WhistleAI

Full Technical Version

Full Technical Version 0.6

Overview

WhistleAI seeks to attack corruption on the world stage by creating a decentralized platform wherein whistleblowers can safely and anonymously “blow the whistle” without fear of retaliation. WhistleAI is a decentralized marketplace that provides whistleblowers anonymity and allows them to monetize information that, until now, would never have seen the light of day. For buyers of information, such as journalists, new media, or watchdogs, WhistleAI uses a unique combination of artificial intelligence (AI) and crowdsourcing to verify the accuracy of the information being sold, while keeping the substance of the information private and unknown to anyone but the seller.

The Problem

When crime or corruption take place, there are often witnesses or people who have knowledge of the act. These witnesses fall into two categories: the bystander and the whistleblower.

Being a bystander is easy. Do nothing or say nothing and life will continue on as usual. In many cases, there may even be a financial or other benefit for remaining silent.

Being a whistleblower is hard. There’s very rarely a real incentive to expose the corruption of powerful people or organizations, and the prospects of career suicide and even violent retaliation are real.

First, a solution would be to protect the whistleblower’s anonymity. That would protect him from retaliation.

Second, fighting corruption requires that blowing the whistle be more lucrative than remaining a bystander. How can we make it more economically appealing to be a whistleblower rather than a bystander?

These two problems can be solved at once with a blockchain solution that simultaneously protects the identity of the whistleblower while making it economically feasible to expose corruption. Introducing WhistleAI, the solution to corruption in the world.

How do we protect whistleblowers while financially incentivizing them to do the right thing? How do we build a model that will turn more bystanders into whistleblowers? Until fairly recently, an immediate answer to questions like these was not forthcoming.

First is the problem of anonymity. We must make it so that the whistleblower cannot be identified. Let’s take a look at a typical information release where the whistleblower is kept anonymous.

An employee of XYZ corporation discovers evidence that XYZ Corp is dumping hazardous materials illegally. The corporation is saving millions of dollars by doing this, but the whistleblower can’t stand what is going on, so she decides to reveal her information to the world. How does she do this while protecting her identity?

In the current paradigm, our brave whistleblower can go to the press. She can take documents that prove the fraud beyond doubt and show them to a reporter. However, the whistleblower is now faced with the fact that she is risking everything: her career, her reputation and possibly her safety. Everything depends on the reporter, a person whom she has never met. Will this person protect her identity at all costs? Would this person turn down a million dollar payment from XYZ Corp in return for revealing his source? Would this person go to jail for the whistleblower if necessary?

Unfortunately, most whistleblowers decide that putting their lives in the hands of another human being is too risky. While we do hear about the occasional reporter who is willing to do jail time to protect a source, this is the exception, not the rule.

Media Markets Are Changing

There are also many situations today where a newspaper may not have any interest in covering the information the whistleblower is trying to expose. This is where WhistleAI can have far more impact than what’s currently available. WhistleAI can act as a connector between whistleblowers and a worldwide community of prospective buyers. In our scenario, perhaps the journalist isn’t interested in covering the story, but there are environmental groups that would be extremely interested. Perhaps it’s a vlogger or an online magazine who wants the story.

While media organizations and watchdog groups are clear examples of interested parties, we believe the real, true market potential for whistleblowers' information remains largely untapped.

This untapped market is new media. Today, traditional or legacy media are being disrupted before our eyes. Millions of citizen journalists now exist throughout the world, many of whom specialize in thousands of different subject matters. Many garner more views online than their legacy media counterparts. We believe that new media is ultimately a force for good and another step toward a more decentralized world. Whatever the subject matter is, there an alternative outlet searching for information that can break a story. We believe that WhistleAI can empower new media in ways never before seen. It will give them a marketplace to shop for stories and information that has always existed but has never been readily available until now.

WhistleAI stands to be a conduit between two massive markets that have never connected. As of right now, there are millions of new media journalists and content creators all over the world. The subjects they create content around are almost infinite. Some deal with politics, some with various social issues, others with food, others with animal rights, and the list goes on. This represents a massive market of untapped buying potential for information that is relevant to their passion or cause. We see new media as a smoldering bundle of sticks, waiting for fuel to fully ignite the fire. WhistleAI, in this scenario, will be gasoline. By connecting these content creators with millions of people in everyday situations where they witness wrongdoing, we can give them potentially endless sources of content to build their following. The age-old market of gossip finally has a buyer. It's easy to see the potential for WhistleAI at a higher level of crime, Wall Street, government, big business, but it also has many possibilities on a more micro scale. Think about the waiter at the restaurant who repeatedly sees the cook intentionally dropping people's food on the ground before serving it. A relatively small issue in the grand scheme of things. But what if that waiter, who's tired of seeing this happen, can now sell pictures he took of this to a local food blogger for $50? Don't confuse micro for small however, when it comes to impact. Imagine this scenario playing out all over the world, in every industry? We believe that the untapped market potential that exists between new media and everyday witnesses to wrongdoing is enormous.

