Detecting Mobile Malware Associated with Global Pandemics

Alfredo J. Perez, University of Nebraska at Omaha, Omaha, NE, USA

David Tan, Valdosta State University, Valdosta, GA, USA

Abstract— The existence of more than 6 billion smartphones worldwide can enable governments and public health organizations to develop mobile apps to prepare, track, and treat global pandemic diseases and their effects. However, rogue actors can take advantage of this opportunity during global health emergencies to target the public in nefarious ways (such as the use of various types of malware). We briefly review and analyze use cases of smartphone apps to enhance public response to pandemics and disease outbreaks. Next, we discuss how malicious actors could exploit worldwide pandemic emergencies to develop malware apps disguised as legitimate pandemic apps. Considering the worldwide COVID-19 Android malware, we propose the use of app permissions to develop Machine Learning (ML) models to enable the fast-detection of pandemic-related malware in smartphones. Our results show that pandemic-related malware mobile apps can be detected to a considerable extent using app permissions, thus helping the public to avoid pandemics-related smartphone malware in future pandemics and other worldwide health emergencies and outbreaks.

Introduction

The emergence of new viruses affecting humans can cause not only death and long-term suffering, but also chaos in societies. Recently the CoronaVirus-19 (COVID-19) pandemic caused by the Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) is an example on how an emergent virus can bring the world into situations that were unthinkable for many only a few years ago. The advent of global, real-time telecommunications and the growth of mobile cellular technology in the last 25 years has also helped to develop new alternatives to prepare for emergent diseases and their epidemics (and possibly pandemics). More than 6 billion smartphones and wearable/portable sensors that can be connected via Bluetooth to a smartphone provide an alternative to inform, diagnose, track, treat and manage epidemics and global pandemics [1]. With the emergence of the COVID-19 pandemic, more than 2,000 COVID-themed mobile apps (not including malware) were developed for different purposes around the world as of December of 2020 [2].

Modern pandemics serve as backdrops for the emergence of rogue actors who exploit individuals, attack governments, and demand ransoms to organizations via global telecommunication networks, computers, and smartphones for financial gain. The COVID-19 pandemic has been exploited by cybercriminals using different threats, attacks and channels including Distributed Denial of Services Attacks (DDoS), malicious domains and websites, malware, ransomware, spam emails, malicious social media messaging, business email compromise, mobile apps, and browsing apps [3], impacting healthcare systems, financial services, government and media outlets, and the public. Cybercrime has increased dramatically during the COVID-19 pandemic, with an estimated impact of more than $6 trillion USD worldwide during 2021 [4]. This major increase in cybercrime activity during 2021 was due to the massive online activity caused by worldwide

lockdowns and restrictions in movement to mitigate the COVID-19 pandemic disease [4]. Thus, to prepare for future pandemics we need not only to create systems to help mitigate the effects of emergent diseases, but also to protect the global cyberinfrastructure during the containment and mitigation of pandemics.

Our contributions are as follows:

1. We review mobile and smartphones' use cases during well-known epidemics and pandemics,
2. We review how cybercriminals exploit mobile apps during a worldwide pandemic
3. We propose the use of app permissions in combination with Machine Learning (ML) to detect pandemics-related mobile malware at the edge

The rest of this paper is organized as follows. In next section we review use cases of mobile phones and smartphones apps in epidemics/pandemics. Then we review mobile malware during pandemics with a focus on the COVID-19 pandemic. Next, we propose the use of app permissions and Machine Learning (ML) as a fast approach to detect malware in Android smartphones. We provide conclusions and future work in the last section.

Pandemics and Cellphones/Smartphones and their Limitations

Smartphone Apps' Use Cases During Pandemics/Epidemics

Recently, the COVID-19 pandemic has highlighted the use smartphones as tools to manage public health emergencies. However, past epidemics and pandemics saw use of smartphones and data generated by mobile cellular communications (Table 1). For example, in 2003 a Hong Kong mobile operator launched a Location-Based Service (LBS) via Short Messaging Service (SMS) and Wireless Application Protocol (WAP) to notify subscribers when a nearby building was contaminated with the Severe Acute Respiratory Syndrome (SARS) during the 2003 SARS outbreak in Asia [5]. Radio Frequency IDentification (RFID) was used during this SARS outbreak in Singapore for contact tracing inside hospitals [5], allowing health officials to identify 10 times faster who an infected person had contact with than using other methods. A similar approach was used during COVID-19 in different parts of the world using Bluetooth Low Energy (BLE) [6].

