS-CoV ausgelöst wird. Anfang 2020 waren diese Viren verantwortlich für die folgende weltweite Corona Pandemie, die aktuell zu weltweit 3 500 000 Toten 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 1-2Metern 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.

B. Corona Warn App

Die deutsche "Corona Warn App"(CWA) wurde im Auftrag des Robert Koch Institutes von SAP, der deutschen Telekom und weiteren Unternehmen entwickelt. 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 heruntergeladen (Stand Mai 2021).

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 hat, erhält man von einer Teststelle zusammen mit dem positiven Testergebnis einen Code, welchem man in seine App eintragen kann. Alle Kontaktpersonen werden dementsprechend 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 Risikobegegnungen 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 Corona Apps entwickelt wurde und das Bluetooth Low Energy (BLE) Feature nutzt, um die Begegnungsdaten der Smartphones anonym untereinander auszutauschen.

Exposure Notification Framework:
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 Identifiers (RPI) abgeleitet. Diese RPI's haben eine Lebensdauer von 10-15 Minuten, bevor sie gelöscht und ein neuer RPI aus dem TEK erstellt wird. Wichtig ist hierbei, dass sich aus den RPI's nicht der TEK ableiten lässt. Die RPI's werden nun unter den Smartphones im Falle einer Begegnung mithilfe von BLE ausgetauscht. Neben den RPI's werden zudem verschlüsselt Informationen über die Begegnungsdauer, sowie dem ungefähren Abstand ausgetauscht. Falls sich nun eine Person mit dem Coronavirus infiziert, werden dessen TEK (nun Diagnosis Keys genannt) der letzten 14 Tage auf einen zentralen Server hochgeladen. Jedes Smartphone hat Zugriff auf diesen zentralen Server und kann die Diagnosis Keys aller Infizierten Personen herunterladen. Mithilfe der Diagnosis Keys können dann die gespeicherten RPI's 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 das Infektionsrisiko der Person.

Transmission Risk Level (TRL): Das TRL wird in der CWA lokal berechnet und auf dem CWA Server zu jedem Diagnosis Key abgespeichert, um diesem ein Übertragungs Risiko (TRL) zuzuordnen. Dieses Risiko Level beruht auf der Annahme, dass eine infizierte Person ein Übertragungspotenzial hat das anfangs niedrig ist, dann auf ein Maximum steigt und innerhalb von 14 Tagen wieder abnimmt . Um daraus ein Risiko Level abzuleiten müssen verschiedene Zeitpunkte bekannt sein. Darunter zählen der Symptombeginn, das Infektiös sein, das angesteckt worden sein und drei weiter 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(Quelle CWADoku). Allerdings werden diese Zeitpunkte noch nicht bereitgestellt und es wird somit auf einen Standardwert zurückgegriffen. Weshalb im Verlauf des Dokuments das Feature TRL Update vorgestellt wird, dass eine Lösung zeigt die fehlenden Daten zu sammeln.

C. Crowd sensing und Bluetooth Low Energy

Bezüglich Crowd Sensing und Density-Maps 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 [2]. Die Standortbasierten Daten werden dann auf einen Server geladen, damit Nutzer die aktuelle Belastung in den Regionen (oder auch Räumen in Gebäuden) abfragen kann. Eine ähnliche Umsetzung gibt es hinsichtlich der Lärmbelästigung [3].

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. [4] Benutzen unter anderem GPS Daten von Smartphones oder Fahrzeugen, um Informationen über die Existenz von Ampeln an Kreuzungen zu extrahieren.

BLE Crowd Sensing: Auch im Bereich BLE existieren Crowd Sensing Anwendungen zum Erfassen von Daten und Erstellen von Density-Maps. Tschuhiya et al. [5] 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 [6] werden Smartphones und vorinstallierte Raspberry Pies mit CICN (Community Information-Centric Networking) verwendet, um eine IoT Sensoren Map innerhalb geschlossener Gebäuden, wie etwa Shopping Malls, 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 kann.

Basalamah [7] 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 [8]. In dessen Suchmaschine lässt sich zu fasst jeder öffentlichen Einrichtung, sowie Restaurants oder Einzelhandel Läden neben den Öffnungszeiten die aktuelle Besucherzahl anzeigen. Zudem wird noch die Durchschnittliche Aufenthaltsdauer von Personen, sowie eine Abschätzung der Auslastung über den Tag verteilt mitgegeben. 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 den konkreten Wert 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 [9]. 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. Auch bei der Nutzung von BLE wurden diese 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 Wissen demnach nicht zu geben.

III. Lösungsvorschläge

A. Übersicht aller Erweiterungen

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 digiatalen 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 immer ein relevanter Nutzen, da es das Ziel ist, die Corona-Warn-App von 100% der Bevölkerung nutzen zu lassen.
Implementiert man zusätzlich dazu eine Art Krankenakte, in der man Merkmale wie Gewicht und Vorerkrankungen einträgt, liese 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 wirklich nutzbar außer für den Endnutzer, da man auf die Eigenverantwortung des einzelnen Nutzers angewiesen wäre, die Daten also einfach gefälscht werden können.
Stattdessen müsste man, um die Daten als sicher zu garantieren, die Bearbeitung des Impfpass beschränken, sodass nur offizielle Stellen wie Krankenhäuser oder Arztpraxen diesen bearbeiten können.
Somit wird Richtigkeit der Daten gewährt, welche dann für das Krankenwesen genutzt werden können.

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 simple 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 getroffenen 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 ernsten 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.

B. Bewertungskriterien

Um einen möglichst einheitlichen Vergleich der Features zu garantieren, haben wir sechs Bereiche festgelegt, nach denen die Features beurteilt 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 ein 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?

C. Bewertung Nebenfeatures

IV. Hauptfeatures

A. Erweiterte Risikoberechnung

Mit der Risikoberechnung wird das Übertragungspotenzial der Begegnungen ermittelt, um damit die Gefahr infiziert worden zu sein abzuschätzen. Dafür wird die Risikoberechnung der CWA erweitert, um damit einen Risikowert zu berechnen, der die Gefahr widerspiegelt.