The Importance of Anonymity

One of the problems with making whistleblowing safe and lucrative is that once the damaging information is revealed to anyone, it loses its value. How do whistleblowers expose information to the marketplace while simultaneously retaining anonymity and also being able to sell this information to the highest bidder, or outlet of their choice - all while keeping the information secret from any one person until the information has been successfully sold?

At WhistleAI, we believe we have the framework for a solution to this problem. A peer-to-peer network serves three powerful functions that will both protect and compensate whistleblowers. First, it can protect the anonymity of a whistleblower more securely than any existing method through a blockchain-based coin geared toward privacy. Second, through a combination of artificial intelligence and crowdsourcing, we can verify the information without any one source having access to the totality of the information. And third - and most important - this anonymous verification process can be used to provide powerful financial incentives for exposing corruption.

Having sensitive information exposed to multiple potential buyers presents a big problem though. It's not like selling a car or a laptop. I can show my car online and let people bid on it, but I can't do the same thing with a document proving malfeasance. Once the document's information is disclosed to anyone - especially if that information is newsworthy - the document loses most of its value. How does a whistleblower shop this document around without destroying its value? The solution is *minimum disclosure authentication*.

A Privacy Coin

In order to preserve a seller's anonymity and allow them to be paid for their contributions, WhistleAI requires an anonymous compensation method. A precedent already exists for private payments in the form of privacy based cryptocurrencies such as Monero, Dash, Zcash, Verge, and most recently PIVX. The WhistleAI platform uses its own type of privacy coins, named WISL. WISL is a fork of the PIVX privacy coin with adaptations to their governance model to better fit our needs. WISLs serve dual purposes as they will be the single payment method for buying and selling assets on the network, but they will also be used by network participants to stake as collateral while they help authenticate information on the network.

Minimum Disclosure Protocols

The idea behind minimum disclosure is the concept of breaking up a secret into fragments, and allowing anonymous community members to verify the secret piecemeal, so that no one person knows the secret, but the secret is still verified and its value authenticated.

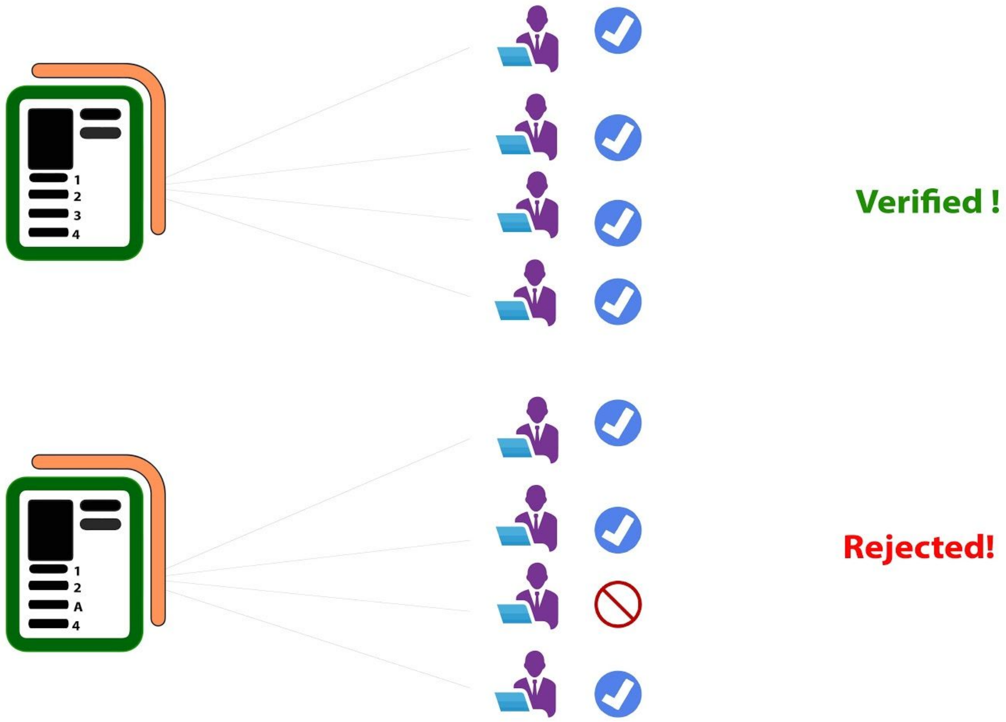

Figure: A Simple Minimum Disclosure Protocol

Crowdsourcing algorithms work well for authenticating general information, but what if that data is sensitive and requires privacy? In these cases, we need to figure out how to prove authenticity without full disclosure of the information.

In computer science, this problem set is referred to as a *zero-knowledge proof*. The idea behind the zero-knowledge proof is to create a verification mechanism that allows a separate party to verify underlying information without having to view the information in its entirety.

Minimum-disclosure protocols, or minimum-disclosure proofs of knowledge, are a subset of zero-knowledge proofs. This protocol takes information and chops it into smaller discernible yet incomplete pieces of information. The algorithm then discloses each piece, one at a time, to a verifier until confidence in the underlying information is reached.