Table 1. Mobile and smartphone's use cases during epidemics and pandemics in healthcare-related applications

Use case	Epidemic/pandemic disease	Approach example
LBS for building infection notification	Severe Acute Respiratory Syndrome (SARS) outbreak	Use of SMS and WAP to notify subscribers during the 2003 SARS outbreak in Hong Kong [5]
Contact tracing	SARS and COVID-19	Use of RFID to conduct contact tracing in Singapore hospitals during the 2003 SARS outbreak [5]

		Use of Bluetooth Low-Energy during the COVID-19 pandemic in contact-tracing apps worldwide [6]
Surveillance and tracking of virus spread	Cholerae	Surveillance of cholerae in wide areas using of anonymized mobile cellular operators' data from the 2010 Haiti cholerae outbreak [7]
Disease detection	Zika, chikungunya, and dengue Malaria, Ebola, and Marburg virus disease	Use of portable devices and sensors used with smartphones to detect pathogens in human specimens [8][9][10]
Treatment adherence and long-term disease management	HIV and tuberculosis	Use of SMS text messages in Kenya to remember patients to take ART medication [11] Similar approaches in other African nations for both HIV and tuberculosis [12]
Health education	HIV, tuberculosis, COVID-19	Mobile applications used by public health organizations to inform the public about infectious diseases, their symptoms, and effects [12]
Digital health passports (DHPs)	COVID-19	Mobile applications used by airlines, the European Union (EU), USA, private organizations, and other countries (such as Israel) to grant access to services for COVID-19 vaccinated individuals [13]
Telemedicine and communication between patients and families	COVID-19	Used extensively in the world during the COVID-19 pandemic for patients to contact practitioners due to lockdowns and safety precautions [1] Used by hospitalized patients in the world to contact their families to minimize contamination risks

Using only anonymized mobile phone data from cellular operators, Bengtsson et al. [7] created a model to survey and track the spread of cholerae in the 2010 Haiti epidemic. Their research showed that mobile operators' data can help to track and contain the spread of infectious diseases and serve

as a surveillance mechanism for wide areas. In 2017, Priye et al. [8] reported on the rapid detection of Zika, chikungunya, and dengue viruses using a portable device called the "LAMP Box", a smartphone's camera, and an app to detect and analyze samples of human specimens (e.g., blood, urine, and saliva). In a similar fashion, Yu et al. reported on the development of a smartphone app that captures photos taken from a portable microscope with a sample of human blood to detect the presence of malaria parasites (*P. falciparum*) [9]. Natesan et al. [10] reported a similar approach for the detection of Ebola and Marburg viruses.

For adherence to treatment using Antiretroviral Therapy (ART) for Human Immunodeficiency Virus (HIV) management, Horvath et al. [11] studied in 2012 the use of mobile phone SMS text messaging and they found based on two Randomized Controlled Trials (RCTs) studies in Kenya that weekly mobile phone text-messaging improved HIV viral load suppression by remembering patients to take their medications, thus helping patients to adhere to their therapy. For long-term treatment, Rama Devi et al. [12] found in a literature review from 2005 to 2015 that mobile phones were successfully used for long-term care and management of HIV and tuberculosis in developing countries. They reported that 73.3% of their reviewed papers (66 papers) reported positive effects on using mobile phones for HIV/tuberculosis management. Finally, during the COVID-19 pandemic other use cases of smartphones (and tablets) apps for public health settings included telemedicine/patient communication, health education, and apps implementing Digital Health Passports (DHPs) [13].

Limitations of Smartphones and their applications during Pandemics

While there have been great advances on the use of smartphones for epidemics/pandemics, there are also limitations for the successful use of smartphone apps during epidemics/pandemics in aspects such as interoperability, effectiveness, politics, design choices and marketing, and security and privacy.