Man unterscheidet zwischen zwei Gruppen: CWA-User, die von der CWA somit aufgezeichnet werden und Not-CWA-User, die die CWA nicht nutzen und somit nicht aufgezeichnet werden. CWA-User werden in drei Untergruppen aufgeteilt: *CWA-Positive* sind alle Personen, die einen positiven Test in die CWA eingetragen haben, *CWA-Likely-Positive* sind alle Personen, die mindestens eine Begegnung mit CWA-Positive hatten und einen roten Bildschirm auf der CWA angezeigt bekommen und *CWA-Likely-Negative* sind alle Personen, die meist keine Begegnung mit CWA-Positive hatten und somit einen grünen Bildschirm auf der CWA angezeigt bekommen. Ebenso wird die NOT-CWA-User in drei Untergruppen aufgeteilt: Not-CWA-Positive sind alle Personen, die ein positives Test Ergebnis haben, dieses aber nicht in die CWA eingetragen haben (Es ist möglich, dass sie die CWA trotzdem nutzen), Hidden-Positive sind alle Personen, die infiziert wurden, jedoch kein positives Testergebnis nachweisen können und Hidden-Negative sind alle Personen, die nicht infiziert sind.

Für die erweiterte Risikoberechnung werden Features (Familienfeature, Feature TRL Update, Feature PK-Server) benötigt, die in Abschnitt 1 bis 3 vorgestellt werden. Im Abschnitt 4 wird auf die erweiterte Risikoberechnung genauer eingegangen und im 5. Abschnitt wird die erweiterte Risikoberechnung bewertet.

1) Familienfeature: Im Folgenden wird ein Familienfeature vorgeschlagen, welches später dafür genutzt wird die Risikoberechnung innerhalb der CWA zu verbessern. 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. 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. Insbesondere heißt dies, dass Begegnungen mit Freunden bzw. Verwandten nicht in $R_{\text{CWA-Likely-Positive}}$, Abschnitt (CWA-Likely-Positive) und $R_{\text{CWA-Likely-Negative}}$, Abschnitt (RLikely-Negative), miteinberechnet werden.

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 möglich mit hoher Wahrscheinlichkeit 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:

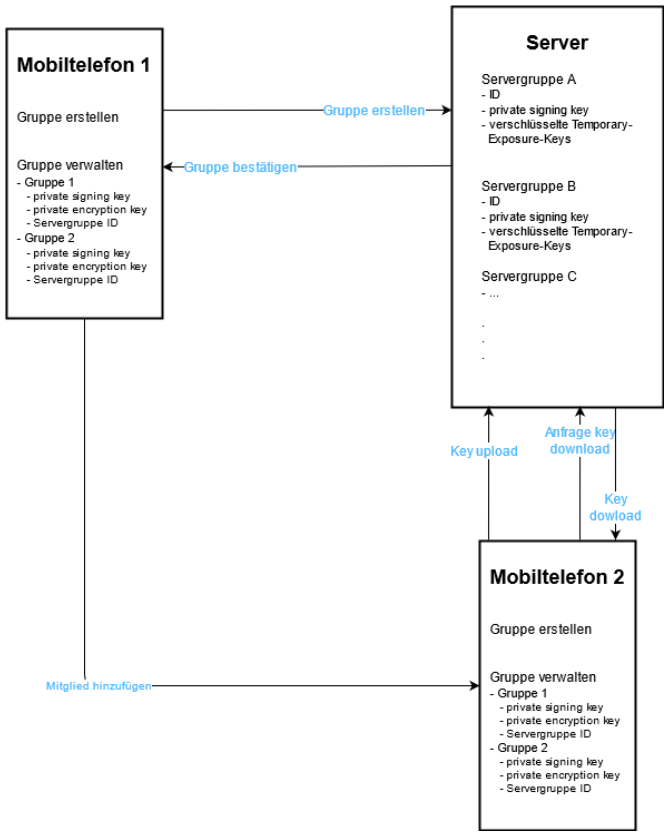

Abbildung 1. Familienfeature

Als Servergruppe wird eine Gruppe auf einem speziell für das Familienfeature angelegten Server bezeichnet. 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. Innerhalb einer lokalen Gruppe wird ein private signing key, ein private encryption key und eine Servergruppen ID gespeichert. Die Verwendung des private signing keys, private encryption keys und der Servergruppen ID wird in den nächsten Absätzen genauer erläutert. Der private signing key und insbesondere private encryption key einer Gruppe müssen geheim gehalten werden, damit die Privatsphäre keines Mitglieds verletzt wird. Es gibt zwei verschiedene Möglichkeiten den private signing key, private encryption key 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 private signing key und ein private encryption key generiert. Der private signing key wird zusammen mit einer Anfrage, eine neue Servergruppe zu erstellen, an den Server gesendet. Ein Verlust des private signing keys kann dazu führen, dass ein Angreifer willkürlich TEKs hochladen und herunterladen kann, die dieser Gruppe zugeordnet werden. Allerdings reicht der private signing key nicht aus, um die TEKs von Gruppenmitgliedern lesen zu können, da diese mit dem private encryption key verschlüsselt sind. Der Server erstellt auf die Anfrage hin eine neue Servergruppe mit einer noch nicht verwendeten ID und schickt die ID an den Anfragesteller zurück. Der lokal erstelle private signing key, private encryption key, sowie die erhaltene Servergruppen ID werden der Gruppe zugehörig abgespeichert.

Im Falle der zweiten Möglichkeit, einer bereits existierenden Servergruppe beizutreten, benötigt man ein Mitglied der Servergruppe. Das Mitglied der Servergruppe teilt den private signing key, private encryption key 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 ein oder mehrere Gruppen lokal erstellt und die zugehörigen private signing keys, private encyrption keys 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 private encryption keys einer Gruppe und eines sicheren, symmetrischen Verschlüsselunsverfahrens verschlüsselt und dann mit dem private signing key 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 private signing key 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 nach 14 Tagen aus der Liste der verschlüsselten TEKs gelöscht. 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 private signing keys 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 private encryption key der Gruppe entschlüsselt werden. Auch hier geschieht dies iterativ für jede lokale Gruppe.