Using a peer-to-peer network, we take the idea of minimum disclosure and combine it with crowdsourcing to achieve a decentralized, private information authentication algorithm. The algorithm splits information into chunks prior to any encryption of the information and sends those pieces to different nodes within the peer-to-peer network, asking for independent verification of each chunk. No user will be able to verify the object is absolutely what the asset owner states it is; however, if any chunk is obviously different from what the owner states it should be, then the entire asset's authenticity is invalidated.

Here’s a simple example. Suppose a whistleblower has a photo of John Doe stealing a cookie out of a jar. For the purposes of this hypothetical, let’s assume that the cookie theft is of great importance to a lot of people and that the fact that John Doe is the culprit is a valuable secret.

Once the world knows that John stole a cookie, the information is no longer valuable and will be impossible to sell. But if the network can verify the secret without actually revealing it, then the photo can be sold at a good price, thus keeping the financial incentive for whistleblowing intact.

The network would have built-in protocols that would be able to parse the expression “This is a photograph of John Doe stealing a cookie from a jar.” It would then break that statement up into two parts: “This is John Doe” and “This is a cookie being stolen.”

It would then create two versions of the photo. The first would blur out John Doe’s face. Members of the community would be compensated with cryptocurrency for merely answering the question: “Is this a photo of someone stealing a cookie from a jar?” If a sufficient number of users answer yes, then the first part of the secret is then considered authentic.

The second version would blur out everything but John Doe’s face, and ask the members of the community “Is this a photo of John Doe?” If a sufficient percentage answers yes, then the second part of the secret will be considered authentic.

If both parts of the secret are successfully authenticated, then all of a sudden we have a verified secret with great monetary value, but without having revealed the secret to any one person. This secret would then be offered to a network of buyers with the simple statement “WhistleAI has verified a photo of John Doe stealing a cookie from a jar.” With the network having authenticated the photo, the monetary incentive for disclosing criminality will have been preserved.

An AI Network for Information Authentication

Traditional whistleblowers give value to information by disclosing their personal identity as well as their relationship to the information so that it can be branded as trustworthy. With the anonymous sale of information, we must find another way to replace this idea of trustworthiness to extract the value from the information without attaching it to any individual as the source. The answer is the WhistleAI Network.

Under the hood, the WhistleAI nwtwork is a peer-to-peer network that connects machine learning with human crowdsourcing to facilitate the mechanics of the minimum disclosure protocol algorithm.

