Should I Answer? Measuring User Responses To Anti-Robocall Application Indicators

Imani N. Sherman
Jasmine D. Bowers
Keith McNamara Jr.
Juan E. Gilbert
Patrick Traynor

University of Florida
Gainesville, FL 32608, USA

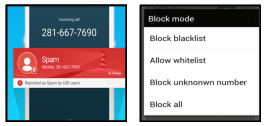
shermani@ufl.edu
jdbowers@ufl.edu
kmcnamara1@ufl.edu
juan@ufl.edu
traynor@cise.ufl.edu

Abstract
Spam calls, like robocalls, can create communication hardship, and irritability for everyday telephone users. Third-party applications have been developed to reduce and block spam calls by using lists to detect spam and warnings to alert users when they receive a spam call. Researchers have investigated the use of lists for detecting spam calls, but there is a lack of research on the effectiveness of spam call warnings. This research investigates user spam call management needs and evaluates the effectiveness of various warnings based on user response. We analyzed 10 popular spam call management applications to identify the current warning design trend, held focus groups to understand user experience with spam calls, and conducted user studies to evaluate the effectiveness of various designs. The results of our user studies provide evidence to warn users of both legitimate and malicious calls.

Author Keywords
spam, robocalls, mobile, security

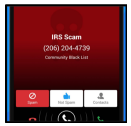
ACM Classification Keywords
H.5.2 [User Interfaces]: User Centered Design;

Introduction
Telephone users are often tasked with determining if an incoming call is safe to answer. Since spam is believed to

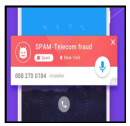


(a) CallApp: Caller ID (A1)

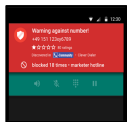
(b) Call Blocker (A2)



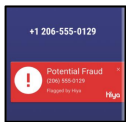
(c) Call Control (A3)



(d) Caller ID and Call Blocker - DU Caller (A4)

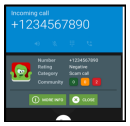


(e) Clever Dialer (A5)

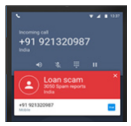


(f) Hiya (A6)

(g) Mr.Number (A7)



(h) Should I Answer? (A8)



(i) Truecaller (A9)

(j) Who's Calling (A10)

Figure 1: The warning designs used for each app.

account for 50% of all calls in 2019, this will likely turn into a daily task [12]. Deregulation, telephony infrastructure and lack of end-to-end authentication make it easy and inexpensive to send spam calls at a large scale. The Federal Communications Commission (FCC) and Federal Trade Commission (FTC) have implemented and enforced rules regulating the use of autodialers, Caller ID spoofing, and spam calling in general [6]. However, the number of complaints they receive has increased every year. This is partly because, even though they are " urging phone companies to implement Caller ID authentication", Caller ID remains unauthenticated [6].

The global telephony infrastructure includes cellular networks, Voice Over Internet Protocol (VOIP), and the Public Switched Telephone Network (PSTN). These networks are connected via gateways, which allow calls made in one network to reach endpoints in another. Each technology generates its own associated metadata; however, we cannot guarantee that any of this data can be delivered end-to-end except voice and Caller ID, neither of which is authenticated. Caller ID authentication is crucial. VOIP provides a cheap way to make calls, and in some VOIP systems an easy way to change the Caller ID information that is sent to the callee. Those VOIP systems allow robocallers and other malicious entities to appear as a familiar or trusted contact, and increase the number of calls sent.

The increase in spam calls has inspired solutions from heuristics to cryptography. Using Caller ID (assuming no spoofing), black or whitelisting, call back verification [11], content and audio analysis [3, 10], provider-based solutions (e.g., SHAKEN/STIR [9]), end-to-end solutions (e.g., AuthentiCall [14, 15]), and mobile applications that implement some of these solutions have been suggested to combat the problem. This work aims to identify the popular indicators in the top anti-spam call applications, user experiences with spam calls and their indicator preferences, and results of warning design user testing.

