AppMoD: Helping Older Adults Manage Mobile Security with Online Social Help

ZHIYUAN WAN, Zhejiang University, China and University of British Columbia, Canada
LINGFENG BAO*, Zhejiang University City College, China
DEBIN GAO, Singapore Management University, Singapore
ERAN TOCH, Tel Aviv University, Israel
XIN XIA, Monash University, Australia
TAMIR MENDEL, Tel Aviv University, Israel
DAVID LO, Singapore Management University, Singapore

The rapid adoption of Smartphone devices has caused increasing security and privacy risks and breaches. Catching up with ever-evolving contemporary smartphone technology challenges leads older adults (aged 50+) to reduce or to abandon their use of mobile technology. To tackle this problem, we present AppMoD, a community-based approach that allows delegation of security and privacy decisions a trusted social connection, such as a family member or a close friend. The trusted social connection can assist in the appropriate decision or make it on behalf of the user. We implement the approach as an Android app and describe the results of three user studies (n=50 altogether), in which pairs of older adults and family members used the app in a controlled experiment. Using app anomalies as an ongoing case study, we show how delegation improves the accuracy of decisions made by older adults. Also, we show how combining decision-delegation with crowdsourcing can enhance the advice given and improve the decision-making process. Our results suggest that a community-based approach can improve the state of mobile security and privacy.

CCS Concepts: • **Security and privacy → Usability in security and privacy**; *Malware and its mitigation*; • **Human-centered computing → Empirical studies in ubiquitous and mobile computing**.

Additional Key Words and Phrases: Mobile smartphones, security, older adults, decision delegation

ACM Reference Format:
Zhiyuan Wan, Lingfeng Bao, Debin Gao, Eran Toch, Xin Xia, Tamir Mendel, and David Lo. 2019. AppMoD: Helping Older Adults Manage Mobile Security with Online Social Help. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 154 (December 2019), 22 pages. https://doi.org/10.1145/3369819

*This is the corresponding author.

Authors' addresses: Zhiyuan Wan, Zhejiang University, Hangzhou, China, University of British Columbia, Vancouver, Canada, wanzhiyuan@zju.edu.cn; Lingfeng Bao, Zhejiang University City College, Hangzhou, China, baolf@zucc.edu.cn; Debin Gao, Singapore Management University, Singapore, dbgao@smu.edu.sg; Eran Toch, Tel Aviv University, Tel Aviv, Israel, erant@tauex.tau.ac.il; Xin Xia, Monash University, Melbourne, Australia, xin.xia@monash.edu; Tamir Mendel, Tel Aviv University, Tel Aviv, Israel, tamirmendel@mail.tau.ac.il; David Lo, Singapore Management University, Singapore, davidlo@smu.edu.sg.

2474-9567/2019/12-ART154 $15.00
https://doi.org/10.1145/3369819

1 INTRODUCTION

Smartphone ownership is rapidly growing worldwide [58], quickly becoming the primary way people interact with each other, their banks, shops, and health providers. At the same time, contemporary mobile applications have access to highly sensitive information, such as users' location, contacts, private messages, and browsing information [30]. If the apps are malicious, they can transfer personal and confidential information to unauthorized third parties, enabling adversaries to infer personal information [59], and cause inconvenience or even financial and physical harm [40]. While smartphones pose new privacy and security risks to users, people's awareness of those risks is significantly low [8]. This is especially true for vulnerable populations, such as the elderly, children, and minorities [62]. Specifically, while older adults, defined as people who are 50 or older [16], are the focus of many mobile technologies that aim to help independent living and health for older adults [15], their ability to control this technology is questionable [20, 22, 42].

Several studies have shown that older adults are more vulnerable to security and privacy threats than the general population [36, 68]. [58] reported that 67% of older adults in the United States own a smartphone in 2018. Meanwhile, 73% of older adults claimed that they need additional help to use smartphones as reported in [4]. While older adults cannot be addressed as a homogeneous population, aggregation of survey data shows that privacy and security poses additional and difficult challenges to this particular population. Older adults perceive security and privacy as important issues [67], but most of them feel that they have low self-efficacy in addressing them [36]. Older adults generally exhibit lower levels of technical understanding in comparison with younger adults [43]. The consequences of these challenges can be severe. Older adults have negative attitudes toward the risks involved in using new technologies, and therefore they tend to underuse those technologies [13, 39, 41]. The growing importance of mobile technologies indicates that the population of older adults now suffers from reduced access to social networking, commerce, and mobile health [25].

Crowdsourcing is one of the promising approaches that can help users manage their security and privacy. In prior studies, various crowdsourcing mechanisms were proposed, e.g., to find minimal sets of permissions that still preserve application usability for diverse users [24], to detect users' expectations of sensitive resources that applications use [33], and to enable expert users to make right permission granting decisions for inexperienced users [46]. However, the proposed crowdsourcing mechanisms did not consider how willing individuals are in disclosing their personal information. Past studies have indicated that privacy concerns are barriers to the adoption of modern technologies for older adults [18, 41]. Additionally, past studies have also reported that older adults have different cognitive and technical abilities as compared to younger adults [9, 14]. Hence, any approach that tries to help older adults should take into account their privacy concerns, as well as cognitive and technical abilities [62].

In this work, we propose a community-based approach to help older adults manage mobile security. We look at the way close social contacts (e.g., family and friends) can be used to help older adults with security management for mobile phones. The approach takes the privacy concerns, as well as cognitive and technical abilities of older adults into account. Prior studies have shown that older adults prefer to receive help from their social connections [16]. The main challenge in leveraging social connections is to reduce the obstacles due to the physical distance. To demonstrate our vision, we have proposed AppMoD (**App**lication **M**ediate-**o**n-**D**emand), an Android app that helps older adults tackle physical distance and make security and privacy decisions. AppMoD enables collaboration between an advisee (an older adult) and and advisor (her social contact). Specifically, when the advisee encounters a security and privacy anomaly on her phone, she can notify the designated advisor via AppMoD; the advisor can either provide a suggestion or make the decision for the advisee. AppMoD is implemented as a mobile app and installed on the mobile phones of both the advisee and the advisor.

To examine the effectiveness of our approach, we conducted three user studies and recruited 50 pairs of participants that fall into two groups: (1) participants between the ages of 18 and 40 as advisors, and (2) participants

above the age of 50 as advisees. In the user studies, we simulated security and privacy anomalies and sent corresponding notifications to AppMoD that were installed on the phones of our *advisee* participants. The correct responses to simulated anomalies were confirmed by security experts, serving as ground truth. To evaluate the extent to which participants make correct decisions, we measure the response *accuracy*, which is the percentage of correct responses among a set of responses.

We investigated three research questions in the user studies:

Study 1: we investigated whether delegation improves our older adult participants' ability to make correct security and privacy decisions. The delegation of security and privacy decisions is the process wherein the advisees ask advisors to make decisions for them.

Study 2: we explored if crowdsourcing facilitates security and privacy decision making. We aggregated the decisions made by the crowd in the first study, and fed to AppMoD in crowdsourcing mode. When encountering a particular anomaly, the participants could consult those aggregate historic decisions.

Study 3: we looked into the learning effect in the process of delegating decisions. We refer to learning effect as the phenomenon in which advisees acquire security-related knowledge in the process of delegating decisions; the acquired knowledge helps them make correct decisions on similar anomalies. Note that while our studies are heavily influenced by Android platform, we believe our results could generalize to other mobile platforms and settings.

2 BACKGROUND AND RELATED WORK

2.1 Older Adults and Mobile Security

Older adults are adopting mobile technology at a rapid pace, from 35% in 2015 to 67% in 2018. Smartphones are widespread among young adults for years, but older adults are catching up [58]. Nevertheless, older adults face barriers to using and adopting new technologies. Approximately 34% of older adults users declare they have low confidence in their ability to use electronic devices [4]. Moreover, they struggle more to handle smartphones security threats than younger adults [13, 39], and they experience difficulties with managing privacy setting to protect their personal information [20, 22, 41, 42, 67].

Having technology tailored and simplified allow use by older adults in ubiquitous and mobile computing [15], but the approach could quickly become outdated as the technology moves on and does not adapt to fit the learning capabilities of older adults. With the ongoing technological innovations, the need to constantly learn and adapt to new user interfaces and risks is a constant frustration to the older adult population [20]. As Damodaran et al. report based on a survey of 323 older adults [16]:

> *Although some of the frustrations and difficulties reported by participants in this study may be reduced or avoided through improved design, given the diversity of skills and capabilities amongst older people as well as the diversity and complexity of technology and the rapid rate of change and development, there is likely to be an on-going need for learning and support. Older people report wanting help and sympathetic support on demand and at point of need. All participants stated that they valued timely support to solve technology related problems.*

Older adults prefer to get support from close social environment, such as friends and families, in ongoing engagement with technology [13, 16]. For example, in the case of mobile health, training and support were found to be the strongest facilitators in mobile technology use [44]. A recent study has shown that family members of older adults are willing to provide more help in mobile security and privacy to their related older adults than they currently do [38]. However, the older adults and their social contacts that can provide help may not be located at the same place. The physical distance could limit the ability to communicate the problem and the solution.

2.2 Malicious App Behavior

One of the most pressing issues of mobile security and privacy is malicious app behavior [6, 21]. Defining the mobile security and privacy problem as a way to separate malicious from benign behaviors allows us to detect activities that do not comply with the expected pattern [5, 10], such as malicious advertising, compromising personal data and malicious chargeback [57]. However, it can be a difficult task to identify in a precise and automatic way whether an activity of an application is a threat or not. In many cases, the app behavior cannot be classified as absolutely malicious and contains security or privacy threats in which the user may not be aware. For example, applications may access sensitive information such as locations and send it for ambiguous purposes [28]. In this case, the user should decide the legitimacy of the activity related to the application.