From the interoperability perspective, applications developed during pandemics with a healthcare (or fitness) focus use a particular architecture (in hardware or software) that forbids (or makes almost impossible) for users to switch components (e.g., wearables for monitoring), health providers, or move healthcare data collected through them. While limitations may be related to laws, others are related to the lack of standardization and business models that makes difficult the interoperability among systems [1].

Many apps developed during pandemics are not evaluated on their effectiveness before or after deployment. For example, Rama Devi on their research about long-term treatment with mobile apps for HIV and tuberculosis [12] found that many research studies lack statistical evaluations on app effectiveness and rather used qualitative measures (e.g., perceptions) and empirical observations (without experiments). The lack of evaluation is exacerbated by a fast development of many mobile apps that are created as a public health response towards an emergent disease (e.g., COVID-19 case), thus impacting an app's efficacy, reliability, and privacy/security.

Additionally, the implementation of certain types of smartphone apps for pandemics may be subjected to politics. For example, during the COVID-19 pandemic, vaccination passports and their smartphone implementations (through DHPs) were subjected to policy decisions that varied

between U.S. states. DHPs were implemented in the state of New York when COVID-19 vaccination became widely available [13. However, in Florida any kind of vaccination passport was explicitly forbidden by an executive order from Governor Ron DeSantis in April 2021 [14].

From the perspective of the design and marketing of apps, the approach used to design, implement, promote, and give choices to the public about pandemics-related apps may affect their installation. In this perspective and using a survey in the U.S. with 1963 respondents to study why somebody would install a contact tracing app for COVID-19 (by exploring the design space of contact tracing apps), Li et al. [15] found that app design choices and individual differences affect more the intentions to install a contact tracing app than perceptions about an app's security and privacy. They recommend paying attention in the design and promotion of contact tracing apps to essential workers, healthcare professionals, and individual belonging to rural communities.

Finally, fast software development cycles used to develop and launch applications/systems during pandemics can account for technical debt that can leak data considered private and make software systems developed during pandemics to be vulnerable to cyberattacks. Results from a survey done during the COVID-19 pandemic in 2021 showed that 78% of the companies surveyed believed their technical debt increased during 2021, with most of the technical debt believed to be arising from the development of new products [16]. In the same survey, 86% of respondents mentioned that launching new digital products/services justified the expense of incurring in technical debt.

Mobile Malware during Pandemics

Mobile apps developed before COVID-19 for epidemics/pandemics were mostly applications developed by well-known organizations as part of health campaigns or prototype systems. However, the worldwide availability of smartphones and other wearables at the start of the COVID-19 pandemic, and their further increase in use as the pandemic progressed [1], made smartphones to be target of malware which grew quickly during the COVID-19 pandemic.

More than 2 million installations of mobile malware packages were performed worldwide during fourth quarter of 2020, which almost duplicated the number of malware package installations during the third quarter of the same year (around 1.1 million in the third quarter of 2021) [17]. These numbers began to decrease during 2021, reaching around nine hundred thousand installations by the second quarter of 2021 (Figure 1). Hackers also exploited users through malware camouflaged as legitimate COVID-19-themed apps. There were at least 370 unique COVID-19-themed mobile malware apps developed worldwide as of mid-November 2020 targeting the Android operating system with most apps released after March 2020 [2].

Hackers targeted smartphones during the COVID-19 pandemic not only because of their use, but also because of the lack of cybersecurity hygiene of smartphone users around the world. Misinformation and mobile malware distribution methods (different to the use of app stores), and vulnerabilities such as SMS phishing (by which SMS messages are used to distribute malware) and Zero-Click attacks (by which no input from users is needed before deploying an attack, but exploitation of vulnerabilities in apps already installed) were exploited by hackers to attack smartphone users during COVID-19 [18].

Other distribution mechanisms for mobile malware during COVID-19 included messages sent via social media apps (e.g., WhatsApp, Instagram, and others), and camouflaged malware distributed via app stores for both Android and iOS devices (i.e., Google Play Store and Apple App Store), even though app stores blocked more than 1 million attempts to circumvent security measures to publish mobile apps [18].