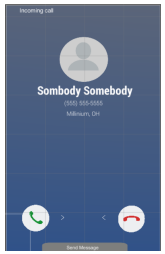
Related Work

"The ultimate criterion of warning effectiveness is, of course, whether the warning actually modifies human behavior [13]". Researchers have investigated the effectiveness of SSL warnings [17, 7, 1], software download warnings [5], warning fatigue [4], indicators [8], browser warnings [16], and malware warnings [2].To our knowledge, there has been no publicly available study on the effectiveness of such indicators for spam call warnings. This research attempts to address this issue.

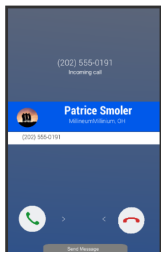
Survey of Anti-Robocall Applications

We begin with a study of the top 10 state-of-the-art anti-robocall applications for Android shown in Figure 1. These applications were selected using the Google Play store. The first 10 applications that 1) appeared as a search result for *spam call blocker* in October 2018, 2) were free to download, 3) had an average rating of at least 4 stars, 4) had at least one million downloads, and 5) were not owned by a telephone carrier were selected. Each application's Google Play Store page, website, and privacy policy were reviewed to determine how they detect spam calls and the warnings used to notify users of spam calls.

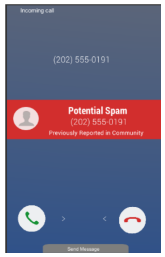
Call Detection: All of the apps use blacklists to identify spam. A2 follows the list created by the user and only references that list. A1 does not disclose their source, but uses data from sixty sources. A3 uses data from the FCC, FTC, IRS, State of Indiana, and their community of users. The remaining apps do not state where the information used in their global database is from, but they do include information from the reported spam calls within their community of



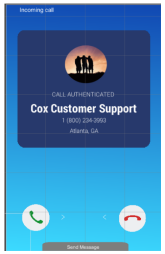
(a) Control



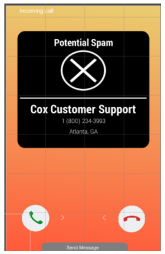
(b) Pop-CID



(c) Pop-Spam



(d) BSky-Auth



(e) BSky-Spam

Figure 2: The five designs used in the interactive survey.

users.

Warning Design: All of the apps, except A2 and A8, meet Wogalter's warning design guidelines [18]. Salience and personal factors are a part of the guidelines but were not considered because they require feedback from users. We instead focused on wording, layout, placement, and pictorial symbols. The majority of the apps used the word *Spam*, placed information where users could easily see them, and implemented indicators commonly used to convey danger. A2 does not actually warn users of danger, and therefore does not meet the requirements. A8 does warn users, but presents cluttered information, which can make it difficult for users to detect the warning.

Focus Groups

After identifying the methods anti-robocall applications use to detect and notify users of malicious calls, we then conducted focus groups to get the users' perspectives on spam calls. We use the phrase spam calls to include all types of malicious calls, including robocalls. The goal of this work was to collect user experiences with caller ID and identify the visual cues they desired.

Methodology. This study was approved by the Internal Review Board at the University of Florida. Eighteen subjects were recruited using an online research administration system. Participants were between the age of 21 and 28 years of age and most (72%) were male. The participants identified themselves as being a part of five ethnic groups: African American (11%), Caucasian (22%), Latinx/Hispanic (28%), South Asian (17%), and East Asian (22%). The participants reviewed and signed the consent, completed a demographic survey, and answered questions about their experiences and preferences related to spam calls.

Results. Participants were asked to recall how they detect spam calls and the notifications they like to receive when a call is malicious. All of the participants mentioned they looked at Caller ID, area code, the time, and reviewed their personal call expectations to detect malicious calls. They use Caller ID and area code to determine if the call is from a known number or area. They refer to their personal call expectations and time to determine if they are expecting a call from an unknown number during that time in their life. As many participants mentioned, they are more likely to answer a call from an unknown number if actively on the job market. But even with these techniques, participants recalled experiences when their techniques failed. As a follow-up, we asked participants to tell us how they would like to be warned about spam calls. The participants agreed that the background color, icons and Caller ID were important aspects of a call notification. They desired a notification with a background color that covered the entire screen, a check mark or "X" mark as icons, and some form of Caller ID validation.