Smartphone platforms, such as Android or iOS, delegate privacy decisions directly to users, asking them to grant or deny applications access to resources. However, users often neither read nor understand the permissions [19, 29]. Prior studies investigate the users' preference regarding their access to sensitive data [3, 23, 54] and propose permission models with more contextualized permissions and on a more granular level [60, 64, 65]. Studies also suggest automating decision-making in security [34, 35, 52, 69] and increasing users' awareness of resource usage [49–51].

2.3 Managing Mobile Security and Privacy

Previous works have looked at ways to enhance permission systems as one of the main tools to address privacy and security systems. However, permission systems require that users understand and make decisions about the permissions, but few users actually read the permissions of an app and even fewer understand the impact of permissions [19, 29]. Users are often surprised by the ability of background applications to collect personal data, and do not understand the connections between permission types and sensitive resources [26, 63]. Interacting with mobile security and privacy interfaces is inherently difficult. Mobile users could become desensitized to future requests and make poor decisions after receiving a significant number of notifications [65].

To prevent users from *habituation*, researchers conducted studies to reduce the user's involvement in decision making. Prior research developed techniques to cluster users [35, 52] and built recommender systems [69]. Liu et al. subsequently predicted user preferences in future permission requests using inferred user clusters and developed a privacy assistant to recommend privacy settings to users [34]. Motivated by contextual integrity, Wijesekera et al. [64] proposed applying machine learning techniques to dynamically grant app permissions, and Tsai et al. [60] proposed a user interface design for that system. Furthermore, Wijesekera et al. [65] implemented the system for the Android platform and performed a field study to evaluate the effectiveness of the system.

Mobile privacy and security infrastructure can be improved by employing privacy-by-design approaches. Crowdsourcing was suggested as a way to evaluate the privacy of mobile app features [7]. Privacy-by-design approaches were proposed to address the privacy in the design of apps to from an early stage of the development process [55]. *Access Control Gadgets* were proposed as a mechanism to associate sensitive resource accesses to particular user interface elements [49–51]. The mechanism enables users implicitly control access to sensitive resources and increases the awareness on resource usage.

Several recent papers have suggested ways that leverage social influence and support to help users to manage mobile security and privacy. Social influence and support have several significant advantages over other methods in helping older adults to manage mobile security and privacy. Social ties were found to have more impact in privacy and security advice than authoritative sources [37]. Older adults are more comfortable to receive help from their close social circle [13, 16]. Even more importantly, the close social circle could have a better understanding of the preferences of older adults [38]. To this end, prior studies have shown how security behavior on Facebook can spread through social visibility mechanisms [17], how users seek security advice from others [47], and how security advice can be selected [48]. Aljallad et al. have designed a prototype which helps individuals

collaborate with people they know, in order to make decisions regarding app permissions [2]. Rashid et al. have used crowdsourcing to provide users with online recommendation of permission control [46].

2.4 Research Objectives

To address the challenges facing older adults, we propose a community-based security management approach to enhance security and privacy decision-making processes. The approach primarily relies on the existing social contacts of older adults to help them effectively obtain trusted assistance, support, and guidance. Specifically, an older adult can delegate the privacy and security decisions to a third party in her social circle. For example, if one does not know if an application behaves in a normal way or if a permission should be granted to an app, the decision can be delegated to their offspring for an advice or an actual decision. Our approach is more general in nature than existing technologies. It provides assistance to address a wide range of security and privacy challenges, rather than focusing on app permissions [2, 46]. We test the feasibility of our approach with the particular population of older adults. We then quantitatively evaluate how the response accuracy is associated with the use of our approach, and related to different types of security and privacy challenges. Finally, we test how combining the help from family members and the crowdsourcing information can result in accurate responses.

3 APPMOD: DELEGATING DECISIONS

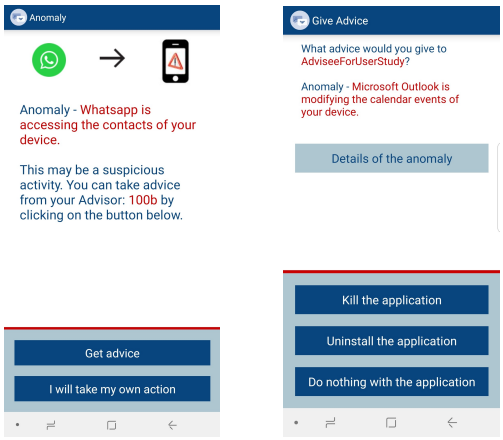
To demonstrate our approach for community-based security management, we have designed and implemented AppMoD, a system that enables delegating security and privacy decisions to other people. Specifically, the delegation involves a pair of participants, an advisee and an advisor. The advisee first finds a person from her circle of trust as the advisor via AppMoD. When a security and privacy anomaly arises, AppMoD summarizes the risk in an informative manner and alerts the advisee. The advisee can ask for a recommendation via AppMoD by forwarding the informative notification to her advisor. Once the advisor receives the request, she can refer to the detailed context of each anomaly provided by AppMoD, and suggest action by choosing among predefined options responding to the anomaly.

We provided three options of actions, i.e., “Kill”, “Uninstall” and “Do nothing”. The options are suggested by prior studies in intrusion detection for mobile devices (e.g., [53]). Specifically, the action “Kill” allows users to kill the process(es) of an app - temporarily shutting it down; the action “Uninstall” empowers users to uninstall an app - permanently shutting it down; the action “Do nothing” indicates that users prefer to do nothing in response to the anomaly. AppMoD provides options of actions, instead of options of perceptions (i.e., normal/abnormal behavior), in order to capture the explicit reactions of the users to the anomalies. After the advisee receives the suggestion, she can choose to accept or reject it.

AppMoD supports two modes of delegating decisions:

Non-crowdsourcing mode. Fig. 1 displays the two major user interfaces of AppMoD in non-crowdsourcing mode: the interface presented to the advisee (a) and the interface for the advisor (b). Once receiving the alert of a security and privacy anomaly, the advisee could choose to take her own action or to ask for advice. As soon as the advisee clicks “Get advice”, the advisor would receive the request and give advice by choosing from the action options.

Crowdsourcing mode. AppMoD also supports the crowdsourcing ability for community-based security management. As growing pairs of participants interact via AppMoD, AppMoD continues to collect advisors’ selection between the predefined options for each specific anomaly in the community. When an advisee encounters an anomaly that previously occurred in the community, AppMoD provides the statistical information for the selection of each option. Both advisees and advisors could leverage the information to make decision. Fig. 2 displays the user interface parts of AppMoD with crowdsourcing information: the interface presented to the advisee (a) and the interface for the advisor (b).



(a) Advisee (b) Advisor

Fig. 1. User interfaces in non-crowdsourcing mode with options in response to anomalies.

(a) Advisee (b) Advisor

Fig. 2. User interfaces in crowdsroucing mode with options in response to anomalies.

We present a typical usage scenario of how an advisee and an advisor interact via AppMoD in Fig. 3. The typical usage scenario consists of five steps. First, the advisee receives the notification of a security and privacy anomaly of a particular app that happens on her smartphone via AppMoD. Second, the advisee seeks advice for this anomaly from her advisor (*Advice* response), instead of taking her own action (*Own Action* response). Third, the advisor immediately receives the request, and checks the details about the anomaly via AppMoD. Fourth, the advisor gives the advice by choosing between a predefined options. Five, the advisee immediately receives the advice and chooses to follow it, instead of ignoring it and taking her own action.

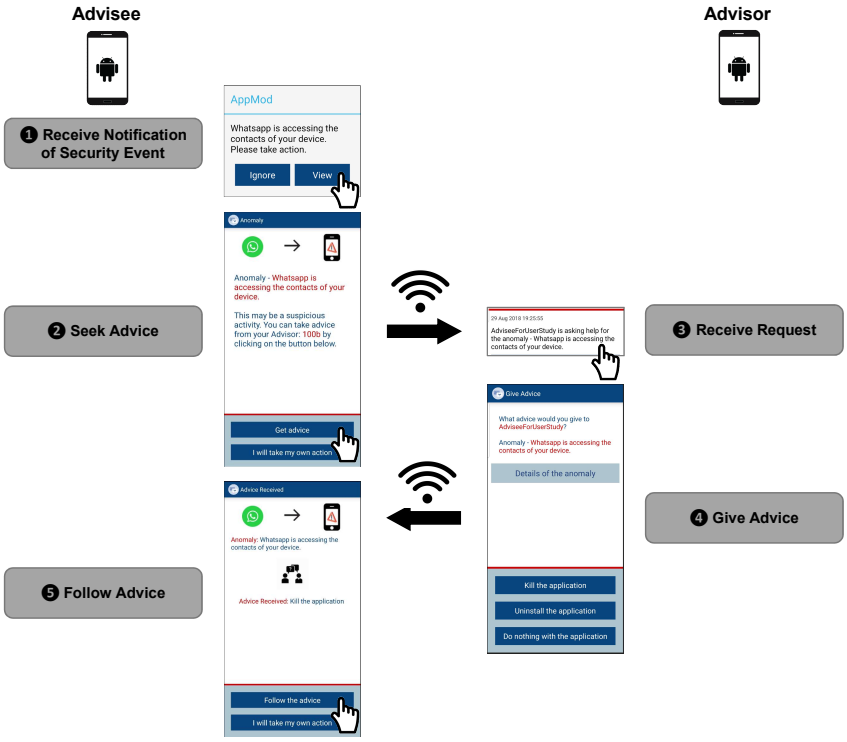


Fig. 3. A usage scenario of AppMoD.