Peer-to-Peer Network Node Types

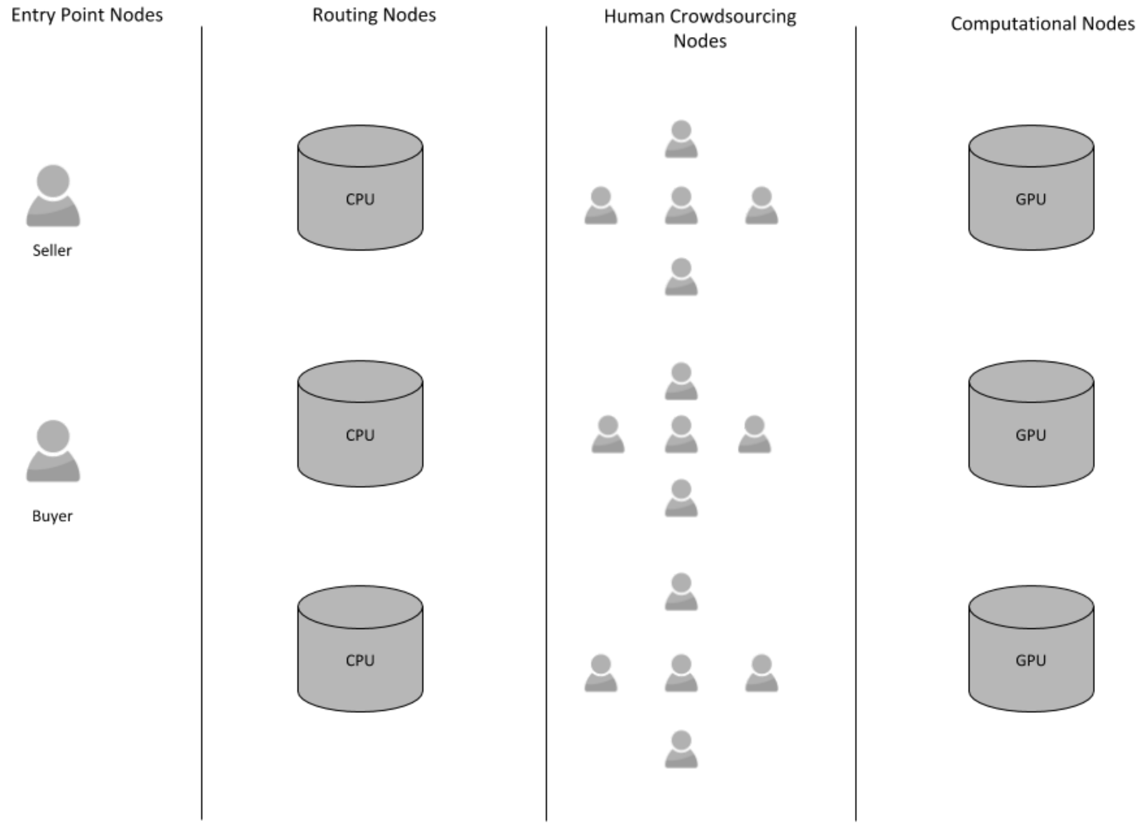

Figure: Visual Representation of WhistleAI Participants

Entry Point Nodes

This node type is the entry point into the network for users to initiate authentication requests and interact with the marketplace. Each user or application that wants to use the service must first download and configure this request node and wallet. Once installed, the client can begin submitting authentication requests with new information, as well as information bids on existing pieces of information. The authentication requests are submitted through the routing nodes, which provide status updates on their requests as progress is made.

Routing Nodes

Routing nodes are the expansion of the masternode concept used by the PIVX and Dash projects. Like masternodes, routing nodes must keep a complete and up-to-date copy of the working blockchain in order to validate all new transaction blocks, which secures the integrity of the network. Additionally, routing nodes evolve past masternodes by also functioning as the connection points for all task delegation for other participating nodes. For example, the routing node connects and handles a seller's data authentication request and then routes and tracks the progress through its various stages. In this instance, they would receive a request from an entry point node and facilitate the selection and guide the progress of the computation and crowdsourcing required for verification.

In addition to the standard blockchain financial transactions, these nodes are also tasked with storing the data for the marketplace bidding, progress of each asset authentication, and basic user data for all buyers and sellers on the network. All of the data needs to remain consistent across the entire network, so the state transitions of each object type are also tracked. These features are enabled by creating smart contracts using the opcodes supplied by the Bitcoin scripting system, although most of the marketplace data is stored off-chain using decentralized storage networks to give transaction volume higher scalability.

Distributed consensus on the blockchain is achieved and distributed using a unique private Proof-of-Stake[1] implementation put forward by the PIVX project.

Alternative Storage

On a blockchain, block sizes are limited so not all the data required for a fully functioning marketplace can be stored in the ledger. In order to increase both scalability and storage space, we use distributed storage networks to link to larger, externally stored data. These networks are a data storage abstraction that allows larger data to be stored and later retrieved using a unique hash key that can be stored on the blockchain as a reference to the larger piece of data.
To guarantee the security of our network data, we use a secure internal data storage protocol called SecureMessaging, originally written by the Particl Project[2] team. This protocol uses the already existing public key encryption to allow all data to be stored on every routing node, but is only decryptable by the intended recipient with the matching asymmetric private key.

Human Crowdsourcing Nodes

The human nodes interact with the network using a simple mobile interface that presents users with a question and a visual representation of the information. The user will be presented with the interactive option to swipe right if validating yes and swipe left if flagging as invalid or inaccurate.

The human nodes communicate only with routing nodes and not with other humans. Each human node sends a ping message to its list of routing nodes telling them it is available for a

[1] https://en.wikipedia.org/wiki/Proof-of-stake
[2] https://github.com/particl

task. In a response message, the routing node discloses to the human node a list of other routing nodes it can connect with as well. This is how routing nodes and human nodes become aware of each other, so that work can be properly crowdsourced as requests come in, and it also helps protect against individual nodes dropping off the network and isolating human nodes that had only a connection with that single routing node.

Accuracy Score

The routing nodes collectively keep track of two accuracy scores for each human node on the network. The first one is a cumulative score for all the work they have done throughout their history on the platform. The score goes up and down incrementally through every interaction with data. For example, if a human participates in a crowd vote and flags a piece of information as inaccurate, but the majority of the crowd approves it for authentication, then that human node's accuracy score for the charged routing node will decrease by one. Conversely, the nodes that vote along with crowd consensus will be rewarded by incrementing their cumulative accuracy score by one. This score is used by routing nodes to help determine a placement score, which determines the order of a human node's place on the work assignment queue.

$Cumulative\ Accuracy\ Score\ =\ \Sigma\ (a_n)$

* where each a_k *is either* 1 *or* -1 *depending on the accuracy result of each participating vote.*

The second accuracy score tracked is a time-weighted accuracy score, which gives more value to accuracy scores within the current thirty-day cycle than the true cumulative score. This version of accuracy is used in the calculation to determine shares of the *Participation Profit Pool* in order to give stronger pull to the results provided within the recent cycle.

$Time\ Weighted\ Accuracy\ Score\ =\ \ (.333\ \bullet\ Cumulative\ Accuracy\ Score\ )\ +(.667\ \bullet\Sigma(\ b_n))$

$*\ \ where\ each\ \ b_k\ is\ an\ accuracy\ result\ within\ the\ current\ 30\ day\ window.$

Participation Multiplier

Active participation will be rewarded by counting a user's continuous streak of activity. Human nodes broadcast their availability to routing nodes. They must continually do so after every completed message is sent from a routing node or the node will reset that human node's counter. The counter can also be reset if a participation request is sent to a node and a busy signal response is sent back, or the request timed out before a response is given.