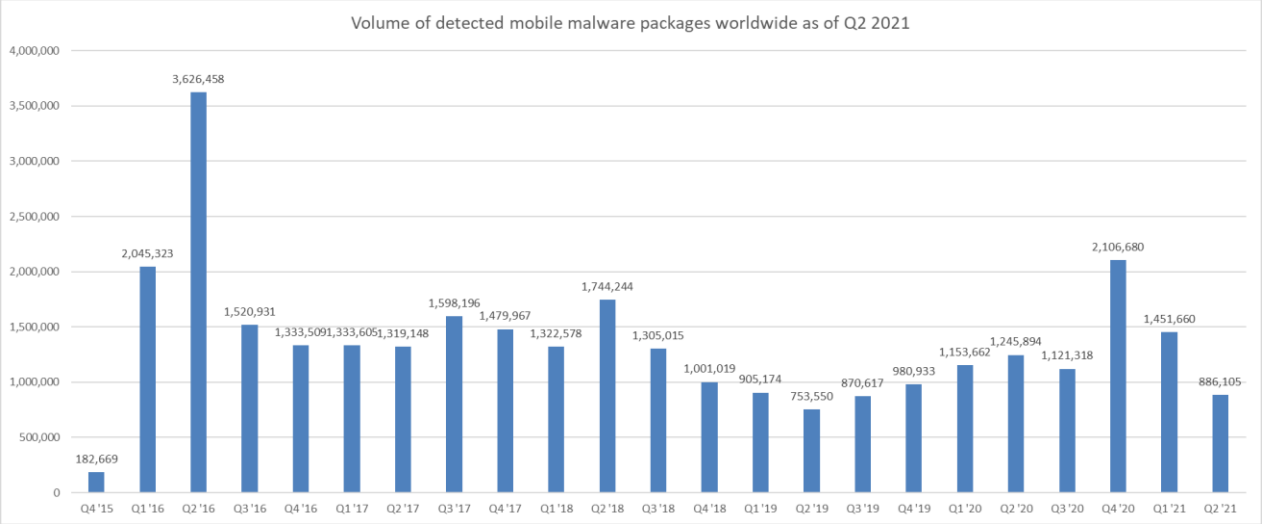

Figure 1. Detected mobile malicious installation packages from Q4 2015 to Q2 2022 based on data from Karpersky Lab's Securelist [17]

According to Karpsersky data, most of the new worldwide mobile malware in 2021 was in the form of AdWare (42.42% of the total), RiskTool (by which malware conceals files, run apps silently, or terminate active process, 35.27% of the total), and trojans (programs that claim to perform some function while doing something else, 8.86% of the total) [19]. For COVID-19-themed malware, Wang et al. [2] reported that trojans (56%) and spyware (29%) made most of the COVID-19-themed malware in Android as of November 2020. Ransomware made about 7% of mobile malware in their study.

Detecting Pandemics-related Malware Using App Permissions and ML

In this section we describe the use of app permissions and Machine Learning (ML) to detect COVID-19-themed malware in Android. We conducted our research based on the COVID-19-themed Android Package Kits (APKs) dataset curated by Wang et al. [2] with malware and non-malware COVID-19 Android apps samples. While there has been similar research based on permissions for COVID-19-themed Android apps [20], our research differs in three aspects: (1) our used dataset has more samples of malware and non-malware COVID-19-themed Android apps; (2) we used an extra feature (generated based on app permissions); and (3) and the use of Synthetic Minority Oversampling Technique (SMOTE) to balance the difference in the number of samples in the minority class (malware app samples in our case).

Permissions Dataset

We obtained our dataset by extracting permissions used by the apps from the COVID-19-themed dataset curated by Wang et al. [2]. This dataset (made publicly available by its curators) has 2,500 unique APKs with 370 unique APKs belonging to malicious apps collected by mid-November 2020. We used the AndroGuard to extract each apps' permissions.

Due to errors generated by AndroGuard when extracting data from some of the APKs, we ended with permissions of 2016 unique apps (80% of the original dataset) with 277 labeled as COVID-19-related malware samples (75% of the original malware samples) and 1739 labeled as non-COVID-19-themed malware (81% of the original non-malware samples). We created a spreadsheet with columns specifying if a specific permission was used by an app, and each row had the permissions used by an app. Thus, the spreadsheet had 2017 rows (first rows specifying the name of a permission) and 205 columns (203 corresponding to permissions, 1 column corresponding to the name of the app, and 1 column corresponding to the class (malware/non-malware)). If an app used a permission, the corresponding permission column for the app had a 1, otherwise, it had a 0.