4 METHODOLOGY

4.1 Recruitment and Incentives

We recruited paired participants in two age groups: (1) advisors: participants between the ages 18 and 40, and (2) advisees: participants above the age of 50. We sent emails to our contacts in each school of Singapore Management University (SMU), and asked their help to disseminate our user study within the university. All participants were at least 18 years old and located in Singapore. Each Participant was paid 40-60 Singapore dollars for completing the user study. The user study was approved by SMU IRB. Participants were instructed about how our app works and how to use our app (e.g., the difference between the options “Kill” and “Uninstall”), and informed that all the notifications are simulated and harmless. The participants above the age of 50 could be the friends and relatives of the students who met the requirements (e.g., parents and grandparents). The participants agreed on our online informed consent form before the user study. The consent form indicates that they are entirely voluntary and they participate in our user study in a paired form.

We recruited a total of 50 pairs of participants (34 unique). Specifically, 18 pairs of participants took part in study 1 (batch 1, 67% of the advisors perceived themselves as being proficient in mobile security), 16 pairs in study 2 (batch 2, 75% of the advisors perceived themselves as being proficient in mobile security), and 16 pairs in study 3 (batch 3). Note that the 16 pairs of participants in study 3 were invited from the 18 pairs from study 1. We surveyed the advisors about their proficiency in mobile security by providing a statement: *I am proficient in*

mobile security. We asked them to respond on a 5-point Likert scale (*strongly disagree*, *disagree*, *neutral*, *agree*, *strongly agree*) to this statement. We regard those who selected *strongly agree* and *agree* as proficient users in mobile security.

4.2 Experimental Setup

To test user responses, we have compiled four sets of anomalous and normal behaviors of 23 Android apps (A1, A2, A3 and A4). We selected apps based on their popularity on Google Play. The apps fall into five categories and access various sensitive resources. We carefully reviewed the declared permissions of those apps and compiled anomalous and normal behaviors for each app accordingly. Please refer to Appendix A for the detail of the anomaly notifications. Furthermore, we confirmed the true/false positiveness and appropriate response for each notification with security experts. Specifically, "*Kill*" or "*Uninstall*" is the accurate response to a real anomaly; meanwhile, "*Do nothing*" is the accurate response for a false-positive anomaly.

The anomaly notifications are categorized into difficulty levels, i.e., "Hard", "Medium", and "Easy", based on how difficult it is for a regular Android user to determine whether an anomaly is true or false positive. The nature of the target app and the resource accessed affects the difficulty of an anomaly. The first author made an initial difficulty categorization; to reduce subjectivity from the categorization, the second and third authors performed a reassessment before determining the final categorization.

To capture participants' responses when making security and privacy decisions at the moment anomalies are occurring, we use the Experience Sampling Method (ESM) [31]. ESM has been used in a number of studies on mobile devices [11, 61]. It refers to a method for collecting data from a participant in the natural context of everyday life. In an ESM study, participants are reminded randomly during fixed windows of time. During each of our user studies, we sent out the simulated notifications to our participants via AppMoD. We followed a random process to pick notifications from a set of the anomaly notifications (i.e., A1, A2, A3 or A4 as listed in Appendix A) until each anomaly notification in the set has been responded by each pair of participants at least once. Once at least one response for each anomaly was collected, we stopped sending out notification to avoid annoyance due to unnecessary repetition. Due to the random process, each pair of participants might receive a particular anomaly notification multiple times. We indicate the set of responses that a pair of participants made to an anomaly notification for the first time as "*first time*" responses, which are part of "*all*" responses from that pair. The number of "first time" responses received for all participants is equal to the number of participants $\times$ the size of an anomaly notification set (i.e., 10). In total, we sent 1,422 notifications and received 1,370 responses from advisees and advisors. The response rate is 96.3%. Out of the 52 non-responded notifications, 14 correspond to requests for advice that did not get answered.

This work revolves around the following research questions:

- **RQ1.** Does delegation of security and privacy decisions achieve better performance in terms of mitigating security risks for older adults?
  If so, delegation could protect older adults against security breaches from untrusted or potentially malicious apps, and reduce the burden of making decisions for older adults when using mobile phones.
- **RQ2.** Does crowdsourcing help achieve better decisions in terms of mitigating security risks?
  If so, crowdsourcing information of security and privacy decisions should be collected and provided to users while making decisions.
- **RQ3.** Can older adults acquire security-related knowledge in the delegation process and achieve better accuracy in responses to anomalies?

4.3 Measures

Our analysis is based on several measures that were used throughout the three user studies. On the basis of AppMoD's delegation system, we first portray the advisee's behavior in terms of whether they have decided to take their own action (*Own Action* response) or to ask for advice (*Advice* response). To figure out the preference of the participants between taking their own action and seeking advice, we measure the *Own Action* ratio and *Advice* ratio of a set of responses.

To measure the extent to which participants make correct responses, we define the response *accuracy* of a set of responses as the percentage of correct responses among that set of responses. "*Kill*" or "*Uninstall*" is the correct response for a real anomaly, and "*Do nothing*" is the correct response for a false-positive anomaly. The true/false-positive label of an anomaly is determined by security experts beforehand.

4.4 Study 1: Assessing Delegation

Simulated Notification. For each pair of participants , AppMoD randomly picked one anomaly in A1 and sent the corresponding notification on an hourly basis from 8am to 6pm during one week until all the 10 anomalies in A1 were covered. For study 1, we use the non-crowdsourcing mode of AppMoD. In total, we have sent 576 anomaly notifications via AppMoD and received 554 responses across 10 distinct anomalies in A1 (Median: 56[1], Min: 25, Max: 106, SD: 23.6). Note that we provided detailed context for each anomaly to advisors. Specifically, advisors could review the detailed context by clicking on the "Details of the anomaly" button as shown in Fig. 1(b) and Fig. 2(b). For instance, for the anomaly "Whatsapp is accessing the contacts of your device", the detailed context we provided is "This is to allow users to search for their friends and send them messages from inside Whatsapp".

Data Analysis. To answer RQ1, we evaluated the effect of delegation by computing the following measures for "all" and "first-time" responses: (1) *Own Action* and *Advice* ratios, and (2) *accuracy* of *Advice* and *Own Action* responses, respectively.

4.5 Study 2: Assessing Crowdsourcing

Crowdsourcing Information. We leveraged the crowdsourcing mode of AppMoD in study 2. Specifically, we collected the selection percentages of each option for each anomaly from proficient advisors (based on self-reported data) in study 1. The details of the selection percentages are shown in Table 1. In the crowdsourcing mode, AppMoD supplemented action options for each anomaly with crowdsourcing information as selection percentages.

Advisees and advisors both can refer to the crowdsourcing information for making decisions. The user interfaces with crowdsourcing information for advisees and advisors are presented in Fig. 2.

Simulated Notification. For each pair of participants, AppMoD randomly picked and sent one anomaly A1 or A2 on an hourly basis from 10am to 3pm during one week until all 10 anomalies in A1 and all 10 anomalies in A2 have been responded at least once. Note that for anomalies in A1, crowdsourcing mode was used; for anomalies in A2, non-crowdsourcing mode was used. In total, we have sent 221 anomaly notifications and received 214 responses across 10 distinct anomalies in A1 (Median: 19.5, Min: 17, Max: 34, SD: 5.0). Additionally, we have sent 342 anomaly notifications and received 327 responses across 10 distinct anomalies in A2 (Median: 31.5, Min: 18, Max: 58, SD: 13.2).

Data Analysis. In pursuit of answering RQ2, we evaluated the effect of crowdsourcing information by computing the following measures for the responses to anomalies in A1: (1) *Own Action* and *Advice* ratios for "all" responses; (2) *accuracy* for "all" *Own Action* and *Advice* responses.

[1]Notifications per anomaly

Table 1. Option selection percentages from proficient advisors in study 1. The most frequently selected options for anomalies are marked in bold.

Anomaly (A1)	Uninstall	Kill	Do nothing
Candy Crush is reading your contacts and sending it to www.hackme.com.	**89%**	11%	0%
Clock is accessing your geolocation and sending it out.	40%	**50%**	10%
Facebook is accessing the location of your device.	36%	24%	**40%**
Gmail is modifying the calendar events of your device.	**58%**	17%	25%
Gmail is sending emails to everyone in your address book.	46%	**51%**	3%
Instagram is accessing the camera of your device.	22%	0%	**78%**
Sudoku is reading your sim card info and sending it to www.abnormal.com.	**73%**	27%	0%
Whatsapp is accessing the contacts of your device.	39%	6%	**56%**
Whatsapp is making phone calls to 61234567 (unknown).	**50%**	**50%**	0%
YouTube is accessing the microphone of your device.	10%	40%	**50%**

4.6 Study 3: Assessing Learning Effects

Returned Participants. To evaluate the learning effect, participants from study 1 were asked to return and engage in study 3. To eliminate the noise induced by simple memorization of prior responses, we provided brand new anomalies (A3 and A4 as shown in Appendix A) to our participants. To eliminate other sources of noise, the new anomalies retain some features of the anomalies in A1, e.g., app popularity, difficult level, and resource accessed. The 10 anomalies in A3 are derived from 7 subject apps from the free apps with comparable downloads that falls into the same 5 categories as A1. The apps access the same sensitive resources, with same difficulty level and true/false positiveness as the corresponding anomalies in A1. In comparison with A3, the 10 anomalies in A4 come from the same 7 apps but access different sensitive resources and with opposite true/false positiveness.
Simulated Notification. For each pair of participants, AppMoD in non-crowdsourcing mode randomly picked and sent one anomaly notification in A3 and A4 on an hourly basis from 10am to 3pm during one week until all 10 anomalies in A3 and all 10 anomalies in A4 were covered. In total, we sent 283 anomaly notifications and received 275 responses across 10 distinct anomalies in A3 (Median: 27.5, Min: 20, Max: 36, SD: 4.1). Additionally, we sent 269 anomaly notifications and received 260 responses across 10 distinct anomalies in A4 (Median: 26, Min: 19, Max: 40, SD: 6.2).
Data Analysis. In order to answer RQ3, we computed *accuracy* for "all" the responses to anomalies in A3 and A4. We then compared the computed accuracy with the corresponding measure in study one. Furthermore, we evaluated the effect of resource accessed on response accuracy, by comparing the response accuracy of anomalies in A3 with the response accuracy of anomalies in A4.