This participation streak counter is multiplied by the Cumulative Accuracy Score in order to determine a user's *Placement Score*. This score as briefly mentioned above, and it gives faster access to new tasks within the network.

$Participation\ Multiplier\ =\ 1\ \bullet\ n$

$* \ where\ n\ is\ the\ number\ of\ times\ a\ user\ has\ consecutively\ participated$

$$Placement\ Score\ =\ (Cumulative\ Accuracy\ Score) \bullet (Participation\ Multiplier)$$

Routing Distribution to Available Human Verification Nodes

The Placement Score gets sorted in a first-in-first-out (FIFO) queue to distribute work to the human nodes. An availability notice sent by a human node will initiate the router node to include that peer in its current placement queue. Each routing node will keep its own version of a participation multiplier and placement queue. The order does not need to be shared or agreed upon through consensus by all of the routing nodes. When it is time for the routing node to delegate a task to humans, it will distribute information to the first N participants taken from the queue. The number of participants, N, will always vary depending on the problem set, as well as current network configuration settings.

The remaining nodes not selected will remain on standby in case of response timeouts from other participants, or in the case of a node delegation conflict. By being included on the waiting list, they will receive the participation multiplier effect as long as they remain active for the next round of work delegation from that routing node.

Human Node Conflicts

In the case where a human node receives two requests simultaneously from two different routing nodes, their node protocol will accept the one that it received first and send a busy signal back to the other routing node. In the event of a tie down to the millisecond, the node will choose a random one to select. When a busy signal is received, the routing node will then extend an invitation to the next available user in its placement queue.

Computational Nodes for Machine Learning

Most machine learning algorithms require data-heavy computations. Neural networks in particular require multiple parallel processing units. For this purpose our network requires specific nodes that do the computational heavy lifting. These nodes will work with the routing nodes which will help organize the parallel processing work needed by the machine learning algorithms that have been trained within WhistleAI. In order to meet the system calculation requirements these computational nodes will all use Graphics Processor Unit (GPU) instead of the standard Centralized Processor Unit (CPU) used on most personal computers. GPUs are optimized for processing multiple computations in parallel and therefore can help a neural network complete the required calculations faster and reach convergence.

An example of a useful artificial intelligence node within our system would be a convolutional neural network (CNN) trained specifically to classify, recognize, and identify human faces. A CNN is a specific category of deep learning neural networks adept at classifying visual imagery.

The convolution layer is a filtering step that precedes analysis from the neural network layers, with the purpose of identifying depth in an image by identifying local maximum values while also shrinking the total inputs required to be analyzed by the neural network layers. The edge detection trick allows the neural network to estimate three-dimensional measurements of a face extracted from a two-dimensional image. A unique set of facial feature measurements is how the neural networks can compare faces to identify a unique facial match within other images.