We created the chart shown in figure 2 using our dataset. The chart in figure 2 shows that around 30% of non-malware apps used one permission, while most of the malware apps used four or more app permissions. The average number of permissions used by non-malware apps was 1.8 ±1.5 permissions per app, and the average number of permissions used by malware apps was 7.09 ±3.6 permissions per app. The difference between the number of permissions used per app suggests that the number of permissions can be added as an extra feature to detect COVID-19-themed malware.

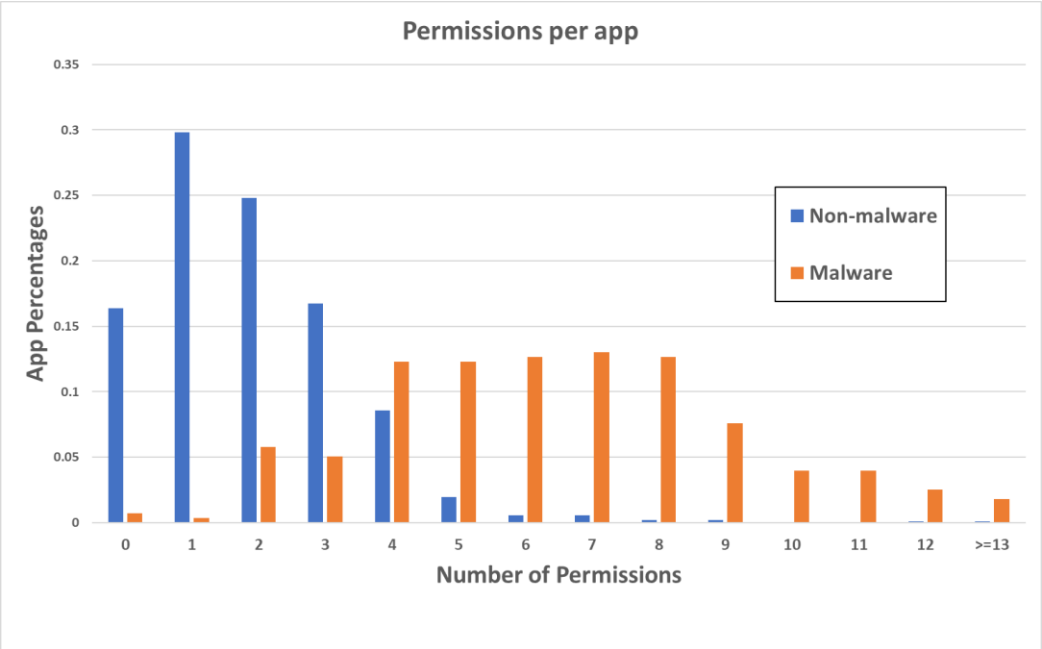

Figure 2. Distribution of app permissions per app class (malware/non-malware) for COVID-19-themed Android apps

Model Creation

The problem of malware detection can be modeled as a machine learning classification problem with two target classes (malware/non-malware). To create the input data for the models we added the total number of permissions used by an app as a extra feature (in addition to each app's permissions).

We used Weka 3.8.6 to create the ML classification models. We created two types of ML modes, using the dataset we obtained in the earlier section (2016 instances), and using SMOTE. Because of the difference in the number of instances (app samples) for each class (1739 non-malware vs 277 malware instances/samples), we use the Synthetic Minority Oversampling Technique (SMOTE) technique to augment the number of samples of the malware class (the smallest of the two classes).

Our nput dataset to train the classification models in Weka after applying SMOTE to balance the classes consisted of 3955 instances (1739 for the non-malware class and 2216 for the malware class) with 205 features (203 app permissions, 1 feature for the class, 1 feature for the total number of permissions). We created classification models with both datasets using OneRule, J48, and Naïve Bayes algorithms. The OneRule and J48 algorithms build decision tree classification models, while the Naïve Bayes builds a probabilistic model for classification. OneRule creates a simple classification tree based only on the attribute/feature with smallest total error as the selected attribute/feature to build the classification model. We selected these models because once the models are trained, these models can be executed very quickly on a smartphone.