5 RESULTS

In this section, we present the results of our user studies and answers to the research questions.

5.1 RQ1. Delegation

5.1.1 Accuracy of Responses. Out of all the 554 responses, the overall response accuracy is 66.1%. Out of the 278 *Advice* responses, the correct responses account for 69.4%. Among the 276 *Own Action* responses, the correct responses account for 62.7%. Generally, participants who seek advice achieve a higher accuracy. We plot the accuracies of *Advice*/*Own Action* responses across anomalies in A1 as shown in Fig. 4: the accuracies of *Advice* responses vary from 23.8% to 96.3% across distinct anomalies (mean: 67.4%, median: 71.0%); the accuracies of *Own Action* responses vary from 18.8% to 94.4% across distinct anomalies (mean: 62.0%, median: 62.5%). To analyze the significance of this improvement, we performed a Wilcoxon signed-rank test [66] between the accuracies of *Own Action* responses and those of *Advice* responses across anomalies, and observed a significant difference (p = 0.005859, $p < 0.05$). We further analyze the accuracy by looking at the effect size between the accuracies of *Own*

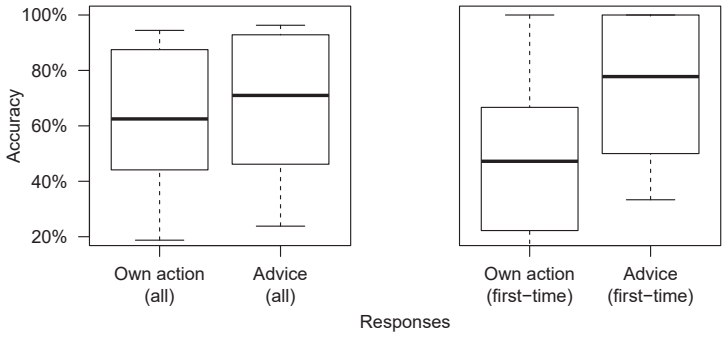


Fig. 4. Accuracies of *Own Action* and *Advice* responses to anomalies in study 1.

Action and those of *Advice* responses. The effect size, as measured using Cliff's delta, is 0.16, which corresponds to a small but non-negligible difference.

We then explore the accuracy for the 180 first-time responses. The accuracy of first-time responses is 64.4%. Out of the 102 first-time *Advice* responses, the correct responses account for 73.5%. Among the 78 first-time *Own Action* responses, the correct responses account for 52.6%. Generally, participants who seek advice for the first-time response achieve a higher accuracy. We plot the accuracies of first-time responses across anomalies in A1 as shown in Fig. 4: the accuracies of first-time *Advice* responses vary from 33.3% to 100% across distinct anomalies (mean: 73.7%, median: 77.8%); the accuracies of first-time *Own Action* responses vary from 14.3% to 100% across distinct anomalies (mean: 50.8%, median: 47.2%). To analyze the significance of this difference, we performed a Wilcoxon signed-rank test between the accuracies of first-time *Own Action* and those of first-time *Advice* responses across anomalies, and observed a significant difference ($p = 0.009152$, $p < 0.05$). We further analyze the accuracy by looking at the effect size between the accuracies of first-time *Own Action* and those of first-time *Advice* responses. The effect size, as measured using Cliff's delta, is 0.42, which indicates a medium difference.

5.1.2 Delegation Preference. Out of all the 554 responses, we have 50.2% responses that seek advice from advisors and 49.8% responses that performed the action independently. The *Advice* ratios vary from 43.8% to 56.8% across distinct anomalies in A1. For the 180 first-time responses, we have 56.7% responses that seek advice from advisors and 43.3% responses that take own action. The *Advice* ratios vary from 44.4% to 72.2% across distinct anomalies in A1.

Generally, the *Advice* ratio decreases since the first-time response for each anomaly occurred as shown in Fig. 5. This may be because our advisee participants tended to take their own action after receiving advice from their advisors. Conversely, for the two anomalies "*Gmail is modifying the calendar events of your device.*" and "*Candy Crush is reading your contacts and sending it to www.hackme.com.*", our participant tended to seek more advice after receiving advice from their advisors.

The accuracy of first-time Advice *responses is 1.4x higher in comparison with first-time* Own Action *responses.*

Participants tend to seek advice at the beginning: Own Action *responses increase and* Advice *responses decrease as user study proceeds.*

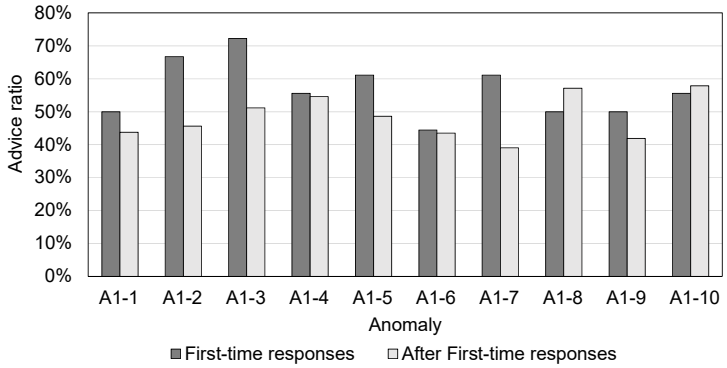


Fig. 5. *Advice* ratios of first-time and after first-time responses across various anomalies. A1-*i* denotes the *i*th anomaly in the list of anomalies A1.

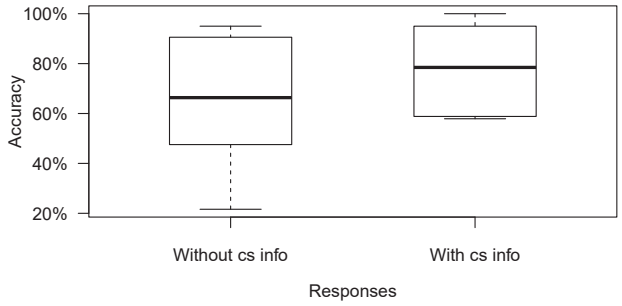


Fig. 6. Accuracies of all responses with and without crowdsourcing (cs) information per anomaly.

5.2 Crowdsourcing

5.2.1 Accuracy of Responses. The overall accuracy of the received 214 responses with crowdsourcing information is 79.9%, which is higher than the accuracy 66.1% of the responses without crowdsourcing information. We plot the overall accuracy of responses with/without crowdsourcing information across anomalies in A1 as shown in Fig. 6. The accuracy of responses varies from 57.9% to 100% across distinct anomalies (mean: 78.7%, median: 78.5%) for responses with crowdsourcing information, which are higher than those of responses without crowdsourcing information (min: 21.6%, max: 95.0%, mean: 64.6%, median: 66.4%). Crowdsourcing information improves the accuracy of responses across distinct anomalies. To analyze the significance of this improvement, we performed a Wilcoxon Signed-rank test between the accuracy of responses with and without crowdsourcing information, and observed a significant difference ($p = 0.001953$, $p < 0.05$) in response accuracy between the two group. We further analyzed the accuracy by looking at the effect size between the accuracy of the responses with and without crowdsourcing information. The effect size, as measured using Cliff's delta, is 0.41, which indicates a medium improvement.

We further explore how crowdsourcing information affects the accuracy of *Own Action* and *Advice* responses respectively. The overall accuracy of *Own Action* responses with crowdsourcing information is 76.8%, which is higher than that of *Own Action* responses without crowdsourcing information (62.7%). The overall accuracy of *Advice* responses with crowdsourcing information is 85.5%, which is higher than that of *Advice* responses

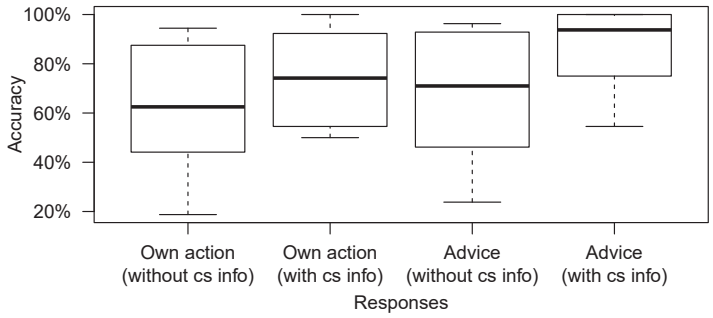


Fig. 7. Accuracies of *Own Action*/*Advice* responses with and without crowdsourcing (cs) information per anomaly.

without crowdsourcing information (69.4%). We plot the overall accuracies of *Own Action* and *Advice* responses with/without crowdsourcing information across anomalies in A1 as shown in Fig. 7. We made the following observations:

(1) With crowdsourcing information, both advisees and advisors achieve higher accuracies across distinct anomalies;
(2) With crowdsourcing information, advisees achieve higher accuracies than advisors without crowdsourcing information in response to identical anomalies;
(3) With crowdsourcing information, the median of accuracies across distinct anomalies for advisors becomes 93.8%. Interestingly, if the advisors were to simply follow the most popular option for each decision in the crowdsourcing information, we would observe a drop in accuracy down to 90%, as following the most popular options favors one incorrect choice across the 10 decisions.