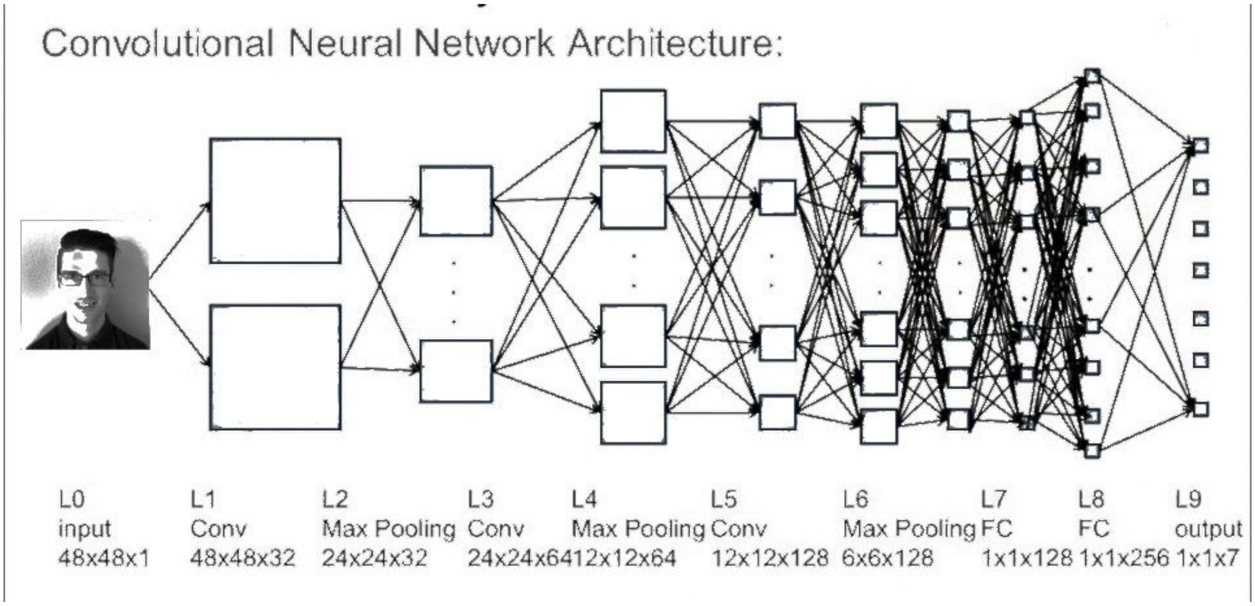

Figure: Convolutional Neural Network Architecture for Facial Recognition

Network Incentives and Fees

The WhistleAI incentive system for participants happens in two phases. On the front end, participants receive compensation for their role in the authentication process as their work is completed. These participants will also receive a portion of the sale of assets in the future. If we connected the compensation to the direct sale of the asset, however, it would cause a conflict of interest for the authenticator to flag a piece of information as invalid. Instead, therefore, we combine the portion of proceeds for successfully sold items into a pool. The distribution of this *Participant Profit Pool* happens at the end of a thirty-day cycle, and the earnings are paid out based on the profit units earned by each participant within the window of that cycle.

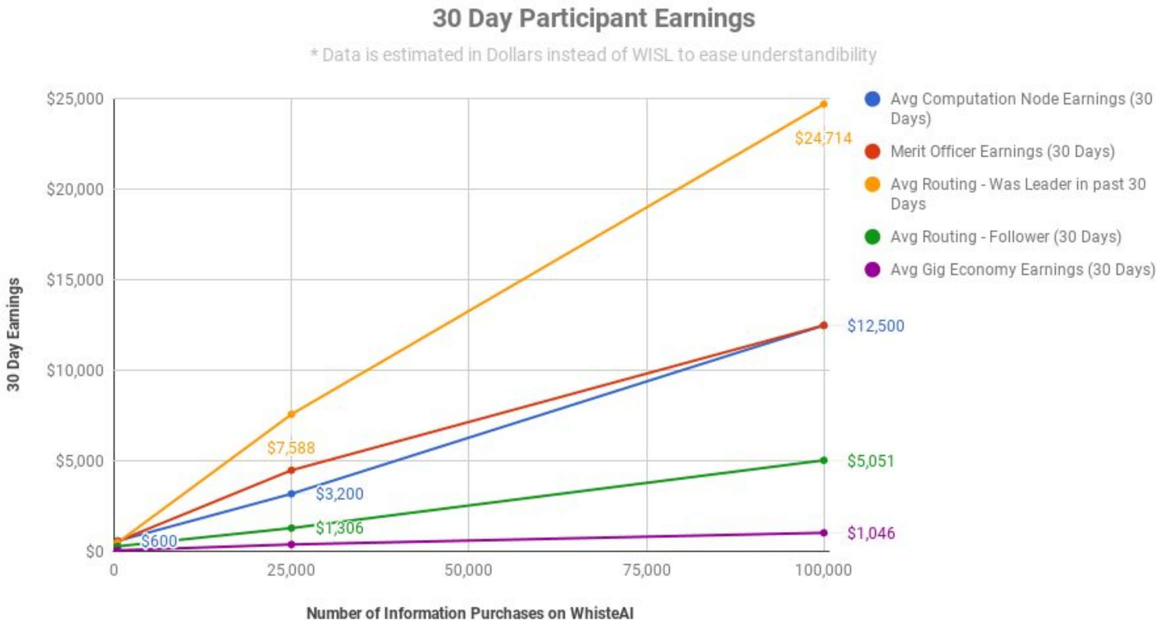

Figure: Potential Participant Classes Earnings Based on Information Transaction Volume

Marketplace Seller Authentication Fees

In order to be able to sell an asset, the seller must have it authenticated by the network. A flat fee, denominated in WISL, is established for every thirty-day cycle. The fee is adjusted at the beginning of every cycle, based on network demand as well as price fluctuation, with the general goal of keeping the fee stable and reasonable for sellers worldwide.

Marketplace Asset Purchase Fees

The transaction fee on the sale of an asset will be set at seven percent of the purchase price. One percentage point of that fee goes toward a pool to compensate Merit Officers. The remaining six percentage points go into the Participation Profit Pool.

Human Crowdsourcing Participation Incentives

A quorum of at least 10 participants is required for each minimum disclosure item authentication within the crowdsourcing algorithm. In simpler terms, this means if an image is split into three pieces, then 30 human participants are required to authenticate a piece of information. Each quorum votes, and the voters in the majority group earn the reward. The dissenting voters receive nothing. The combined quorum participants equally split 60% of the authentication fee, and a portion of shares in the Participation Profit Pool, based on a score calculated relative to the others in their quorum. This score is based on a number of factors including their accuracy score, the amount they have cumulatively staked, and the hours active during the current cycle.

Routing Node Hosting Incentive

Every node online earns a portion of the authentication fee on the front end, and shares from the Participation Profit Pool on the back end. Routing nodes also earn block rewards for their role as masternodes, helping secure the network by validating new transactions. The distribution of the Participation Profit Pool is based on the number of hours a node is active. This is tracked by recording the current stake each hour and adding it to the accumulated amount for the current thirty-day network cycle.