2) Feature PKServer: Das Feature PKServer ist Voraussetzung für die in Abschnitt (Risikoberechnung) eingeführte 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 die Risikoberechnung in Abschnitt (R-CWA-Likely-Positive) 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 Potenziell Kranke Server (PKServer) bezeichnet. Wird in der CWA ein erhöhtes Risiko mit einem Risikowert ≥ 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 PKServer 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 PKServer, die älter als 14 Tage sind, gelöscht.

Die CWA lädt regelmäßig alle neuen TEKs mit den zugehörigen Risikowerten vom PKServer herunter und leitet daraus die RPIs ab. Die Verwendung der RPIs und Risikowerte in der Risikoberechnung wird im Abschnitt (CWA-Likely-Positive) 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 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 PKServer hochladen zu können. Dies wirft verschiedene Probleme hervor, die zwar im Folgenden genauer erläutert, jedoch nicht gelöst, werden.

Die vorgeschlagene Implementierung dieses Features sieht vor, dass ein TEK nicht mehrmals innerhalb 14 Tage auf den PKServer hochgeladen wird. Dies führt dazu, dass nur der erste Risikowert, der zu einem roten Bildschirm führt, zusammen mit den TEKs auf den PKServer hochgeladen werden kann. Da die Risikowerte in direkter Verbindung mit der Risikoberechnung stehen, kann dies zu einer ungenaueren Risikoberechnung führen. Sollten bereits auf dem PKServer 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 PKServer nicht eingeschränkt ist.

Es ist des weiteren möglich falsche TEKs auf den PKServer hochzuladen, da weder die Existenz eines TEKs, noch das Vorhanden sein eines roten Bildschirms, überprüft wird. Dies ermöglicht Denial of Service Angriffe mit verhältnismäßig geringem Aufwand und Ressourcen.

3) Feature TRL Update: In diesem Feature wird das volle Potenzial des TRLs 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 dafür existiert bereits in der CWA, jedoch existieren die benötigten Daten noch nicht, wie im Abschnitt (TRL) genauer beschrieben. Durch Umfragen, die beim Eingeben eines positiven Corona Tests aktiviert werden, sollen die Infizierten ihren Infektionszeitpunkt und den Symptom-Start-Zeitpunkt angeben können. 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 Infektiös seins kalkuliert werden und das Übertragungspotenzial mit dem TRL für jeden Diagnosis Key abgeschätzt werden, damit Nutzer besser gewarnt werden können.

4) 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 dieses zu berechnen wird das Übertragunspotenzial aus Begegnungsdauer und Begegnungsnähe mit einer Person, die ein positives Testergebnis in die CWA eingetragen hat, mit einbezogen. Und zum anderen die Wahrscheinlichkeit, dass eine Person infiziert ist, die im folgenden als Positiv-Begegnungs-Wahrscheinlichkeit bezeichnet wird. Wenn das Übertragunspotenzial sich insgesamt zu einer Begegnungsdauer von 13 Minuten in weniger als 8 Metern Abstand aufsummiert, wird der Bildschirm rot, ein erhöhtes Risiko angezeigt. Die Positiv-Begegnungs-Wahrscheinlichkeit ist in der aktuellen CWA immer 1, da es sich in jedem Fall um Begegnungen mit infizierten Personen handelt. Diese Berechnung hätte einen hohen Nutzen, wenn mehr Personen die App herunterladen würden und mehr Tests zur Verfügung gestellt werden könnten. (Quelle 123) Je mehr Tests durchgeführt werden und auch in die CWA eingetragen werden, umso genauer wird die momentane Risikoberechnung. Aufgrund der aktuellen Test-Möglichkeiten und Downloadzahlen sollte die Risikoberechnung jedoch angepasst werden.

Probleme und Lösungsansätze:
Ein Problem der Risikoberechnung der 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 positiven 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 höhere Wahrscheinlichkeit haben infiziert zu sein, da sie direkten Kontakt zu Infizierten hatten, werden in der neuen Berechnung stärker als die CWA-Likely-Negative 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.

Zudem sollte es eine geringere Abstufung der Risikostufen geben, denn nur grün und rot wecken den Eindruck, dass man

