[mCaptcha] Open Technology Fund | Internet Freedom Fund

Project Title*

mCaptcha

Describe your project in 1-3 sentences.

Existing CAPTCHA systems track their users across the internet using cookies and IP logging, in addition to human abilities to identify objects better than bots to prevent spam. They use opaque, non-idempotent "prediction" mechanisms which can be taken advantage of to deny access to specific groups trying to access websites.

In free worlds, transparency is critical in guarding access to information. mCaptcha's PoW mechanism is transparent and idempotent, making validation a verifiable.

Existing CAPTCHAs are inaccurate in Tor, it will deny access rely or require multiple challenges to pass. mCaptcha's IP independence provides same quality UX, even in Tor.

What problem will your project address?

The existing CAPTCHA systems like reCAPTCHA and hCaptcha use opaque decision-making mechanisms to validate spam. The decisions made by these systems are not reproducible and not verifiable as they are non-idempotent and opaque. So it is impossible to guarantee that these systems are non-discriminating. Censors can take advantage of the opaqueness of these systems to block certain groups from accessing certain types of information.

mCaptcha's Proof of Work mechanism uses strong cryptographic principles that guarantee idempotency and transparency. Censors can't use mCaptcha to deny access to information without evading detection. With mCaptcha, successful validation and validation failure and failure are absolute states and are fully verifiable outcomes.

Also, existing CAPTCHAs depend on humans' superior abilities to recognize objects. They use distorted photos of things or text and require the users to identify them successfully to pass the challenge. This nature of the existing CAPTCHA imposes grave difficulties on people with cognitive, visual, and auditory impairments. mCaptcha's Proof of Work mechanism is idempotent, fully automated, and highly accessible.

Furthermore, the ubiquitous nature of reCAPTCHA and hCaptcha makes them a tempting target for nation-states to compromise these systems. A security breach in these systems will be enough to expose the internet usage history of all the users that have ever come in contact with them. mCaptcha is Free Software and is self-hostable. Even if a mCaptcha instance that provides CAPTCHA service to multiple third-party services is compromised, it will not have the same effect as it doesn't log user activity and IP addresses.

reCAPTCHA and hCaptcha are generally bad for internet privacy, especially for users living in countries with oppressive governments. Tor is the safest way to access the internet when living in oppressive regimes. The IP addresses used by VPNs and Tor are dirty because hundreds if not thousands of users' traffic is proxied through them simultaneously. Current CAPTCHA systems rely on IP-based tracking to gauge and determine the possibility of a request being spam, which makes VPN and Tor traffic resemble bot activity. So this makes accessing websites from anonymizing technologies like VPN hard and Tor impossible.

mCaptcha's PoW mechanism doesn't log IP addresses and doesn't require tracking user activating across the internet. This aspect of mCatpcha allows it to combat spam effectively and accurately, even in anonymizing networks like VPNs and Tor.

If this project is funded, what form will it take?

Technology Development

Give a brief overview of the activities in this project.

Objective 1: Proof of Work accessibility: PoW within mCaptcha is configurable to take anywhere between 200ms to +10s to compute. But the time will vary across devices, as some might have more computing power than others. Older devices might be too slow to generate PoW and fail validation.

I'm working on a survey to benchmark PoW on various devices, which will shed light on optimal PoW difficulty configuration. The survey results will help mCaptcha admins pick a difficulty factor that works for most visitors and improve PoW accessibility.

Activity 1.1: Run WASM and JavaScript polyfill benchmarks Activity 1.2: Collect statistics Activity 1.3: Periodically publish results and determine 90th percentile and 99th percentile average

Objective 2: Horizontal scaling to handle websites with large visitors: mCaptcha implements actor model in which each website is an actor. While the current system can scale horizontally to accommodate more websites, it doesn't scale horizontally to accommodate more visitors on a single website. If popular services are to switch to mCaptcha, single website horizontal scalability should be implemented.

Activity 2.1: Create distributed cache that implements a leaky bucket algorithm Activity 2.2: Verify correctness of the implementation Activity 2.3: Create multi-node DDoS attack simulation for benchmark(mCaptcha/dos is WIP, needs Infrastructure as Code to deploy multiple attacking nodes) Activity 2.4: Benchmark and compare performance against the semi-distributed, Redis-based cache implementation that is currently used

Objective 3: Integration tests: While there is extensive unit testing(+90% code coverage on all repositories), there are no integration tests that span all repositories.

Activity 3.1: Create Selenium-based integration testing with the following configuration options: 1. PoW options: i. WASM library ii. JavaScript polyfill library 2. Database options: i. Postgres ii. MariaDB 3. Cache options: i. embedded ii. Redis-based(uses mCaptcha/cache, a custom Redis-module) iii. Custom distributed cache 4. On browsers: i. Firefox ii. Chromium

Are there similar projects that exist already? How is your project different or complementary to those projects?

- Friendly captcha: PoW-based CAPTCHA but non-free. A libre, self-hostable version is available but is stripped down and doesn't have critical features like variable difficulty scaling(which improves UX). Little to no documentation on how the CAPTCHA system works.
- pow-captcha: PoW based CAPTCHA libre implementation. Doesn't support variable difficulty scaling. Uses Scrypt against SHA256, unsuitable for large-scale deployments as Scrypt is more resource intensive. Integration libraries and client libraries don't exist for popular stacks yet.

mCaptcha is fully libre and aims to be usable by even the most popular web services.

How long do you estimate this project will take?

6 months to 1 year

How much funding do you estimate you will need? (In US Dollars)

$26970

Who would benefit from this project?

People with visual, auditory, and cognitive impairments will benefit from the superior accessibility offered by mCaptcha. Also, people dependent on anonymizing technologies like VPNs and Tor to access the internet will benefit from the IP-independent validation mechanism that mCaptcha uses.

More broadly, wide-scale adoption of mCaptcha will improve internet privacy globally as existing CAPTCHA systems like reCAPTCHA and hCaptcha take advantage of their wide deployment to track users across the internet.

Where are your intended users, or audiences located?

Global

What is your name?*

Aravinth Manivannan

What email address should we use to contact you?*

realaravinth@batsense.net

Why are you, and your team members, the right people to work on this project?

I am the creator of mCaptcha.

mCaptcha has been in active development since March 2021. Over the year, I have spent more than 1,000 hours on project writing code, documentation, project planning, deploying infrastructure, and providing support to switch to and integrate mCaptcha.

Codeberg, based on Gitea, is the first popular service that will adopt mCaptcha to solve accessibility issues and replace their current CAPTCHA. I will work with Codeberg and Gitea maintainers to integrate mCaptcha in Gitea. This work has just started. This first production deployment is a huge opportunity for mCaptcha and will provide proof of its accessibility claims.

Aside from mCaptcha, I actively work on ForgeFlux, a project that is involved in software forge federation, and Hostea, which is a project trying to offer 100% libre, managed Gitea hosting. I'm also a strong believer in Free(as in freedom) Software, so all my work is exclusively licensed under Free Software licenses.

Internet privacy is not accessible to everyone, but I'm in a privileged position to self-host services, which are available for anyone to use free of charge at batsense.net that are open to all, free of cost.

Please upload any supporting documents to your application.

None

My application will be dismissed if it does not fit within OTF's mission, values, principles statements.

on

I have read and understand OTF's Terms and Privacy policy.

on

I understand that all intellectual property created with support for this application must be openly licensed.

on