Fighting Merchant Fraud: Using simple crowdsourcing techniques to identify bad actors in real time

June 2022

Overview

One of the major challenges in building a scalable and democratized Cost-per-Action advertising infrastructure is ensuring that fraudulent behavior is minimized or non-existent.

There are three primary participants in Asttro's CPA (cost-per-action) ecosystem: users, publishers, and merchants. Users are prohibited from committing fraud since in order to activate a payout, it would require a user to spend money with the merchant. Since the spend is greater than the payout - fraud would occur at a net loss for the user and therefore they should not commit it. Similarly, publishers cannot commit fraud because any fabricated action would require actual spending of money which would in turn result in a net loss.

The final participant in the ecosystem is merchants. In order for merchants to successfully launch a campaign and for Asttro to execute source attribution - it is required that merchants submit all of their wallet addresses in order for Asttro to be able to identify said transactions. However, if a merchant submits an incomplete set of addresses to Asttro it is possible that they will receive transactions that should be attributed to an advertisement but will go undetected by Asttro since the platform did not know to look for transactions assigned to that address.

In order to prevent this from happening, Asttro uses a crowdsourced flagging model to identify when a merchant has submitted an incomplete set of addresses.

User Reward Mechanism

Attro targeted ads are designed to reward users with cash back when they later transact with a merchant whose ad they interacted with. For example, when a user sees an ad through the Asttro display network - the ad itself is an interactive UI that when clicked logs a connection between the user, the merchant, and the publisher of the ad. Additionally, transaction rewards are attached to each ad and made visible to the user on display. An ad might offer 0.05 ETH back to the user upon purchase - this value will be made known to the user at time of ad interaction.

How Crowdsourced Flagging Works

Since all ads are associated with rewards, users can expect to receive a reward upon transacting with the merchant if they had activated an offer from that merchant. The mechanism for paying out user rewards is handled by a smart contract which is initially deployed by the merchant upon launching a campaign.

In order to recognize a transaction between a user and a merchant as one that is associated with a previous offer activation, Asttro identifies transactions that are associated with a merchant's known set of addresses as new

blocks are parsed. If a transaction is found where the recipient is a known address associated with a merchant that has launched a campaign on Asttro, then the purchasing party in those transactions is then checked against the set of user addresses that have activated offers for that particular merchant.

However, if a user transacts with a merchant but the merchant address is unknown in Asttro's system, then Asttro will not be able to associate the transaction with a previous ad interaction despite it being the case. In this situation, the user will not receive their expected reward (which would have otherwise been triggered by the smart contract) since source attribution cannot be completed. This can result in free advertising for the merchant and no payments made to the publisher or Asttro.

In order to flag merchants who are using unknown addresses to conduct business, Asttro takes advantage of the reward mechanism for users to incentivize notification of such situations. When a user transacts with a merchant who they were expecting to receive cash back from but did not - the user will have the ability to notify Asttro of the situation.

Once notified, Asttro will conduct an investigation to determine if the recipient address did indeed belong to the merchant and was in fact not included in the set of known addresses. If it is determined that the address was hidden from Asttro, the user will be rewarded for correctly flagging the merchant and the merchant will be publicly penalized. Penalties can range from decreased reputation rankings, which can adversely impact ad exposure and targeting, to being permanently banned from using the Asttro platform, which can have serious business ramifications for the merchant as the reach and network of Asttro expands.

How to flag a transaction
Users can submit their transactions to Asttro along with the relevant merchant information. If there are any receipts of further proof of transaction associating the transaction with the merchant they should be included here. This may include screenshots or other forms of evidence.

How Asttro determines if a flag was correct
Asttro will have a dedicated team for investigating reports of undisclosed addresses. Another option is to have a democratized voting body of token holders who will vote to decide if a particular case was fraudulent or not. Token holders will receive a profit share of Asttro earnings and are therefore incentivized to vote correctly since mistakes will adversely impact merchant trust of the platform.

Each case will be visible to the public and all decision making criteria will be documented for others to see. The following are steps that can be taken by the Asttro Anti Fraud team to determine if a flagged transaction was indeed fraudulent:

- Reach out directly to the merchant first for explanation or further detail. Ideally a solution can be quickly reached - though if found to be true will likely hurt the merchant's reputation. Since ramifications can be dire - it is in the best interest of the merchant to cooperate

- For other user addresses that have activated an ad from the merchant, Asttro can analyze blockchain data to identify if they have also transacted with the same flagged merchant account
- If the merchant has been reported by multiple users then there is an increased chance of violation
- Asttro can walk through the merchant's transaction process to see if transactions with different addresses are triggered

How users are rewarded for correct flagging

There are various ways that a user can be rewarded for correctly flagging a merchant:

- Receive their expected reward amount, as defined by the merchant for their campaign, through Asttro
- Receive a different amount, as defined by Asttro. in stablecoin
- Receive an Asttro native token
    - The value of this token can be derived in a variety of ways such as profit sharing or usage

Questions?

Please email hello@asttro.xyz