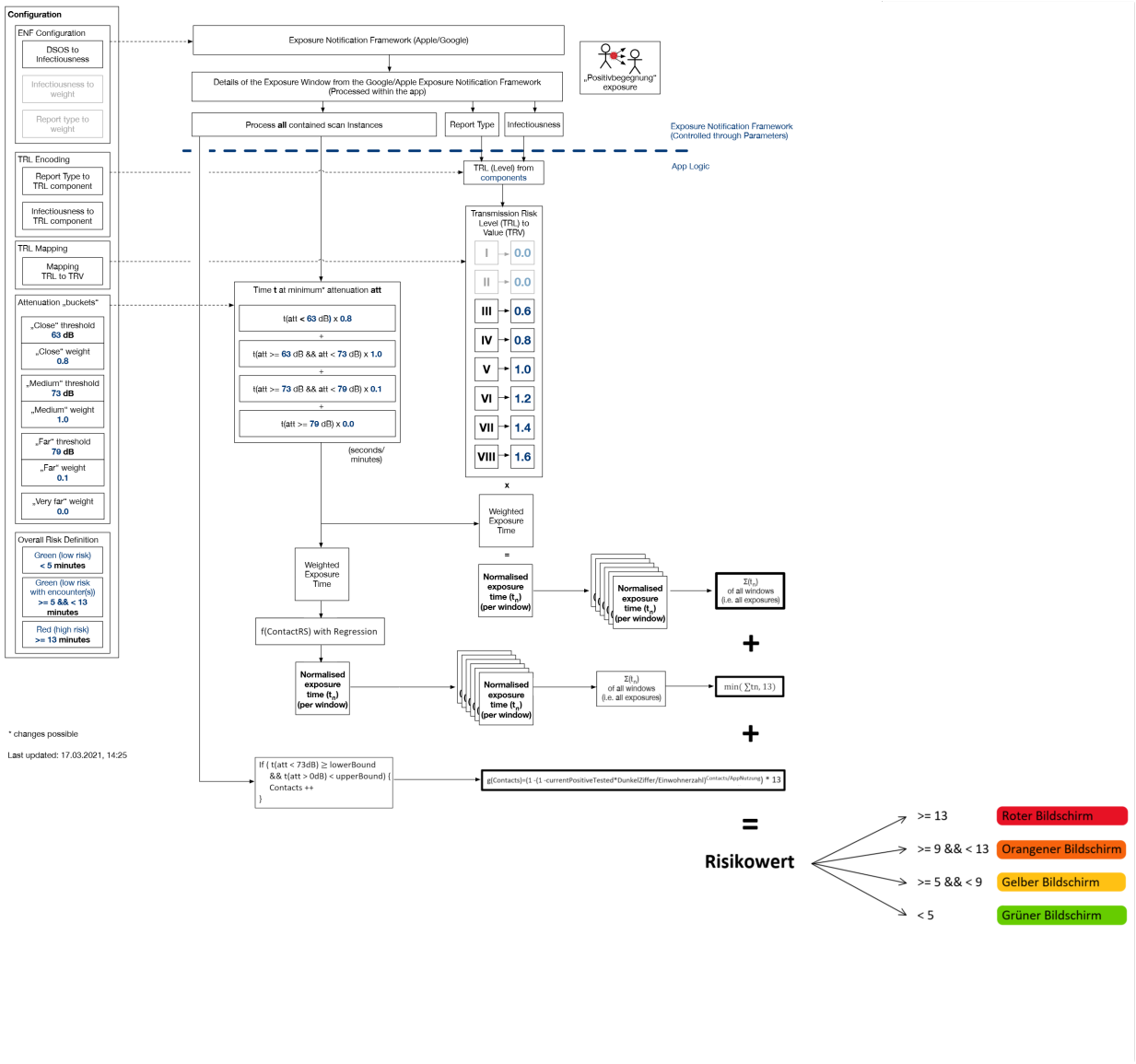

Abbildung 2. erweiterte Risikoberechnung

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.

Überblick:
Dazu wurde eine erweiterte Risikoberechnung ausgearbeitet, die Daten des RKIs und aus anderen Features, sowie Daten aus Umfragen, nutzt, um die fehlenden Daten mit einzubeziehen oder zu beschaffen. Die neue Risikoberechnung wird transparent am Ende neben einer feineren Granulierung des Risikowerts angezeigt. Die neue Risikoberechnung behandelt drei Punkte die zu einem neuen Risikowert zusammenaddiert werden.

Neuer Risikowert :=

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

Der erste Summand ist $R_{CWA\text{-}Positive}$ und berechnet das Risiko, indem nur Begegnungen mit CWA-Positive Personen einbezogen werden. Dabei wird jede Begegnung gewichtet, je nach dem wie hoch das Übertragungspotenzial der Begegnung ist. Der zweite Summand $R_{CWA\text{-}Likely\text{-}Positive}$ 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 $R_{CWA\text{-}Likely\text{-}Negative}$, 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 dann als höhe des

Risikos interpretiert werden kann. Dabei werden anstatt der ursprünglichen zwei Farben nun vier Farben Rot, Orange, Gelb und Grün verwendet.

$0 \leq Risikowert < 5 \rightarrow$ grüner Bildschirm
$5 \leq Risikowert < 9 \rightarrow$ gelber Bildschirm
$9 \leq Risikowert < 13 \rightarrow$ orangener Bildschirm
$13 \leq Risikowert \rightarrow$ roter Bildschirm

Damit die Risikoberechnung transparenter für die Nutzer wird, werden die Anteile der einzelnen Summanden zum farbigen Bildschirm angegeben, sowie die Begegnungsdauer und Begegnungsnähe zu CWA-Positive Nutzern. Im Folgenden werden die drei Summanden näher beschrieben, wobei die angepasste Grafik in Abbildung 2 dies veranschaulicht.

a) $R_{CWA\text{-}Positive}$: Das Risiko $R_{CWA\text{-}Positive}$ 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 konstant 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 attFactor durch folgende Intervalle aus der Distanz att bestimmt wird:

$att < 63dB : attFactor = 0.8$
$63dB \leq att < 73dB : attFactor = 1.0$
$73dB \leq att < 79dB : attFactor = 0.1$
$79dB \leq att : attFactor = 0$

Dabei kann man die Werte 63dB bzw. 73dB, 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 (Quelle cwa-risk-assessment). Aus der Weighted Exposure Time wird dann die Normalised Exposure Time tn mithilfe des Transmission Risk Levels TRL ermittelt und des daraus resultierenden Transmission Risk Values TRV. Das Transmission Risk Level wird in Abschnitt TRL genauer beschrieben und bezieht die Wahrscheinlichkeit mit ein, dass eine Person die Krankheit auch wirklich überträgt. Der TRV ergibt sich aus diesem nach der folgenden Vorschrift:

$TRL = I \rightarrow TRV = 0.0$
$TRL = II \rightarrow TRV = 0.0$
$TRL = III \rightarrow TRV = 0.6$
$TRL = IV \rightarrow TRV = 0.8$
$TRL = V \rightarrow TRV = 1.0$
$TRL = VI \rightarrow TRV = 1.2$
$TRL = VII \rightarrow TRV = 1.4$
$TRL = VIII \rightarrow TRV = 1.6$

Mit dem TRV lässt sich dann t_n bestimmen:

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

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

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

b) $R_{CWA\text{-}Likely\text{-}Positive}$: Das Risiko $R_{CWA\text{-}Likely\text{-}Positive}$ 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 Familienfeature) nicht mit einbezogen, da diese sonst aufgrund der potenziell langen Begegnungsdauer und kurzen Distanz ein hohes $R_{CWA\text{-}Likely\text{-}Positive}$ 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 $R_{CWA\text{-}Positive}$ 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, die keinen passenden Diagnosis Keys auf dem CWA Server liegen haben. Zuerst wird, wie beim Summand $R_{CWA\text{-}Positive}$, 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 $R_{CWA\text{-}Positive}$ der Fall ist, da dafür ein positives Testergebnis vorausgesetzt wird, welches CWA-Likely-Positive Personen nicht vorweisen können. Anstelle davon wird als nächstes aus der Weighted Exposure Time jeder Begegnung die Standardized Exposure Time t_s ermittelt. Da die Begegnung der CWA-Likely-Positive Personengruppe angehört, ist es nicht sicher, dass diese auch Corona hat. Aus diesem Grund wird die Positiv-Begegnungs-Wahrscheinlichkeit benötigt, welche die Wahrscheinlichkeit angibt, dass die Begegnung infiziert ist. Um die Positiv-Begegnungs-Wahrscheinlichkeit zu erhalten wird jedoch erst eine Funktion f benötigt, die diese Wahrscheinlichkeit angibt. 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 TEK zusammen vom PKServer (Abschnitt Feature Potenziell Kranke) heruntergeladen. Um f zu bestimmen wird folgendes Verfahren angewandt: Zunächst werden Risikowert-Intervalle gebildet, da Risikowerte nicht nur natürliche Zahlen als Werte annehmen können. Anhand der gesammelten Daten (Abschnitt Feature Potenziell