To analyze the significance of the improvements, we performed a Wilcoxon Signed-rank test between the accuracies of *Own Action*/*Advice* responses with and without crowdsourcing information, and observed a significant difference (p = 0.001953/0.003906, $p < 0.05$) in response accuracy between the groups with and without crowdsourcing information. We further analyze the accuracies by looking at the effect sizes between the accuracies of *Own Action*/*Advice* responses with and without crowdsourcing information. The effect sizes, as measured using Cliff's delta is, 0.37/0.52, which indicates a medium difference for *Own Action* responses and a large difference for *Advice* responses.

5.2.2 Delegation Preference. Out of the 214 responses, we have 35.5% responses that seek advice from advisors and 64.5% responses that take own action. The advice ratios vary from 20.6% to 50% across distinct anomalies in A1. Out of the 160 first-time responses, we have 36.9% responses that seek advice from advisors and 63.1% responses that take own action. The advice ratios vary from 18.8% to 50% across distinct anomalies in A1.

We further evaluate how crowdsourcing information affects the delegation preference. We consider the first-time responses and compare the *Advice* ratios for the notifications with and without crowdsourcing information as shown in Fig. 8. The *Advice* ratios dramatically decreased when we provided notifications with crowdsourcing information except for one anomaly "*Clock is accessing your geolocation and sending it out*", which obtains an identical *Advice* ratio when providing crowdsourcing information. We further analyze the significance of the decrease in *Advice* ratios by performing a Wilcoxon signed-rank test. We observe a significant difference in *Advice* ratios between the first-time responses with and without crowdsourcing information (p = 0.009152, $p <$ 0.05). We further look at the effect size between the *Advice* ratios of the first-time responses with and without crowdsourcing information, as measured using Cliff's delta, is 0.85, which indicates a large difference in response accuracy between the two group.

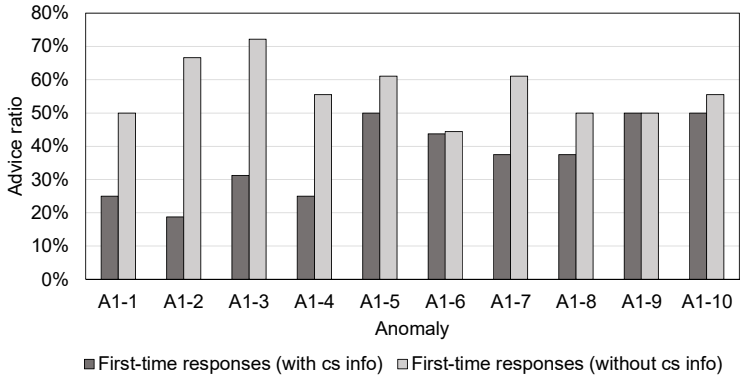


Fig. 8. *Advice* ratios of first-time responses with and without crowdsourcing (cs) information. A1-*i* denotes the *i*th anomaly in the list of anomalies A1.

Crowdsourcing information significantly improves the accuracy of both advisors and advisees across distinct anomalies; the accuracy of all repsonses with crowdsourcing information is 1.2x higher in comparison with the responses without crowdsourcing information.

Advice *ratios dramatically decreased when we provided notifications with crowdsourcing information.*

5.3 RQ3. Learning Effect

5.3.1 Accuracy of Responses. The overall response accuracy of all received 275 responses to anomalies in A3 is 64.4%; while the overall accuracy of all responses for anomalies in A4 is 57.7%. Both are lower than the accuracy 66.1% of all responses for anomalies in A1 in study 1. We plot the response accuracies across distinct anomalies in A1, A3 and A4 respectively as shown in Fig. 9: the response accuracies of anomalies in A3 vary from 35.7% to 92.9% across distinct anomalies (mean: 63.4%, median: 58.8%); the response accuracies of anomalies in A4 vary from 19% to 90% across distinct anomalies (mean: 56.1%, median: 58.5%); the response accuracies for anomalies in A1 vary from 21.6% to 95% across distinct anomalies (mean: 64.6%,median:66.4%). We do not observe an improvement in response accuracy of our participants after study 1. To analyze the significance of the differences, we performed a Wilcoxon Signed-rank test between the response accuracies of anomalies in A1 and A3/A4, and observed non-significant differences (p = 1/0.4922, $p > 0.05$) in response accuracy between the two group.

We further investigate whether accessed sensitive resources in anomalies affect the response accuracy. We consider the same sensitive resources for anomalies in A2 as anomalies in A1. Anomalies that involve the sensitive resource "Calendar" both achieve the lowest response accuracies across distinct anomalies. We consider additional sensitive resources for anomalies in A3 as anomalies in A1. The anomaly that involves the new sensitive resource "Storage" in A3 achieves the lowest accuracies across distinct anomalies.

Participants did not achieve higher accuracy for the anomalies that are arisen by similarly popular apps after previous user study.

5.4 Other Influences

We check whether the advisors' proficiency in mobile security affects the accuracy of their responses. We compared the distributions of response accuracy between two groups of participants, the group with "proficient"

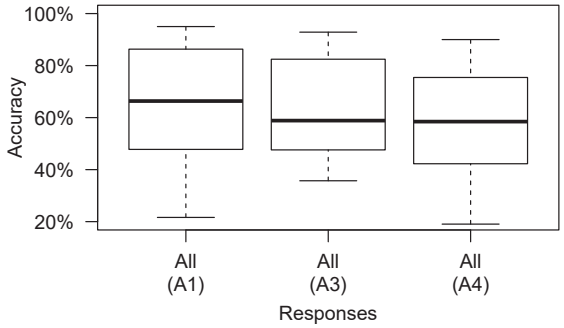


Fig. 9. Accuracies of responses per anomaly in A1 (study 1), A3 (study 3) and A4 (study 3).

advisors versus the group with "non-proficient" advisors, using Wilcoxon rank-sum test. We notice that the group with "proficient" advisors has an average accuracy of 70%, 1.2x higher than the group with "non-proficient" advisors (W = 123.5, p-value = 0.002081). The effect size of Cliff's delta is 0.5459559, which indicates a large difference in response accuracy between the two groups. The results suggest that proficient advisors could significantly benefit the delegation of security and privacy decisions.

In addition, we check whether the difficulty levels of anomalies affect the accuracy of the responses. We compared the distributions of response accuracy across different difficulty levels, using Wilcoxon rank-sum test. To minimize the effect of other factors (e.g., crowdsourcing information and learning effect), we used the responses of anomalies in A1 in study 1 and those in A2 in study 2. We observe that the responses to anomalies in the "Easy" difficulty level elicit an average accuracy of 91%, 1.5x times higher than those in the "Medium" and "Hard" difficulty levels (W = 56, p-value = 0.02188). The effect size (Cliff's delta) is 0.75, which indicates a large difference in response accuracy between "Easy" and "Not Easy" anomalies. The results show that our participants' ability in accurately responding to anomalies in the "Easy" category is statistically significantly and substantially higher than their ability for the other anomalies (those in the "Medium" and "Hard" categories).

We further checked whether app popularity affect the accuracy of responses. We use the responses of anomalies in A2 in study 2, and compare the response accuracy of anomalies in A1 in study 1. The anomalies in A2 come from 8 subject apps that fall into the same 5 categories as A1 but with fewer downloads than A1. Each anomaly in A2 accesses the same sensitive resource and with same difficulty level and true/false positiveness as the corresponding anomaly in A1 to eliminate the effects from other factors. The overall accuracy of responses to anomalies in A2 is 63.0%, which is lower than the overall accuracy 66.1% for anomalies in A1. We plot the accuracies of responses across distinct anomalies in A1 (study 1) and A2 (study 2) as shown in Fig. 10: the response accuracies for anomalies in A2 vary from 25% to 95.1% across distinct anomalies (mean: 60.9%, median: 59.8%); the response accuracies for anomalies in A1 vary from 21.6% to 95% across distinct anomalies (mean: 64.6%, median:66.4%). In general, popular apps achieved higher response accuracies across distinct anomalies. To analyze the significance of the difference due to popularity, we performed a Wilcoxon signed-rank test between the response accuracies of anomalies in A1 and A2, and observed a non-significant difference ($p = 0.375$, $p > 0.05$). Note that we provide descriptions from Google Play to our participants for less popular apps. the descriptions may facilitate making correct decisions.

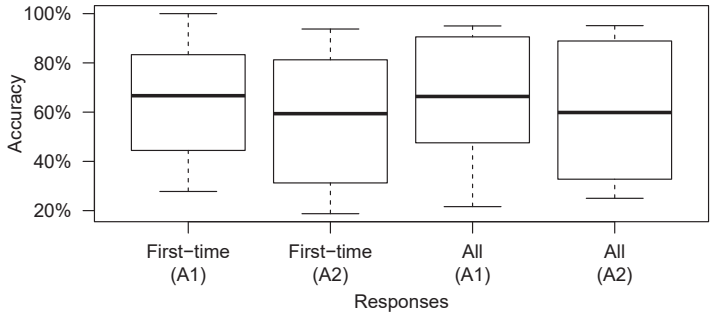


Fig. 10. Accuracies of responses per anomaly in A1 (arise from more popular apps) and A2 (arise from less popular apps).

