July 16, 2018

Crowd-Validated CAPTCHAs and Content Verification

Myra Wong Liu

Kathryn Bush

Crowd-validated CAPTCHAs and content verification

ABSTRACT

Social media, rating systems, trending news, and other online information sources rely on bona fide human contributors in order to function accurately and retain trust amongst information consumers. Inaccurate or false content, quickly and widely disseminated by bots, can cause erosion of trust in these platforms and have wider ramifications.

This disclosure describes techniques to generate CAPTCHA challenges that are difficult, and nearly impossible, for bots to solve. The challenges comprise gestural, emotive, or cognitive micro-tasks that involve physical interaction of the challenge-taker with multiple UI modes such as camera, touchscreen, etc.

KEYWORDS

- CAPTCHA
- Crowd-validated content
- Turing test
- Content validation
- Trusted content
- Cognitive captcha
- Gestural captcha
- Emotive captcha

BACKGROUND

Online information sources, e.g., social media posts, ratings systems, review/ recommendation content, as well as content based on such sources, e.g., trending news, search rankings, etc., rely on bona fide human contributors to function accurately and retain trust

amongst information consumers. Inaccurate, false, or fake content disseminated by bots can cause erosion of trust and can have wider ramifications. At present, bots are available (possibly released by mala fide actors) that can skew the relative popularity of social media posts or comments, or even generate false content such as news and other information. Such bots, if left unchecked, can lead to a shift the information consumer's perception of truth and result in serious damage to institutional trust in the online information source.

DESCRIPTION

This disclosure describes techniques to generate CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) challenges that are comprehended and solved nearly effortlessly by humans, while being difficult, and nearly impossible, for bots to solve. The challenges comprise gestural, emotive, or cognitive micro-tasks involving physical interaction of the challenge-taker with multiple UI modes such as camera, touchscreen, etc. The captchas are illustrated by examples.

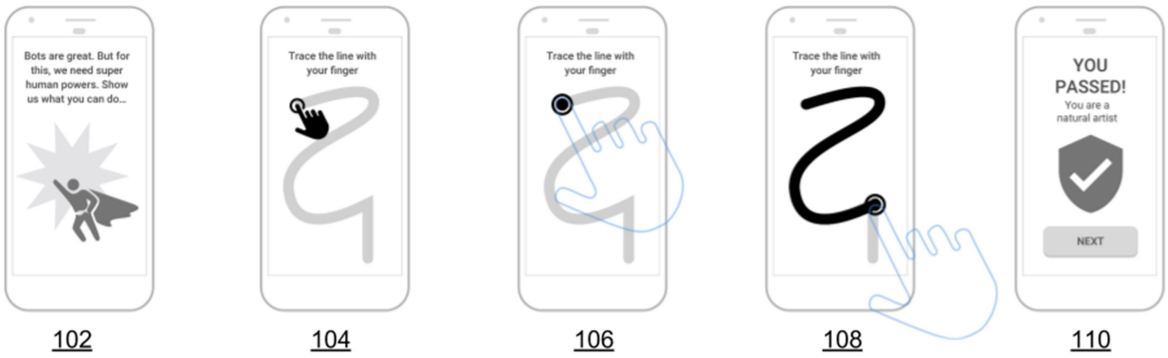

Fig. 1: CAPTCHA comprising gestural micro-tasks

Example 1: Fig. 1 illustrates an example of a captcha comprising a gestural micro-task. After an introductory screen (102) requesting the user to respond to a challenge, a curved line is displayed on the screen (104), the challenge being that the user trace their finger over the

curved line. This micro-task requires physical interaction of the user with the touchscreen. It can change with each invocation, and can vary in challenge level from simple curves to intricate ones. It is easy for most humans to comprehend and solve, yet difficult for a bot to simulate an answer. When a bona fide human traces finger over the displayed curve (106, 108), and in this example, is determined as having passed the captcha challenge (110).

Fig. 2: CAPTCHA comprising emotive micro-tasks

Example 2: Fig. 2 illustrates an example of a captcha comprising emotive micro-tasks. The captcha challenges the user to make respectively an angry face (202), a surprised face (204), and a happy face (206). This micro-task requires interaction of the user with the camera of the mobile device. It can change, e.g., present different emotive or facial-expression challenges, with each invocation. It is easy for most humans to comprehend and solve, yet hard for a bot to simulate a correct response. In the case of emotive, countenance-based, or biometric-based challenges, a user is provided with options to reject such challenges and request in lieu captchas that are not based on facial or biometric capture.

Fig. 3: CAPTCHA comprising cognitive micro-tasks

Example 3: Fig. 3 illustrates an example of a captcha comprising cognitive micro-tasks. The captcha challenges the user to distinguish between objects of differing beauty (302), between instruments that produce sounds of differing euphony (304), between objects of differing edibility (306), etc. Such micro-tasks have a large degree of variability across invocations. They are easy for most humans to comprehend and solve, yet hard for a bot to simulate a correct response.