Kranke) wird für jedes Risikowert-Intervall $x_i miti \in N$ 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 PKServer alle $\leq$ 13 sind. Damit die Wahrscheinlicht 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 hat und die Anzahl an Personen, die sowohl einen Risikowert innerhalb des Intervalls x_i hat, als auch später an Corona erkrankt ist. Dafür werden die TEKs auf dem PKServer und die Diagnosis Keys auf dem CWA Server verwendet. Aus den TEKs auf dem PKServer kann die Anzahl an Personen bestimmt werden, die einen Risikowert innerhalb des Intervalls x_i hat. Werden zusätzlich zu den TEKs auf dem PKServer 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 hat, als auch später an Corona erkrankt ist, indem die TEKs auf dem PKServer und den zugehörigen Risikowerten mit den Diagnosis Keys auf dem CWA Server abgeglichen werden. Daraus lässt sich für jedes Intervall x_i die Wahrscheinlichkeit, infiziert zu sein, wie folgt bestimmen:

$$\frac{\text{Anzahl TEKs auf dem PKServer und CWA Server mit Risikowert}\in x_i}{\text{Anzahl TEKs auf dem PKServer mit Risikowert}\in x_i}$$

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

Zur Anzahl TEKs auf dem PKServer mit Risikowert $\in x_i$ werden alle TEKs auf dem PKServer 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 Bestimmung der Funktion f die richtige Datenmenge genommen werden. Die Anzahl TEKs auf dem PKServer 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 PKServer 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 PKServer hochgeladen worden sein. Verwendet man zur Bestimmung der Funktion f zum Beispiel die TEKs der letzten 7 Tage auf dem PKServer, 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-Begegnungs-Wahrscheinlichkeit zuordnet. Hierbei ist besonders zu beachten, dass die Funktion f etwa alle 7 Tage mit aktuelleren Daten neu berechnet und für die CWA aktualisiert werden muss. Beim Herunterladen von Keys (vom CWA Server oder PKServer) innerhalb der CWA muss dann gleichzeitig die aktualisierte Funktion f heruntergeladen werden. Mithilfe der Funktion f kann dann t_s bestimmt werden:

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

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 RCWA-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 roten Bildschirm hervorrufen kann, die wiederum bei weiteren Personen rote Bildschirme hervorrufen können und so weiter. Durch die Begrenzung kann zwar eine CWA-Likely-Positive Person mit einem hohen Risikowert immer noch rote Bildschirme bei anderen hervorrufen, jedoch können die Personen, die dadurch einen roten Bildschirm erhalten haben, diesen nur durch sehr lange Begegnungen, z.B. einen ganzen Tag, bei weiteren Personen hervorrufen.

c) $R_{CWA\text{-}Likely\text{-}Negative}$: Das Risiko $R_{\text{CWA-Likely-Negative}}$ (Abbildung 3, grüner Kasten) berechnet alle Begegnungen mit CWA-Likely-Negative und Not-CWA-User Personen mit ein. Denn aktuell haben nur ein Bruchteil der Infizierten ein positives Test Ergebnis in die App eingetragen. 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 wurden. 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. Die Dunkelziffer wurde erstmals durch

eine Studie in Kupferzell (Quelle) 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 6 möglicherweise ungenau.

- 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

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

Wie auch beim Risiko $R_{\text{CWA-Positive}}$ werden beim Risiko $R_{\text{CWA-Likely-Negative}}$ die Begegnungen der letzten 14 Tage beachtet. Denn 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 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 Familienfeature) und Begegnungen mit Diagnosis Keys, sowie zusätzlich Begegnungen mit Likely-Positive (Abschnitt Potenziell Kranke) 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 dafür 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 Aunter 73 Dezibel ist, der Abstand somit kleiner als 8 Meter ist und die Begegnungsdauer zwischen einem LowerBound und einem UpperBound ist.

Damit nicht jede noch so kurze Begegnungen aufgezeichnet wird, wird eine 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 wird noch einen UpperBound benötigt. Dafür soll jede Signalstärke mit einberechnet werden, damit kurze Distanzierungen keinen großen Einfluss haben. Für diesen Wert wird 14 Minuten geschätzt, damit Verbindungsfehler keinen großen Einfluss auf die Begegnungsanzahl haben. Jede Begegnung mit Begegnungsdauer zwischen Lower- und UpperBound wird somit gezählt.

$$if(att < 73dB) \leq lowerBound\{$$
$$\&\&t(att > 0dB) < upperBound)$$
$$Contacts++$$
$$\}$$

Um die Wahrscheinlichkeit zu berechnen, bei welcher Begegnungsanzahl man sich ansteckt, wird zuerst die aktuelle Anzahl infizierter ausgerechnet, um damit die Positiv-Begegnungs-Wahrscheinlichkeit auszurechnen. Die aktuelle Anzahl positiv getesteter wird mit der Dunkelziffer multipliziert, um die Anzahl aktuell Infizierter zu berechnen. Damit wird currentCWAPositive subtrahiert, da das Risiko bereits in $R_{\text{CWA-Positive}}$ mit einfließt.