User Testing

Methodology. This study was approved by the Internal Review Board at the University of Florida. Thirty-four subjects were recruited using an online research administration system. The majority of participants (62%) were between the age of 20 and 25, while 29% were between 26 and 30, 9% were over the age of 30, half (50%) were female and 50% male. The racial and ethnic background of participants included East Asian(15%), Caucasian (26%), African American (18%), South Asian (26%), Latinx/Hispanic (6%), Middle Eastern (6%), and Caribbean (3%). There was no overlap in participants between the focus groups and this study. When participants met with the research team at their chosen time slot, they reviewed and signed the consent form, completed a demographic survey, an interactive survey, and were then debriefed about the purpose of the study. We de-

Scene Design (A)	Scene Design (B)	(A-B)*	*p*
C	B-Auth	-5%	.08
	P-Spam	1%	.901
B-Auth	P-CID	6%	.0049
B-Spam	P-Spam	12%	<.0001

* Difference in % of Calls Accepted

Table 1: Comparison of Notice Impact on Response - Tukey Test Results

Scene Design (A)	Scene Design (B)	(A-B)*
C	B-Auth	-15%
	P-CID	-1%

*** Difference in % of participants that accepted calls from unknown numbers**

Table 2: Comparison of the Percent of Participants that Accepts Calls from Unknown Numbers when an Authenticated Call Notice was Present - Tukey Test Results

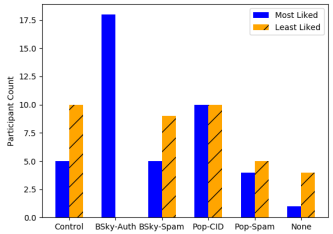


Figure 3: The results from participants picking the design they liked the most and the one they liked the least.

veloped a mobile application to 1) display mock phone calls (screenshots) used for the interactive survey and 2) capture the participants' responses to the mock calls. Five warning designs were used in the survey and can be seen in Figure 2. The Control design (Control or C) imitates what an incoming call looks like on a Samsung Galaxy S7 without a warning. The Popular design (Pop or P) was inspired by the reoccurring design elements used in anti-robocall apps found in the survey of anti-robocall apps. The Blue Sky design (BSky or BY) was inspired by the design elements requested by the participants in the focus group. Each design was shown with an incoming call from six unique numbers, two of which came from the participants saved contacts.

The app analysis results show that a number of anti-robocall applications follow Wogalter's warning design guidelines and use a similar warning layout. The focus group results suggest that users desire a warning with easy to interpret icons, a noticeable background color that fills the entire screen, and trusted Caller ID information. This study evaluates the effectiveness of spam call warning design elements in a best case scenario environment.

Results. In general, participants were more likely to answer authenticated calls and calls from known numbers. They did not spend significantly more time reacting to calls with warnings compared to calls from those same numbers without a warning. The presence of an authenticated call notice did increase the number of accepted calls. In addition, the number of participants that answered authenticated calls from unknown numbers increased by 15%, when compared to the control design. The presence of a spam calls warning significantly decreased the number of answered calls from known numbers when compared to calls from those same numbers without a warning. Participants answered more calls when shown BSky-Auth (61%) than Pop-CID (55%).

Finally, participants were more likely to answer spam calls from known numbers when BSky-Spam (25%) was shown compared to Pop-Spam (13%), where p is less than .0001. The results are due to the presence of Caller ID and the color scheme used in each design. Participants liked the color red, blue, the authenticated call label and the Pop-spam note which told participants why the call was labeled as spam.

Summary

Cell phone users are interrupted by robocalls daily. Telephone providers and third-party organizations have developed applications to solve this problem by detecting and blocking robocalls. We reviewed the top ten anti-robocall apps and found that 1) they all use blacklists to detect robocalls and 2) the majority of spam call warnings used in these apps placed a red bar in the middle of the screen. We then held focus groups which found that all of our participants 1) relied on Caller ID, and 2) desired a spam call warning that uses a check mark and prohibition sign, along with an alerting background color that fills the entire screen. We applied these design elements to the Popular and Blue Sky design, respectively, and compared their effect on users to each other and to the Control design which had no warning. As indicated in similar studies [16, 7, 1], warning design can affect user decision making. The BSky-Auth and Pop-CID increased the number of calls that were answered through the use of *Authenticated Call* label. Anti-robocall apps available today do not determine if Caller ID information is valid. However, the results of this study suggest that users want that capability and would go so far as to trust the notice, answering calls they would usually ignore.