6 DISCUSSION

6.1 Implications

Our findings have several implications to the growing literature of design and analysis of privacy and security mechanisms for older adults. Several recent papers have suggested to use delegation of decisions as a way to mitigate privacy and security concerns of older adults [20, 22, 45]. While delegation naturally stems from existing practices of older adults [20], it was not tested in an experimental environment. We obtain multiple findings on the effects of delegation through three user studies which involved 50 pairs of participants.

Delegation of Security and Privacy Decisions. Delegation of security and privacy decisions had significantly helped our participants achieve 1.4x higher accuracy in making security and privacy decisions (RQ1). This effect is stronger at the first encounters with anomalies. The delegation ratio at the first encounters with anomalies is 56.7%. As suggested in [20] and [42], delegation of privacy and security management occurs frequently among older adults. Our result is consistent with previous findings. Our proposed approach explored the possibility to facilitate the delegation process online. After the first-time delegation of a particular anomaly, our advisee participants are more likely to make their own action when the same anomaly is encountered. We can think of several possible explanations, either originating from some learning effect or fatigue related to technology use.

Delegation with Crowdsourcing. The decisions with crowdsourcing information achieve 1.2x higher accuracy in comparison with the choices without crowdsourcing information (RQ2). Crowdsourcing significantly improved the accuracy of security and privacy decisions for our participants. The improvement in accuracy equally appears in *Advice* responses (1.2x) and *Own Action* responses (1.2x). Researchers have proposed various crowdsourcing models and systems to recommend security and privacy preferences for users [1, 24, 35]. However, the one-size-fits-all solutions could be insufficient to accurately capture users' diverse preference. Furthermore, those proposed crowdsourcing approaches may suffer from data sparsity and the cold-start problem in real-world scenarios. Our proposed approach combines delegation of decisions with crowdsourcing. The combination leverages delegation to prevent the potential issues of crowdsourcing. Meanwhile, the availability of crowdsourcing information encourages advisees to take their own action. Ideally, crowdsourcing would have two positive effects: (1) improving the accuracy of the advice, and (2) reducing the burden on the advisors. However, our findings also imply that crowdsourcing cannot wholly replace social help. Although crowdsourcing improves the advice, the best accuracy is elicited when social support and crowdsourcing worked together. Thus, advisors might be more useful in the cases in which a more nuanced approach to security and privacy is needed, e.g., deciding on an app's permission request.

Learning through Delegation. We were not successful in showing a learning effect in the process of decision delegation (RQ3). Specifically, the participants did not achieve higher response accuracy when they encountered

similar security and privacy anomalies that had never appeared. We found that older adults lack a nuanced understanding of mobile security and privacy, leaving them especially vulnerable to security and privacy violation. Our findings adhere to prior work [20]. The particular concerns and misconceptions of older adults should be addressed through customized training and educational efforts for older adults. While our studies have shown that the behavior of advisees changed over time, we believe that further experiments are needed to establish or rule out learning effect through delegation. Our findings also suggest that assistance technologies need to be designed differently to foster better learning by advisees.

Technology Design in Mobile Security and Privacy. Our main findings highlight the potential, as well as the challenges, of designing support-based technologies in mobile security and privacy. First, our results point to the potential of designing systems that include support built around existing social ties. As our findings show, delegation to family members had improved the performance of the participants, with or without crowdsourcing. While our results were generated in the context of anomaly detection, we believe that these results can be generalized to other fields of security and privacy, as well as to other mobile application domains. Many systems, in various fields, can be improved by adding social support into user interaction that require difficult decisions and monitoring to be carried out. The design of AppMoD points to the way crowdsourcing can be effectively embedded in mobile applications. The use of crowdsourcing for mobile security and privacy is becoming central in mobile applications [7, 24, 33]. However, the way in which crowdsourcing can be effectively leveraged by end-users is not always clear. Our findings show that one impactful use of crowdsourcing was achieved when social support and crowdsourcing worked together. This social-crowdsourcing mechanism can be generalized to various ubiquitous and mobile applications, shaping the way in which crowdsourcing could be used. AppMoD is an example and an exploratory system that uses the social-crowdsourcing mechanism. Via AppMoD, the close social network of the user interprets and monitors the mobile apps, in collaboration with crowdsourcing information, helping the end-users to better understand the apps and make security and privacy decisions.

6.2 Limitations and Future Work

Our work has several limitations that should be taken into account when analyzing its external validity and the impact of our findings.

Our study is based on anomaly detection. While recognizing anomalies is a critical ability, and malware is a severe threat to mobile platforms, the readers should be aware that other privacy and security behaviors might have other delegation characteristics. We rely on a small sample of mostly college students (as advisors), which are probably more technically savvy than the general population. Deploying similar technology in the general population might yield diminishing results. However, we believe that feedback loops can help even less technical advisors learn and get better with time. Our studies occurred in a single country. Future studies should help us understand the effect of culture, family ties, and other factors in delegating security and privacy decisions.

To collect a sufficient amount of responses in a controlled environment, we sent simulated anomaly notifications in our user studies and our participants were informed of the simulated settings. However, in a real life setting, it is possible that advisees may be unwilling to share information with their advisors, especially when working with personal data. Future work can develop privacy preserving mechanisms to facilitate the process of decision delegation.

While this study focused on providing a technical contribution, the way we have designed and tested our technology has deep cultural implications. Different cultures have different ways of interacting with the elderly [27]. In cultures with strong personal growth measures for the elderly, the motivation might be stronger to learn new technologies and to create stronger self-efficacy in computer security and privacy. As the same time, in cultures with closer interpersonal and familial ties, the motivation of advisors to provide support and assistance

can be stronger. Delegation technologies pose interesting and important challenges to elderly care: who are the people that can assist? How can they help?

The analysis of the effect of perceived proficiency is based on self-reported data. Although it is true that self-reported data may suffer from alleged problems, the study [12] reports that there is no strong evidence to conclude that self-reported data is inherently flawed or its use will always impede the ability to interpret correlations. A follow-on study can test the relationship between perceived proficiency and actual proficiency of users in mobile security. In addition, future work could measure whether application familiarity affects users' security and privacy decisions. The familiarity with one app might change how comfortable a user is in making security and privacy decisions due to their experience with that app and the relative time that they have had with it.

Delegating security and privacy decisions opens up new research questions that could be the subject of future studies. Testing and designing AppMoD raises new technical challenges. Users might not even know that they need help at a particular moment because they are not aware of the inner workings of the phone. Therefore, developing technologies to recognize supportable moments can automatically streamline delegation technologies and make it more usable. While our findings reveal some of the opportunities of delegation, it may also carry some risks. For example, delegating decisions might disclose personal and sensitive information to third parties. Furthermore, transferring decisions away from older adults might reduce their autonomy and sense of agency. Theories in fields such as education [32] can help guide the design of technologies that create and transform networks of influence, learning, and support. We observed 5 cases where a participant requested an advice but eventually rejected the suggestion. Future studies can conduct a comparative analysis to answer the question: what is the effect of accepting or rejecting a suggestion?

7 CONCLUSION

In this paper, we propose a community-based security management approach named AppMoD, which allows delegating security and privacy decisions on mobile platforms to trusted social connections. The approach differs from other support-based security and privacy [24, 33], as it is geared towards immediate help loops from close social circles, such as family and friends. AppMoD was implemented to in advocating for using mobile technology as an accessible and usable platform for improving the self-efficacy of older adults, and to foster collective responsibility in their social circles for the technological well-being. While we implemented AppMoD with older adults in mind, but it can generalize to any situation in which help and assistance is required from people with high commitment. Other populations, such as young children, could also benefit from this technology.

ACKNOWLEDGMENTS

This research was supported by the Singapore National Research Foundations National Cybersecurity Research & Development Programme (award number: NRF2016NCR-NCR001-008). We would like to thank Anamika Sawhney and Hwee Lee Tan for their contributions, including in the development of earlier versions of the mobile application used in our study, and in the pilot user study (results are excluded in this paper, but are available in [56]). We would also like to thank Braxton Hall, Anna Maria Eilertsen, Arthur Marques, Jan M. Pilzer, and Giovanni Viviani for their help in proofreading the paper. Most of the work was done when the first author was affiliated with Singapore Management University and supported by the NRF2016NCR-NCR001-008 project. The second author was partially supported by NSFC Program (No. 61902344).

REFERENCES

[1] Yuvraj Agarwal and Malcolm Hall. 2013. ProtectMyPrivacy: detecting and mitigating privacy leaks on iOS devices using crowdsourcing. In *Proceedings of the 11th Annual International Conference on Mobile systems, Applications, and Services*. ACM, 97–110.