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

$R_{\text{CWA-Likely-Negative}}$setzt sich aus dieser Positiv-Begegnungs-Wahrscheinlichkeit zusammen und wird auf 13 beschränkt, da die Person gewarnt werden soll, allerdings nicht als hoch Infektiös gilt. Dabei wird die Anzahl Contacts mit der AppUsage nach oben korrigiert, um die Anzahl wirklicher Begegnungen CorrectContacts besser abzuschätzen. Aktuell beträgt die AppUsage ca. $1/3$ (Quelle), somit werden die Contacts um das Dreifache erhöht.

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

Da mit einem Beschränkten Wachstum gerechnet wird und somit die Positive-Begegnungs-Wahrscheinlichkeit subtrahiert werden muss, muss mit der Gegenwahrscheinlichkeit gerechnet werden. Denn umso höher die Infektionswahrscheinlichkeit, desto geringer ist die Gegenwahrscheinlichkeit und desto näher ist $R_{\text{CWA-Likely-Negative}}$ an der Schranke 13. Um die Wahrscheinlichkeit zu berechnen infiziert worden zu sein, wird die Wahrscheinlichkeit, dass eine Person nicht infektiös ist, mit der Anzahl an getroffenen Personen CorrectContacts exponiert.

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

Somit ist es möglich ein erhöhtes Risiko in der App angezeigt zu bekommen, obwohl man keinen Diagnosis Key getroffen hat.

5) Bewertung: 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 freiwillig ist, benötigt man eine hohe Beteiligung an Nutzern, die Daten teilen. Zudem ist die Berechnung des Risikos $R_{\text{CWA-Positive}}$, 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 PKServers müssen Sicherheitsmechanismen gegen Denial of Service Angriffe implementiert werden, auf die in diesem Dokument nicht weiter eingegangen wird. Im Vergleich dazu wird bei den im Abschnitt (Feature TRL Update) 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 2 weitere Server gibt (PKServer 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 PKServer und dem CWA Server dazu genutzt, die in Abschnitt (R-CWA-LIKELY-POSITIVE) beschriebene Funktion f zentral zu bestimmen.

Durch die Installation von 2 weiteren Server, werden zusätzliche Serverkosten anfallen. Da diese Server einen ähnlichen Aufbau wie der CWA Server haben, werden die Kosten für diese in einen ähnlichen Umfang haben.

B. 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 dieses Feature benutzen, um in nahezu Echtzeit die Besucherzahl an öffentlichen Orten oder Geschäften aufzuzeichnen und grafisch auf einer Karte darstellen. Damit kann eine Nutzer 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.

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 Smatphones auch die RPI's 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 RPI's 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 ebebfalls gezielt nach Verfahren, welche weitmöglichst eine dezentrale Lösung verfolgen. Hierfür haben wir uns für die Umsetzung 2 Verfahren überlegt, mit welchen wir das Problem angehen: 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 wird die Anzahl der Begegnungen auf einen zentralen Server hochgeladen. Anders als beim naiven Ansatz, bei dem der zentrale Server die Daten aller Smartphones erfassen würde, ist hier die Aufzeichung 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 znetralen Servers erheblich.

Die Begegnungsaufzeichnung erfolgt analog zur Aufzeichnung mobiler Endgeräte untereinander 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 Threshhold 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 RPI's oder Personen über die Masten hinweg nachverfolgt und damit identifiziert werden können. Abb. 3 zeigt den groben Ablauf eines zentralen Masten.

Wie bereits erwähnt scannt ein Mast seine Umgebung nach

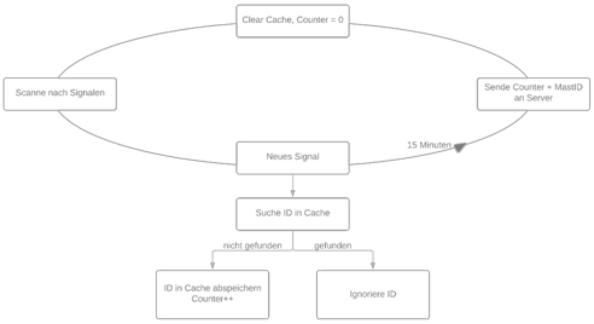

Abbildung 3. Funktionsweise eines zentralen Masten

Endgeräten und zeichnet alle Begegnungen innerhalb eines Zeitintervalls von etwa 15 Minuten (Lebensdauer eines RPI's) auf. Die Aufzeichnung erfolgt analog zur bisherigen Begegnungsaufzeichnung der Corona Warn App über BLE. Es werden demnach nur die RPI's 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 Counter 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, falls dies mit BLE möglich. 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 gleichtzeitig 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 aufhällt und dementsprechend nur einem Masten 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 Smarphones, welches als zentraler Mast fungieren soll, die App installiert werden. Jedoch muss vor der Nutzung des Features das Gerät von einer offizienllen Behörde autorisiert werden. Damit darf dann etwa ein Restaurantbetreiber, über eine zentrale Behörde registriert, seine Einrichtung mit seinem Smartphone aufzeichnen.
- Zentrale Masten müssen währende 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 Aufzeichnungsbereich 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 diese 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.

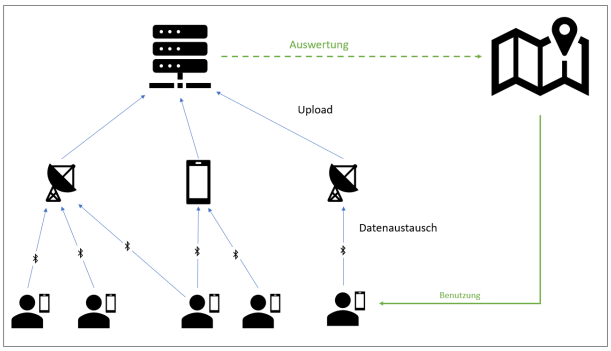

Abbildung 4. Übersicht: Erstellen der Density Map

Wie in Kapitel 2 beschrieben, besitzen die RPI's eine Lebensdauer von etwa 15 Minuten. Dementsprechend muss beachtet werden, dass Smartphones auch während des Aufenthaltes im Umkreis eines Masts 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 Masten 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 RPI's (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 Formel errechnet:

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

Dabei ist „RPIs“ die gemessene Anzahl an RPI's 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:

$$\text{RPIs} = \text{Ids} + \text{IDs} \cdot p$$

Diese Formel nach „IDs“ umgeformt ergibt die oben genannte Formel zur Berechnung der tatsächlichen Anzahl an Endgeräten.