Neural Net Computational Hosting Incentive

The incentive system we have put in place to reward participation is analogous to proof of work mining for a blockchain. Unlike the other participant classes, computational hosts do not require staking of WISL. These nodes are rewarded solely based on the energy they expend to compute calculations for a neural network. The measurement for this is called *work units* and is based on the convergence time of the neural net to reach an inference relative to the processing power of the machines. The Participation Profit Pool allocation is distributed pro-rata based on earned work units during a thirty-day cycle.

Block Reward Breakdown

- Routing Nodes are allocated 45% of each block reward
- Staking vakidator block lottery winner is allocated 45% of each block reward
- Merit Officers are allocated 5% of each block reward
- The developer fund is allocated 5% of each block reward

Participation Profit Pool Breakdown

- The Human Crowdsourcing Nodes are allocated 59%
- The Routing Node are allocated 16%
- The Routing Node followers are allocated 8%
- Computational Nodes are allocated 20%
- The Developer Fund is allocated 5%

Authentication Fee Distribution Breakdown

- The Human Crowdsourcing Nodes receive 60%
- The Routing Node leaders receive 10%
- The Routing Node followers receive of 5%
- Computational Nodes receive 20%

Peer-to-Peer Communication

Messaging

All peer-to-peer messages use a common protocol structured in JSON format. Each node is given a unique ID and messages are sent over TCP. Anonymous marketplace users shall be granted a new ID every time they re-authenticate into the network for additional obscurity. Every task that gets passed between the network has its own unique message type. Message types received that either aren't signed or are signed incorrectly are discarded.

Authentication

In addition to a connected node ID, each node has another unique address that functions as a public key. This key uniquely identifies nodes on the system, and provides the decryption key to verify digital signatures.

For this to work properly, much like existing blockchain projects like Bitcoin and Ethereum, our digital signatures will use asymmetric cryptography, specifically the Elliptic Curve Digital Signature Algorithm (ECDSA)[3]. When a unique private key is created, a corresponding public key is generated as well. This private key is used by nodes to encrypt hashes of information to generate a unique digital signature. These signatures can be verified by decrypting the hash using the public key, and by verifying that the decrypted hash equals the hash of the original information sent. As a fork of PIVX we will continue to use the quark hashing function[4] as implemented and tested by their development team.

Discovery

Each peer in the network keeps a list of all other known active nodes on the network. The nodes communicate with each other in intervals to share information. One of the pieces of information shared is their most up to date list. Each node then adds missing nodes to its own master list, and begins communicating with those nodes as well.

Not all nodes will keep a list of all node types. For example, human nodes need only to communicate with routing nodes, and therefore will not keep a list of other human nodes in the network. Routing nodes, on the other hand, communicate with all network nodes and must keep a comprehensive list. If a node is brand new to the network, it must connect with a known address of an active routing node; otherwise it will remain in a stand-alone state forever.

[3] https://en.wikipedia.org/wiki/Elliptic_Curve_Digital_Signature_Algorithm
[4] https://en.wikipedia.org/wiki/Quark_(hash_function)

Application Security

Peer-to-peer networks face different security challenges than the centralized client/server networks such as Amazon, Facebook, and other modern web applications. Client/server application architectures have a central security model, which means all data remains on a protected server. This generally creates a scenario where only a single access point needs to be secured. The common web protocol used to secure these application access points is HTTPS.

Peer-to-peer networks have no central server, so security is a much harder problem to solve. Every communication between users needs to be independently secured so that it can be trusted. The current best practice of secured peer-to-peer networks is asymmetric encryption, also known more commonly as public key encryption. In this model, two unique keys are created to encrypt and decrypt the messages being sent through the network. If one of these keys remains private to an individual peer within the network, and the other shared with all other users, communications can be privately encrypted by other users. Also, messages can be guaranteed to have been generated by the source. These properties create two useful security protections.

First, a user can send a message to another user encrypted with the public key so that only she will be able to decrypt it with the unknown private key. Second, messages encrypted with the private key are verifiable to anyone with the public key. This reversal is useful if one wants to verify that the message originated from the expected source. This latter capability is called a *digital signature*, so that all messages can be verifiably pointed back to their original destinations. When requiring user-to-user authentication instead of user-to-server authentication, digital signatures are key to secure communications.

Security of peer-to-peer communications and anonymity of payments are both crucial to the success of WhistleAI. We are fully aware that vital information only maintains its value while it remains private. To protect the integrity of the information's value for the seller, it must be stored securely on the network and can be passed along only to trusted nodes. Within the network, the only nodes that should receive a full piece of information are the node of the owner of the information herself, as well as the machine learning computational nodes that obfuscate the information prior to human analysis.