REFERENCES

1. Devdatta Akhawe and Adrienne Porter Felt. 2013. Alice in Warningland: A Large-Scale Field Study of Browser

Security Warning Effectiveness.. In *USENIX security symposium*, Vol. 13.

2. Hazim Almuhimedi, Adrienne Porter Felt, Robert W Reeder, and Sunny Consolvo. 2014. Your reputation precedes you: History, reputation, and the chrome malware warning. In *Symposium on Usable Privacy and Security (SOUPS)*, Vol. 4. 2.

3. Vijay A Balasubramaniyan, Aamir Poonawalla, Mustaque Ahamad, Michael T Hunter, and Patrick Traynor. 2010. PinDr0p: using single-ended audio features to determine call provenance. In *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 109–120.

4. Cristian Bravo-Lillo, Lorrie Cranor, Saranga Komanduri, Stuart Schechter, and Manya Sleeper. 2014. Harder to ignore. *Revisiting pop-up fatigue and approaches to prevent it, USENIX Association* (2014), 105–111.

5. Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. 2013. Your attention please: designing security-decision UIs to make genuine risks harder to ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*. ACM, 6.

6. Federal Communications Commission. 2018. Stop Unwanted Robocalls and Texts. (2018). `https://www.fcc.gov/consumers/guides/stop-unwanted-robocalls-and-texts`

7. Adrienne Porter Felt, Alex Ainslie, Robert W Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL warnings: Comprehension and adherence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2893–2902.

8. Adrienne Porter Felt, Robert W Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson, Mustafa Embre Acer, Elisabeth Morant, and Sunny Consolvo. 2016. Rethinking Connection Security Indicators. In *SOUPS*. 1–14.

9. Alliance for Telecommunications Industry Solutions. 2017. Signature-based Handling of Asserted information using toKENs (SHAKEN): Governance Model and Certificate Management. (2017). `http://www.atis.org/sti-ga/resources/docs/ATIS-1000080.pdf`

10. Federico Maggi. 2010. Are the con artists back? a preliminary analysis of modern phone frauds. In *2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010)*. IEEE, 824–831.

11. H. Mustafa, W. Xu, A. R. Sadeghi, and S. Schulz. 2014. You Can Call but You Can't Hide: Detecting Caller ID Spoofing Attacks. In *2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*. 168–179. `DOI: http://dx.doi.org/10.1109/DSN.2014.102`

12. First Orion. 2018. Nearly 50% Of U.S. Mobile Traffic Will Be Scam Calls By 2019. (2018). `https://firstorion.com/nearly-50-of-u-s-mobile-traffic-will-be-scam-calls-by-2019/`

13. George A Peters. 1984. A challenge to the safety profession. *Professional Safety* 29, 10 (1984), 46–50.

14. Bradley Reaves, Logan Blue, Hadi Abdullah, Luis Vargas, Patrick Traynor, and Thomas Shrimpton. 2017. AuthentiCall: Efficient Identity and Content Authentication for Phone Calls. In *26th USENIX*

Security Symposium (USENIX Security 17). USENIX Association, Vancouver, BC, 575–592. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/reaves

15. Bradley Reaves, Logan Blue, and Patrick Traynor. 2016. AuthLoop: End-to-End Cryptographic Authentication for Telephony over Voice Channels. In *Proceedings of the USENIX Security Symposium (SECURITY)*. (Acceptance Rate: 15.5%).

16. Robert W Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. 2018. An Experience Sampling Study of User Reactions to Browser Warnings in the Field. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 512.

17. J Sunshine, S Egelman, H Almuhimedi, N Atri, and LF Cranor. 2009. USENIX security symposium. *Crying wolf: an empirical study of SSL warning effectiveness* (2009), 399–416.

18. Michael S Wogalter, Vincent C Conzola, and Tonya L Smith-Jackson. 2002. based guidelines for warning design and evaluation. *Applied ergonomics* 33, 3 (2002), 219–230.