[2] Z Aljallad. 2019. Designing a Mobile Application to Support Social Processes for Privacy. In *Proceedings of the NDSS Workshop on Usable Security and Privacy*.
[3] Hussain MJ Almohri, Danfeng Daphne Yao, and Dennis Kafura. 2014. Droidbarrier: Know what is executing on your android. In *Proceedings of the 4th ACM Conference on Data and Application Security and Privacy*. ACM, 257–264.
[4] Monica Anderson and Andrew Perrin. 2017. Tech adoption climbs among older adults. *Pew Research Center* (2017), 1–22.
[5] Daniel Arp, Michael Spreitzenbarth, Malte Hubner, Hugo Gascon, Konrad Rieck, and CERT Siemens. 2014. Drebin: Effective and explainable detection of android malware in your pocket. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, Vol. 14. 23–26.
[6] Vitalii Avdiienko, Konstantin Kuznetsov, Alessandra Gorla, Andreas Zeller, Steven Arzt, Siegfried Rasthofer, and Eric Bodden. 2015. Mining apps for abnormal usage of sensitive data. In *Proceedings of the 37th International Conference on Software Engineering-Volume 1*. IEEE Press, 426–436.
[7] Oshrat Ayalon and Eran Toch. 2018. Crowdsourcing privacy design critique: An empirical evaluation of framing effects. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.
[8] Ron Bitton, Andrey Finkelshtein, Lior Sidi, Rami Puzis, Lior Rokach, and Asaf Shabtai. 2018. Taxonomy of mobile users' security awareness. *Computers & Security* 73 (2018), 266–293.
[9] Tim Broady, Amy Chan, and Peter Caputi. 2010. Comparison of older and younger adults' attitudes towards and abilities with computers: Implications for training and learning. *British Journal of Educational Technology* 41, 3 (2010), 473–485.
[10] Iker Burguera, Urko Zurutuza, and Simin Nadjm-Tehrani. 2011. Crowdroid: behavior-based malware detection system for android. In *Proceedings of the 1st ACM workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM, 15–26.
[11] Daniel Buschek, Sarah Völkel, Clemens Stachl, Lukas Mecke, Sarah Prange, and Ken Pfeuffer. 2018. Experience Sampling as Information Transmission: Perspective and Implications. In *Proceedings of the ACM International Joint Conference and International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. ACM, 606–611.
[12] David Chan. 2009. So why ask me? Are self-report data really that bad. *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organizational and Social Sciences* (2009), 309–336.
[13] Ke Chen and Alan Chan. 2013. Use or non-use of gerontechnology—A qualitative study. *International Journal of Environmental Research and Public Health* 10, 10 (2013), 4645–4666.
[14] Yiwei Chen and Anna Persson. 2002. Internet use among young and older adults: Relation to psychological well-being. *Educational Gerontology* 28, 9 (2002), 731–744.
[15] Michela Cozza, Antonella De Angeli, and Linda Tonolli. 2017. Ubiquitous technologies for older people. *Personal and Ubiquitous Computing* 21, 3 (2017), 607–619.
[16] Leela Damodaran, CW Olphert, and Jatinder Sandhu. 2014. Falling off the bandwagon? Exploring the challenges to sustained digital engagement by older people. *Gerontology* 60, 2 (2014), 163–173.
[17] Sauvik Das, Tiffany Hyun-Jin Kim, Laura A Dabbish, and Jason I Hong. 2014. The effect of social influence on security sensitivity. In *Proceedings of the 10th Symposium On Usable Privacy and Security (SOUPS)*. 143–157.
[18] Ana Correia de Barros, Roxanne Leitão, and Jorge Ribeiro. 2014. Design and evaluation of a mobile user interface for older adults: navigation, interaction and visual design recommendations. *Procedia Computer Science* 27 (2014), 369–378.
[19] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. 2012. Android permissions: User attention, comprehension, and behavior. In *Proceedings of the eighth symposium on usable privacy and security*. ACM, 3.
[20] Alisa Frik, Leysan Nurgalieva, Julia Bernd, Joyce Lee, Florian Schaub, and Serge Egelman. 2019. Privacy and security threat models and mitigation strategies of older adults. In *Proceedings of the 15th Symposium on Usable Privacy and Security (SOUPS)*.
[21] Alessandra Gorla, Ilaria Tavecchia, Florian Gross, and Andreas Zeller. 2014. Checking app behavior against app descriptions. In *Proceedings of the 36th International Conference on Software Engineering*. ACM, 1025–1035.
[22] Dominik Hornung, Claudia Müller, Irina Shklovski, Timo Jakobi, and Volker Wulf. 2017. Navigating Relationships and Boundaries: Concerns Around ICT-uptake for Elderly People. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 7057–7069.
[23] Peter Hornyack, Seungyeop Han, Jaeyeon Jung, Stuart Schechter, and David Wetherall. 2011. These aren't the droids you're looking for: retrofitting android to protect data from imperious applications. In *Proceedings of the 18th ACM Conference on Computer and Communications Security*. ACM, 639–652.
[24] Qatrunnada Ismail, Tousif Ahmed, Apu Kapadia, and Michael K Reiter. 2015. Crowdsourced exploration of security configurations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 467–476.
[25] Jonathan Joe and George Demiris. 2013. Older adults and mobile phones for health: a review. *Journal of Biomedical Informatics* 46, 5 (2013), 947–954.
[26] Jaeyeon Jung, Seungyeop Han, and David Wetherall. 2012. Short paper: enhancing mobile application permissions with runtime feedback and constraints. In *Proceedings of the 2nd ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM, 45–50.

[27] Mayumi Karasawa, Katherine B Curhan, Hazel Rose Markus, Shinobu S Kitayama, Gayle Dienberg Love, Barry T Radler, and Carol D Ryff. 2011. Cultural perspectives on aging and well-being: A comparison of Japan and the United States. *The International Journal of Aging and Human Development* 73, 1 (2011), 73–98.
[28] Lih-Jen Kau and Chih-Sheng Chen. 2015. A smart phone-based pocket fall accident detection, positioning, and rescue system. *IEEE Journal of Biomedical and Health Informatics* 19, 1 (2015), 44–56.
[29] Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. 2012. A conundrum of permissions: installing applications on an android smartphone. In *International Conference on Financial Cryptography and Data Security*. Springer, 68–79.
[30] Matthias Kranz, Andreas MöLler, Nils Hammerla, Stefan Diewald, Thomas PlöTz, Patrick Olivier, and Luis Roalter. 2013. The mobile fitness coach: Towards individualized skill assessment using personalized mobile devices. *Pervasive and Mobile Computing* 9, 2 (2013), 203–215.
[31] Reed Larson and Mihaly Csikszentmihalyi. 2014. The experience sampling method. In *Flow and the Foundations of Positive Psychology*. Springer, 21–34.
[32] Kenneth Leithwood and Doris Jantzi. 2008. Linking leadership to student learning: The contributions of leader efficacy. *Educational Administration Quarterly* 44, 4 (2008), 496–528.
[33] Jialiu Lin, Shahriyar Amini, Jason I Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. 2012. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the ACM conference on Ubiquitous Computing*. ACM, 501–510.
[34] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun Zhang, Norman Sadeh, Alessandro Acquisti, and Yuvraj Agarwal. 2016. Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions. In *Proceedings of the 12th Symposium on Usable Privacy and Security (SOUPS)*.
[35] Bin Liu, Jialiu Lin, and Norman Sadeh. 2014. Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help?. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 201–212.
[36] Wiebke Maaß. 2011. The elderly and the Internet: How senior citizens deal with online privacy. In *Privacy online*. Springer, 235–249.
[37] Tamir Mendel and Eran Toch. 2017. Susceptibility to social influence of privacy behaviors: Peer versus authoritative sources. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 581–593.
[38] Tamir Mendel and Eran Toch. 2019. Social help: developing methods to support older adults in mobile privacy and security. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*.
[39] Tracy L Mitzner, Julie B Boron, Cara Bailey Fausset, Anne E Adams, Neil Charness, Sara J Czaja, Katinka Dijkstra, Arthur D Fisk, Wendy A Rogers, and Joseph Sharit. 2010. Older adults talk technology: Technology usage and attitudes. *Computers in Human Behavior* 26, 6 (2010), 1710–1721.
[40] Alexios Mylonas, Marianthi Theoharidou, and Dimitris Gritzalis. 2013. Assessing privacy risks in android: A user-centric approach. In *Proceedings of the International Workshop on Risk Assessment and Risk-driven Testing*. Springer, 21–37.
[41] Tobias Nef, Raluca L Ganea, René M Müri, and Urs P Mosimann. 2013. Social networking sites and older users–a systematic review. *International Psychogeriatrics* 25, 7 (2013), 1041–1053.
[42] James Nicholson, Lynne Coventry, and Pamela Briggs. 2019. If It's Important It Will Be A Headline: Cybersecurity Information Seeking in Older Adults. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 349.
[43] Kenneth Olmstead and Aaron Smith. 2017. What the public knows about cybersecurity. *Pew Research Center* 22 (2017).
[44] Samantha J Parker, Sonal Jessel, Joshua E Richardson, and M Cary Reid. 2013. Older adults are mobile too! Identifying the barriers and facilitators to older adults' use of mHealth for pain management. *BMC geriatrics* 13, 1 (2013), 43.
[45] Sebastiaan TM Peek, Katrien G Luijkx, Maurice D Rijnaard, Marianne E Nieboer, Claire S van der Voort, Sil Aarts, Joost van Hoof, Hubertus JM Vrijhoef, and Eveline JM Wouters. 2016. Older adults' reasons for using technology while aging in place. *Gerontology* 62, 2 (2016), 226–237.
[46] Bahman Rashidi, Carol Fung, and Tam Vu. 2016. Android fine-grained permission control system with real-time expert recommendations. *Pervasive and Mobile Computing* 32 (2016), 62–77.
[47] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. 2016. How I learned to be secure: a census-representative survey of security advice sources and behavior. In *Proceedings of the ACM SIGSAC conference on Computer and Communications Security*. ACM, 666–677.
[48] Elissa M Redmiles, Amelia R Malone, and Michelle L Mazurek. 2016. I think they're trying to tell me something: Advice sources and selection for digital security. In *Proceedings of the IEEE Symposium on Security and Privacy*. IEEE, 272–288.
[49] Talia Ringer, Dan Grossman, and Franziska Roesner. 2016. Audacious: User-driven access control with unmodified operating systems. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 204–216.
[50] Franziska Roesner and Tadayoshi Kohno. 2013. Securing embedded user interfaces: Android and beyond. In *Proceedings of the 22nd USENIX Conference on Security*. USENIX Association, 97–112.
[51] Franziska Roesner, Tadayoshi Kohno, Alexander Moshchuk, Bryan Parno, Helen J Wang, and Crispin Cowan. 2012. User-driven access control: Rethinking permission granting in modern operating systems. In *Proceedings of the IEEE Symposium on Security and privacy*.