Eine Übersicht infwiefern der gemessene Wert mithilfe der Formel korrigiert 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 Smatphones (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).

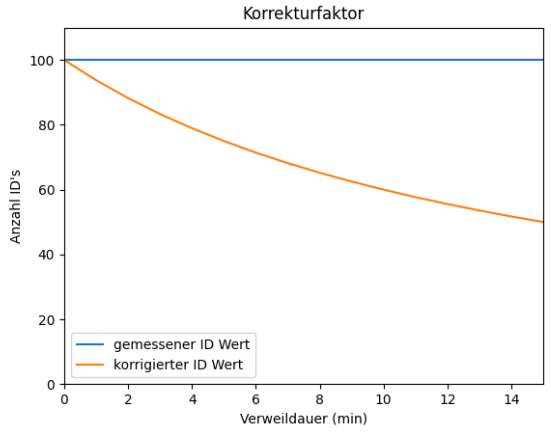

Abbildung 5. Korrekturfaktor in Abhängigkeit der durchschnittlichen Verweildauer

Eine Alternative zu dieser Formel 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 abhängt.
Bei einer Wahrscheinlichkeit p einer doppelten Registrierung müssen *p/2* Begegnungen von den RPI's abgezogen werden, um so den realen Wert zu erhalten.
Diese alternative Formel ist somit wie folgt definiert:

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

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 überehmen kann. Auch bei diesem Ansatz möchten wir znächst 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 RPI's 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. Smatphones, die ihren Standort nicht aktiv haben, werden können von anderen Smartphones mit aktiven Standort aufgezeichnet, 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.

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 entpsrechen Ort der Aufzeichnung eines anderen Gerätes und Ort des Hochladens der Daten nach Ablauf des Zeitintervalles nicht überein und das Bild der Density-Map

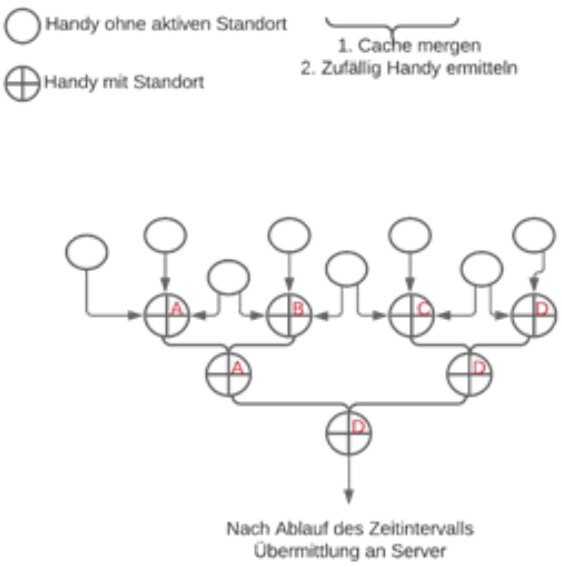

Abbildung 6. Smartphones als Mobile Masten

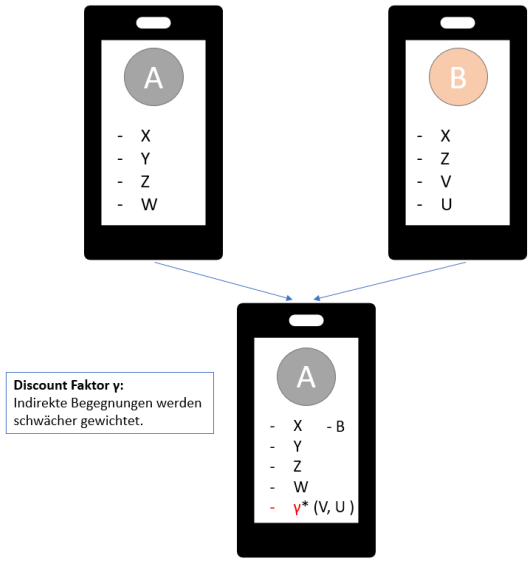

Abbildung 7. Discount Faktor für Mobile Masten

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 die Auswertung der Density Map.

In Abbildung 7 begegnen sich die beiden Smartphones A und B. A hat bereits die RPI's X, Y, Z und W aufgezeichnet und B die RPI's X, Z, V, U. Nun tauschen die beiden Smartphones ihre RPI's 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 RPI's 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). Dementpsrechend werden die Begegnungen RPI's U und V mit dem Discount Faktor γ 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. Sollten etwa die beiden indirekten Begegnungen U und V an ein weiteres Smartphone C übergeben werden, so würden sie bei C doppelt mit dem Discount Faktor versehen werden.

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 infrastrukturbasierenden Ansatz beschriebene spezielle Tag für RPI's 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, was 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 (34%). 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.

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 [1].

Wie bereits in Abschnitt IV.B beschrieben, existiert eine bestimme 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 die Formel

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

und im zweiten Teil dieses Abschnitts die Formel

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

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 Counter hoch, welcher die Dauer des Aufenthalts im Radius des Mastes misst. Am Ende eines Zyklus werden diese Counter addiert und durch die gemessene Anzahl der Handys geteilt.
Zwischenzeitliches Verlassen des Radius des Mastes setzt den Counter nicht zurück. Stattdessen wird nach Wiedereintritt an Hand der ID der alte Counter weitergeführt.
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 Countern 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 RPI's 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:
Abb. 9 stellt die gesammelten Daten in einem Schaubild dar, ohne das 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.
Mit Social Distancing (Abb. 10) ist die durchschnittliche Verweildauer wesentlich höher und der angeglichene Wert liegt näher an dem gemessen Wert als in Abb. 9.
Schlussfolgerung:
Die gesammelten Daten belegen, dass die durch die Formel 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