Model Evaluation

We used a 10-fold cross validation on each algorithm with a 66% split for each fold (66% of the instances used for model training while the rest used for evaluation per fold). For each dataset, algorithm and class, we obtained the following measurements (summarized in table 2):

- True positive rate (TP Rate): This is the probability that an actual positive will be classified positive.
- False positive rate (FP Rate): This is the probability that an actual negative will test positive. In our case this means that an app that is malware is classified as non-malware and vice versa
- Precision: Proportion of actual positives correctly identified
- F-Measure: A measure of a model's accuracy. It is calculated from the precision and recall. Values tending to 1 means better score.

Table 2. Evaluation metric results for our classification models

Algorithm + Dataset	Class	TP Rate	FP Rate	Precision	F-Measure
OneRule + SMOTE	Non-malware	0.877	0.228	0.751	0.809
	Malware	0.772	0.123	0.889	0.826
J48 + SMOTE	Non-malware	0.98	0.031	0.961	0.971
	Malware	0.969	0.02	0.984	0.977
Naive Bayes + SMOTE	Non-malware	0.939	0.136	0.844	0.889
	Malware	0.864	0.061	0.948	0.904
OneRule	Non-malware	0.971	0.343	0.947	0.959
	Malware	0.657	0.029	0.781	0.714
J48	Non-malware	0.983	0.217	0.966	0.974
	Malware	0.783	0.017	0.879	0.828
Naive Bayes	Non-malware	0.965	0.119	0.981	0.973
	Malware	0.881	0.035	0.8	0.838

Discussion

When creating our models using OneRule, we found that the *total number of permissions* was selected as the attribute/feature to build the OneRule models, meaning that this attribute alone is the best one to potentially detect a COVID-19-themed app as malware or not. We expected this result from our analysis on the average number of permissions for malware/non-malware apps (figure 2). We observed that in general all algorithms performed relatively well, however models based OneRule have the worst metrics when compared against J48 and Naïve Bayes models.

The best overall result on the compared models was obtained using the J48 algorithm with SMOTE to increase the number of instances of the minority class (malware instances in our case), with the best classification results in the metrics, especially those associated with the malware class. Our results suggest that a static malware detector specifically targeting pandemic-themed apps can be implemented directly in the Android OS, as the OS detects the permissions during APK install time. This detector can be used during pandemic times especially when apps are attempted to be installed from non-trusted sources, which was the case of COVID-19-themed malware in many cases.

Conclusion

We have reviewed the use of mobile and smartphones during pandemics with their use cases. We also reviewed the growth of malware and pandemic-related malware during the COVID-19 pandemic. Finally, we evaluated the use of permissions and ML methods to detect COVID-19-themed malware. From our review of mobile malware during pandemic, especially from our

review of COVID-19 relate malware, we suggest the following countermeasures to minimize the impact of cyberattacks to smartphone users in future pandemics:

- Increase the awareness of cybersecurity and cyberhygine specifically focused on cybersecurity for smartphones during a pandemic. This could be achieved by cybersecurity education before a pandemic and public announcements about mobile malware risks during a pandemic
- Suggest that any kind of mobile app to be installed during a pandemic to be installed from a trusted source (e.g., Google Play Market, Apple App Store). This makes malware to be harder to be distributed and installed, especially during pandemics
- Implement a static malware detector as part of the mobile OS as a software update during pandemics that can detect and alert about possible malware being installed when a non-traditional/trusted source (e.g., via SMS or social networks camouflaged as a pandemic-related app). Our research suggests that in future pandemics, pandemic-themed apps can be detected using a combination of app permissions and ML methods

Acknowledgement

This work was supported by the U.S. National Science Foundation under grant award # 1950416.

References

[1] Perez, A. J., & Zeadally, S. (2021). Recent advances in wearable sensing technologies. Sensors, 21(20), 6828.

[2] Wang, L., He, R., Wang, H., Xia, P., Li, Y., Wu, L., ... & Xu, G. (2021). Beyond the virus: a first look at coronavirus-themed Android malware. Empirical Software Engineering, 26(4), 1-38.