IEEE, 224–238.
[52] Jialiu Lin Bin Liu Norman Sadeh and Jason I Hong. 2014. Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*.
[53] Asaf Shabtai, Uri Kanonov, Yuval Elovici, Chanan Glezer, and Yael Weiss. 2012. "Andromaly": a behavioral malware detection framework for android devices. *Journal of Intelligent Information Systems* 38, 1 (2012), 161–190.
[54] Bilal Shebaro, Oyindamola Oluwatimi, Daniele Midi, and Elisa Bertino. 2014. IdentiDroid: Android can finally Wear its Anonymous Suit. *Transactions on Data Privacy* 7, 1 (2014), 27–50.
[55] Deógenes P Silva Junior, Patricia Cristiane de Souza, and Thaíres A de Jesus Gonçalves. 2018. Early Privacy: Approximating Mental Models in the Definition of Privacy Requirements in Systems Design. In *Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems*. ACM, 19.
[56] Hwee Lee Tan. 2018. Safety and Privacy of Smart-City Mobile Applications through Model Inference. *MAIS Capstone Project Reports (Main, User Study, System Documentation)* (2018).
[57] Wei Tang, Guang Jin, Jiaming He, and Xianliang Jiang. 2011. Extending Android security enforcement with a security distance model. In *Proceedings of the International Conference on Internet Technology and Applications*. IEEE, 1–4.
[58] Kyle Taylor and Laura Silver. 2019. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. *Pew Research Center (FEBRUARY 5Th 2019)* (2019), 1–46.
[59] Ilaria Torre, Odnan Ref Sanchez, Frosina Koceva, and Giovanni Adorni. 2018. Supporting users to take informed decisions on privacy settings of personal devices. *Personal and Ubiquitous Computing* 22, 2 (2018), 345–364.
[60] Lynn Tsai, Primal Wijesekera, Joel Reardon, Irwin Reyes, Serge Egelman, David Wagner, Nathan Good, and Jung-Wei Chen. 2017. Turtle Guard: Helping Android Users Apply Contextual Privacy Preferences. In *Proceedings of the 13th Symposium on Usable Privacy and Security (SOUPS)*.
[61] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys (CSUR)* 50, 6, Article 93 (Dec. 2017), 40 pages.
[62] Yang Wang. 2017. The third wave?: Inclusive privacy and security. In *Proceedings of the New Security Paradigms Workshop*. ACM, 122–130.
[63] Primal Wijesekera, Arjun Baokar, Ashkan Hosseini, Serge Egelman, David Wagner, and Konstantin Beznosov. 2015. Android Permissions Remystified: A Field Study on Contextual Integrity.. In *Proceedings of the USENIX Security Symposium*. 499–514.
[64] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. 2017. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *Proceedings of the 2017 IEEE Symposium on Security and Privacy*. IEEE, 1077–1093.
[65] Primal Wijesekera, Joel Reardon, Irwin Reyes, Lynn Tsai, Jung-Wei Chen, Nathan Good, David Wagner, Konstantin Beznosov, and Serge Egelman. 2018. Contextualizing Privacy Decisions for Better Prediction (and Protection). In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 268.
[66] Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics bulletin* 1, 6 (1945), 80–83.
[67] Bo Xie, Ivan Watkins, Jen Golbeck, and Man Huang. 2012. Understanding and changing older adults' perceptions and learning of social media. *Educational Gerontology* 38, 4 (2012), 282–296.
[68] Eva-Maria Zeissig, Chantal Lidynia, Luisa Vervier, Andera Gadeib, and Martina Ziefle. 2017. Online privacy perceptions of older adults. In *Proceedings of the International Conference on Human Aspects of IT for the Aged Population*. Springer, 181–200.
[69] Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen. 2014. Mobile app recommendations with security and privacy awareness. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 951–960.

A ANOMALIES IN USER STUDIES

Table 2. Overview of anomalies in user studies.

	Category	App	Installs	Anomaly	Resource	Difficulty Level	True Positive?
A1-1	Communication	Gmail	1,000M+	Gmail is modifying the calendar events of your device.	Calendar	Medium	No
A1-2	Communication	Gmail	1,000M+	Gmail is sending emails to everyone in your address book.	Contacts	Hard	Yes
A1-3	Communication	Whatsapp	1,000M+	Whatsapp is accessing the contacts of your device.	Contacts	Medium	No
A1-4	Communication	Whatsapp	1,000M+	Whatsapp is making phone calls to 61234567 (unknown).	Calls	Medium	Yes
A1-5	Games/Board	Sudoku	10M+	Sudoku is reading your sim card info and sending it to www.abnormal.com.	Phone Status	Easy	Yes
A1-6	Games/Casual	Candy Crush	500M+	Candy Crush is reading your contacts and sending it to www.hackme.com.	Contacts	Easy	Yes
A1-7	Social	Facebook	1,000M+	Facebook is accessing the location of your device.	Location	Hard	No
A1-8	Social	Instagram	1,000M+	Instagram is accessing the camera of your device.	Camera	Medium	No
A1-9	Tools	Clock	100M+	Clock is accessing your geolocation and sending it out.	Location	Hard	Yes
A1-10	Video Players & Editors	YouTube	1,000M+	YouTube is accessing the microphone of your device.	Microphone	Medium	No
A2-1	Communication	Microsoft Outlook	100M+	Microsoft Outlook is modifying the calendar events of your device.	Calendar	Medium	No
A2-2	Communication	Microsoft Outlook	100M+	Microsoft Outlook is sending emails to everyone in your address book.	Contacts	Hard	Yes
A2-3	Communication	Line	10M+	Line is accessing the contacts of your device.	Contacts	Medium	No
A2-4	Communication	Line	10M+	Line is making phone calls to 61234567 (unknown).	Calls	Medium	Yes
A2-5	Games/Board	Mahjong Titans	500K+	Mahjong Titans is reading your sim card info and sending it to www.abnormal.com.	Phone Status	Easy	Yes
A2-6	Games/Casual	Bubble Shooter	50M+	Bubble Shooter is reading your contacts and sending it to www.hackme.com.	Contacts	Easy	Yes
A2-7	Social	Linkedin	100M+	Linkedin is accessing the location of your device.	Location	Hard	No
A2-8	Social	Tumblr	100M+	Tumblr is accessing the camera of your device.	Camera	Medium	No
A2-9	Tools	Compass Galaxy	1M+	Compass Galaxy is accessing your geolocation and sending it out.	Location	Hard	Yes
A2-10	Video Players & Editors	VLC	100M+	VLC is accessing the microphone of your device.	Microphone	Medium	No
A3-1	Communication	Messenger	1,000M+	Messenger is reading calendar events on your device.	Calendar	Medium	No
A3-2	Communication	Messenger	1,000M+	Messenger is sending SMS messages to everyone in your address book.	Contacts	Hard	Yes
A3-3	Communication	Skype	1,000M+	Skype is accessing the contacts of your device.	Contacts	Medium	No
A3-4	Communication	Skype	1,000M+	Skype is making phone calls to 97752222 (unknown).	Calls	Medium	Yes
A3-5	Games/Board	Happy Color	10M+	Happy Color is reading your sim card info and sending it to www.hackme.com.	Phone Status	Easy	Yes
A3-6	Games/Casual	My Talking Tom	500M+	My Talking Tom is accessing the contacts of your device and sending it to www.abnormal.com.	Contacts	Easy	Yes
A3-7	Social	Google+	1,000M+	Google+ is accessing your geolocation of your device.	Location	Hard	No
A3-8	Social	Google+	1,000M+	Google+ is accessing the camera of your device.	Camera	Medium	No
A3-9	Tools	AVG AntiVirus	100M+	AVG AntiVirus is accessing the location of your device and sending it to www.abnormal.com.	Location	Hard	Yes
A3-10	Video Players & Editors	Google Play Movies & TV	1,000M+	Google Play Movies & TV is accessing the microphone of your device.	Microphone	Medium	No
A4-1	Communication	Messenger	1,000M+	Messenger is reading contacts of your device and sending it out.	Contacts	Medium	Yes
A4-2	Communication	Messenger	1,000M+	Messenger is reading text messages of your device.	SMS	Hard	No
A4-3	Communication	Skype	1,000M+	Skype is sending SMS messages to 97752222 (unknown).	SMS	Medium	Yes
A4-4	Communication	Skype	1,000M+	Skype is accessing the location of your device.	Location	Medium	No
A4-5	Games/Board	Happy Color	10M+	Happy Color is viewing network connections of your device.	Network	Easy	No
A4-6	Games/Casual	My Talking Tom	500M+	My Talking Tom is accessing the microphone of your device.	Microphone	Easy	No
A4-7	Social	Google+	1,000M+	Google+ is accessing the contacts of your device and sending it to www.abnormal.com.	Contacts	Hard	Yes
A4-8	Social	Google+	1,000M+	Google+ is reading your sim card info and sending it to www.hackme.com.	Phone Status	Medium	Yes
A4-9	Tools	AVG AntiVirus	100M+	AVG AntiVirus is deleting the contents of the storage on your device.	Storage	Hard	No
A4-10	Video Players & Editors	Google Play Movies & TV	1,000M+	Google Play Movies & TV is reading your sim card info and sending it to www.hackme.com.	Phone Status	Medium	Yes