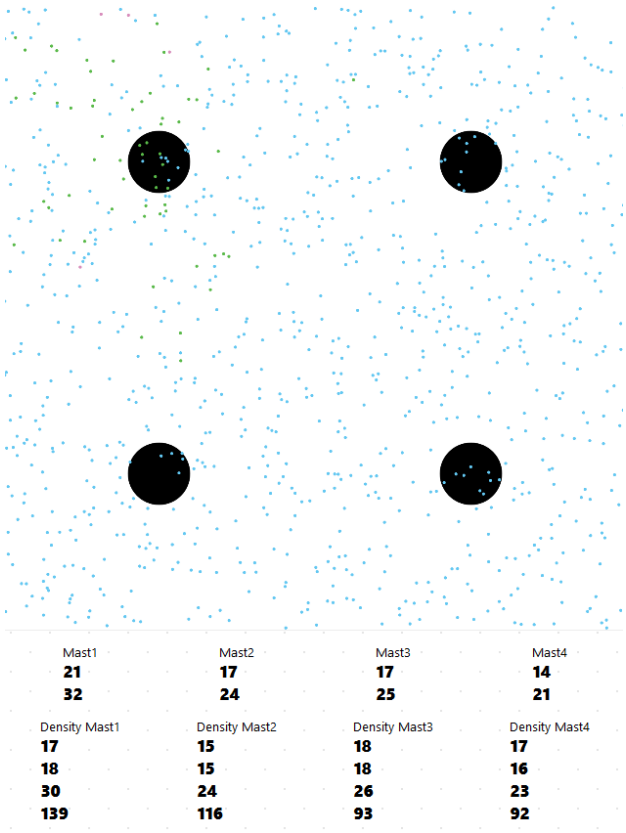

Abbildung 8. Darstellung der durchgeführten Simulation

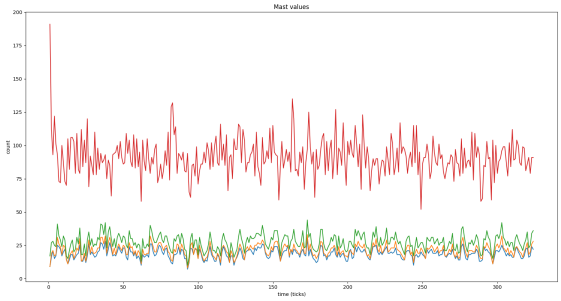

Abbildung 9. Ergebnisse ohne Social Distancing

Fällen die berechnete 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 bisherige Formel nicht optimal ist, da der berechnete Wert stets zu groß ist, sich grundlegend jedoch im Vergleich

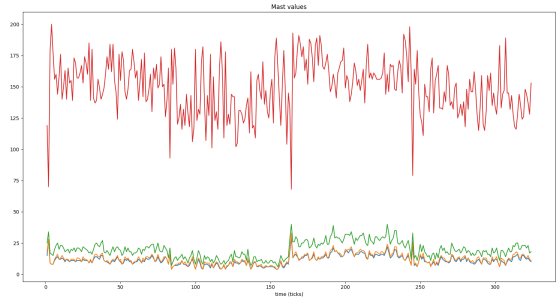

Abbildung 10. Ergebnisse mit Social Distancing

zum gemessenen Wert ein Wert näher an der realen Anzahl berechnen lässt.

Simulation zweiter Formel: Um die Formel $\text{IDs} = \frac{\text{RPIs}}{(1+p)}$ 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 11 und Abbildung 12gezeigten Werte.

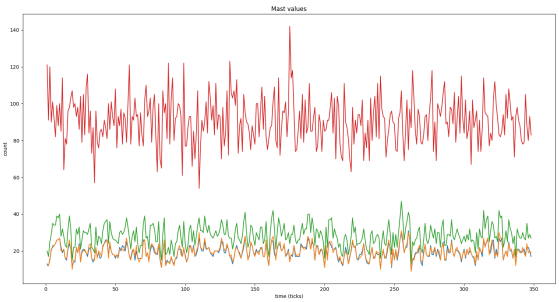

Abbildung 11. Neue Ergebnisse ohne Social Distancing

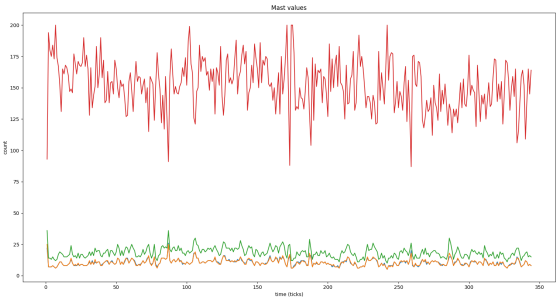

Abbildung 12. Neue Ergebnisse mit Social Distancing

Ergebnisse:
Wie auch im ersten Abschnitt lässt sich feststellen, dass der durch die 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 dies in der vorherigen Simulation der Fall war.
Der durchschnittliche Unterschied zwischen korrigiertem und realen Wert ist zudem niedriger mit dieser Formel als bei der Verwendung der anderen Formel.

Schlussfolgerung:
Durch den geringeren Unterschied zwischen korrigiertem und realen Wert ist die zweite Formel besser als die Formel der ersten Simulation für die Erfüllung unseres Ziels, die Anzahl von individuellen Handys innerhalb eines Mastes zu berechnen.

C. Bewertung der Hauptfeatures

V. Zusammenfassung und Ausblick

Danksagung

Literatur

[1] Miguel Ángel Durán, COVID-19 spread simulator, https://github.com/midudev/covid-19-spread-simulator

[2] J. Dutta, F. Gazi, S. Roy and C. Chowdhury, ÄirSense: Opportunistic crowd-sensing based air quality monitoring system for smart city,"2016 IEEE SENSORS, 2016, pp. 1-3, doi: 10.1109/ICSENS.2016.7808730.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] A. Basalamah, SSensing the Crowds Using Bluetooth Low Energy Tags,ïn IEEE Access, vol. 4, pp. 4225-4233, 2016, doi: 10.1109/ACCESS.2016.2594210.

[8] "Google Stoßzeiten Feature" https://support.google.com/business/answer/6263531/sto%C3%9Fzeiten-wartezeiten-und-besuchsdauer?hl=de (04.06.2021)

[9] "Google Standortverlauf" https://support.google.com/maps/answer/6258979/google-maps%E2%80%91zeitachse-computer?co=GENIE.Platform%3DDesktop&hl=de (04.06.2021)