[3] Khan, N. A., Brohi, S. N., & Zaman, N. (2020). Ten deadly cyber security threats amid COVID-19 pandemic.

[4] TechXplore. Global cost of cybercrime topped $ 6 trillion in 2021: defence firm (2022). Available online: https://techxplore.com/news/2022-05-global-cybercrime-topped-trillion-defence.html

[5] Eysenbach, G. (2003). SARS and population health technology. Journal of Medical Internet Research, 5(2), e882.

[6] Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., ... & Jha, S. K. (2020). A survey of COVID-19 contact tracing apps. IEEE access, 8, 134577-134601.

[7] Bengtsson, L., Gaudart, J., Lu, X., Moore, S., Wetter, E., Sallah, K., ... & Piarroux, R. (2015). Using mobile phone data to predict the spatial spread of cholera. Scientific reports, 5(1), 1-5.

[8] Priye, A., Bird, S. W., Light, Y. K., Ball, C. S., Negrete, O. A., & Meagher, R. J. (2017). A smartphone-based diagnostic platform for rapid detection of Zika, chikungunya, and dengue viruses. Scientific reports, 7(1), 1-11.

[9] Yu, H., Yang, F., Rajaraman, S., Ersoy, I., Moallem, G., Poostchi, M., ... & Jaeger, S. (2020). Malaria Screener: a smartphone application for automated malaria screening. BMC Infectious Diseases, 20(1), 1-8.

[10] Natesan, M., Wu, S. W., Chen, C. I., Jensen, S. M., Karlovac, N., Dyas, B. K., ... & Ulrich, R. G. (2018). A smartphone-based rapid telemonitoring system for Ebola and Marburg disease surveillance. ACS sensors, 4(1), 61-68.

[11] Horvath, T., Azman, H., Kennedy, G. E., & Rutherford, G. W. (2012). Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database of Systematic Reviews, (3).

[12] Devi, B. R., Syed-Abdul, S., Kumar, A., Iqbal, U., Nguyen, P. A., Li, Y. C. J., & Jian, W. S. (2015). mHealth: An updated systematic review with a focus on HIV/AIDS and tuberculosis long term management using mobile phones. Computer methods and programs in Biomedicine, 122(2), 257-265.

[13] Gostin, L. O., Cohen, I. G., & Shaw, J. (2021). Digital health passes in the age of COVID-19: Are "vaccine passports" lawful and ethical?. JAMA, 325(19), 1933-1934.

[14] State of Florida, Office of the Governor. Executive Order Number 21-81 (Prohibiting COVID-19 Vaccine Passports) (2021). Available online: https://www.flgov.com/wp-content/uploads/2021/04/EO-21-81.pdf

[15] Li, T., Cobb, C., Yang, J. J., Baviskar, S., Agarwal, Y., Li, B., ... & Hong, J. I. (2021). What makes people install a COVID-19 contact-tracing app? Understanding the influence of app design and individual difference on contact-tracing app adoption intention. Pervasive and Mobile Computing, 75, 101439.

[16] Doerrfeld, B. A pandemic side effect: Rampant technical debt (2022). Available online: https://devops.com/a-pandemic-side-effect-rampant-technical-debt/

[17] Statistica, Number of detected malicious installation packages on mobile devices worldwide from 4th quarter 2015 to 2nd quarter 2021 (2021). Available online: https://www.statista.com/statistics/653680/volume-of-detected-mobile-malware-packages/

[18] Check Point Blog. The mobile malware landscape in 2022 – Of Spyware, Zero-Click attacks, Smishing and Store Security (2022). Available online: https://blog.checkpoint.com/2022/09/15/the-mobile-malware-landscape-in-2022-of-spyware-zero-click-attacks-smishing-and-store-security/

[19] Statistica, Distribution of new mobile malware worldwide in 2021, by type (2021). Available online: https://www.statista.com/statistics/653688/distribution-of-mobile-malware-type/

[20] Manzil, H. H. R., & Naik, M. S. (2022, January). COVID-Themed Android Malware Analysis and Detection Framework Based on Permissions. In 2022 International Conference for Advancement in Technology (ICONAT) (pp. 1-5). IEEE.