In this manner, captchas described herein provide at least the following features.

- They comprise dynamically-generated gestural, emotive, or cognitive micro-tasks.
- They include time-boxed input requirements, e.g., real-time interaction.
- Physical human interaction with the user interface is an integral component of the challenges, e.g., different device sensors (touchscreen, camera, etc.) are utilized in issuing a captcha challenge and assessing a response.
- They have multi-device support, e.g., can work with devices that have touchscreen capabilities and/or on-device camera.

- They can be changed with every invocation, e.g., a unique design to trace, a new facial-expression challenge, etc.
- They are hard to learn or game by bots.

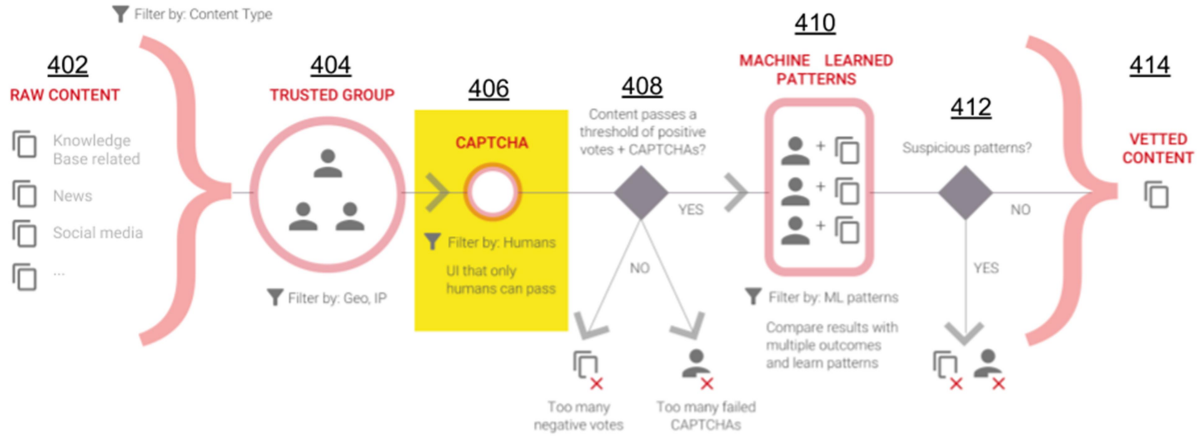

Fig. 4: Chain of procedures to separate vetted content from raw content

Gestural, emotive, or cognitive captchas, as described herein, are part of a larger chain of crowd and machine-validated procedures that transform raw, possibly false, content to vetted content. Fig. 4 illustrates an example of a chain that separates vetted content out of raw content. Raw content (402) is filtered by a trusted group of humans (404). The content is further filtered by passage over a threshold of positive votes and captcha challenges (406).

Too many failed captchas on a given content, or too many negative votes, render the content as possibly false (408). The content is filtered with the use of machine-learning models (410) to identify trustworthy and untrustworthy sources, e.g., by detecting patterns similar to other bot-authored or disseminated content, network addresses that consistently generate dubious content, user identifiers associate with a history of mala fide online behavior, etc. Content that is deemed suspicious, as evaluated by humans or machine-learned models (or both), is rejected (412) to arrive at vetted content (414). Captchas described within this

disclosure can be used throughout, e.g., to verify that the creators or disseminators of raw content are humans, to verify that members of the trusted group are humans, etc.

In this manner, the techniques of this disclosure enable the creation of a crowd-sourced machine learning system that combines human intuition with robust pattern recognition capabilities of computers to filter out false or inaccurate content.

By enabling the verification of users as humans, online information portals, e.g., content providers, news aggregators, e-commerce sites, social media, etc., are empowered to identify and mark or eliminate false stories, inaccurate content, fake news, purchase-and-scalping of tickets by bots, click fraud, false votes, false likes, etc. The possibility of influence, interference, or overpowering of the content portal by bot swarms is thereby foreclosed. Major online tools and information sources such as search rankings, news, trends, comments, ratings, reviews, etc. can thus be protected. User trust in online information sources is thereby preserved and cultivated.

Alternately, online information sources can have dedicated internal human resources to vet content. Such an approach may not be as scalable as the techniques described herein, and are also subject to unconscious biases.

The gestural, emotive, or cognitive captchas are implemented to utilize user-permitted data, e.g., touchscreen input, detection of facial or other features, biometrics, etc. Users can choose to restrict access to data that is used for such captchas. Based on user-permitted factors, one or more different types of captchas may be presented to different users.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network,

social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques to generate CAPTCHA challenges that are difficult, and nearly impossible, for bots to solve. The challenges comprise gestural, emotive, or cognitive micro-tasks that involve physical interaction of the challenge-taker with multiple UI modes such as camera, touchscreen, etc. The challenges can be deployed, for example, to verify that content provided in online platforms is from human users, not bots.