To ensure that the information remains only within the intended nodes, we use digital signatures to safely ensure accuracy of the source and public key encryption to guarantee that the information is decryptable only by the intended recipient. For this functionality, we use a Blake256 hash function[5], which is currently in use by the Decred cryptocurrency project to

[5] https://en.wikipedia.org/wiki/BLAKE_(hash_function)

secure their blockchain. Blake256 is designed as a safer alternative to the SHA256 hash function used by Bitcoin.

The coin used by WhistleAI is deployed using blockchain technology that has been production battle-tested for security as well as proven anonymizer algorithms to protect the identity of the users.

Storage of Sensitive Information

In case the encryption is ever broken through quantum computing, we add an extra layer of security for the storage of sensitive assets. Starting with images, we chunk the files in such a way that no image is ever stored in its full form. Each chunk is then randomly distributed to exactly one-third of the routing nodes. This practice will ensure that it will be highly unlikely that any routing node holds a full copy of the image. Further, only the entry point node that originates the image, owns the copy of the matrix that would piece the image back together.

Proof-of-Stake Consensus

The crux of the decentralized consensus problem is the need to have consistent shared data among a group of parties each holding a copy of the data. Consensus requires the data to be accurately updated for every copy whenever a change is made.

WhistleAI uses a proof-of-stake distributed consensus algorithm to help secure the network. Stakers compete on the network to be selected to be the initial validator a block, which holds a group of transactions. Each staker’s odds of winning the block are proportional to that wallets staked amount versus the total amount being staked on the blockchain. For example if Bob is staking a wallet with 1000 coins, and there are 100000 coins currently being staked on the blockchain, then Bob has a 1% chance of winning the lottery to validate the next block. The more coins being staked on the network, the harder it becomes for a malicious staker to incorrectly validate blocks. Further because the blockchain “cold stakes”, meaning makes coins unspendable and held as collateral, the malicious actors reward must outweigh the potential loss they could accrue by improperly invalidating and forfeiting their amount staked.

The most recent PIVX implementation allows private staking using zerocoin protocol coins. This first ever “stealth staking” implementation allows for coin stakers wallet balances to remain private so that the amount of wealth can remain private to help further increase wallet anonymity.

Stealth staking also increases the amount of zerocoin transactions happening on the network which further helps prevent timing vulnerabilities. Without a high volume of zerocoin transactions, transactions could theoretically be traced to specific wallets based on the timing of transactions from wallets relative to a low number of zerocoin transactions taking place on the network.

Flow of Information

The network itself is comprised of a group of nodes communicating together to authenticate data that has been submitted to the network. The authentication process is made up of four components: input processing, machine learning, human analysis and accuracy oversight.

Input Processing

A request to authenticate a piece of information must be submitted through a peer-to-peer node. The information must be an accepted file format (png, pdf, mp3, mp4, avi ,docx, etc.) in order for the request to be processed. Based on the digital file type, the network then creates a sequence order of required authentication steps to be taken and keeps track of the progress of the request.

Machine Learning

In the first phase of authentication a digital file will be passed through different machine learners using algorithms to filter for specific things on specific types of data. These filters can range from relevancy filters to determine the piece of information will be valuable on the network, all the way to facial or voice recognition that helps identify an individual’s identity.

Human Analysis

As part of the minimum disclosure protocol, the digital data file is altered for privacy and anonymity of the individuals in the pieces of information by machine learning algorithms. It is then passed to the human nodes whose function is to group-analyze the same piece of information and extract a yes or no answer about the piece of information.

Not all participating human nodes will vote on every piece of information that comes through the system. Human work assignments are delegated based on an ordering algorithm, called *Priority Rank*, which takes into account current time spent on the platform as well as an historical accuracy score of work done on the network (see Priority Rank details below under the WhistleAI Gig Economy section).

Accuracy Oversight

As the final step before each piece of information can be deemed fully verified, it must pass through the oversight measures we have put in place in order to increase the rate of accuracy. The primary means of oversight is by a group of pre-defined moderators, called *Merit Officers*. A credentialed Merit Officer has earned her participation level through consistently high-quality authentication work on the network. The main function of a Merit Officer is to monitor for flagged authentication requests as they are processed, whether it is by a machine learning algorithm or a human verifier.

Additionally, we also include an escrow and arbitration service for added assurance that a user gets exactly what she was expecting when she purchases a piece of information. In these cases, a buyer spends an additional 1% in cost for the option to use an arbiter if the piece of information she bought is fraudulent. After the purchase, the funds of the sale will be held in escrow for 48 hours before the seller can claim it. During that time, if the information is deemed “not acceptable” by the buyer, the buyer can open a complaint. Once a complaint is opened, a Merit Officer is randomly assigned to the case through a round-robin algorithm. The Merit Officer then looks at the information in question to determine whether or not the seller acted in good faith. The Officer’s arbitration ruling on the complaint is final. Both parties must acknowledge this finality ahead of time.

Mobile Application & Participants

Human nodes are to interact with a swipe-right, swipe-left interface. The action of swiping right indicates a positive verification and swiping left indicates the piece of information needs to be flagged as fraudulent or not what the seller has indicated.

For each human attempt at authentication, there will be a set number of voters assigned a given task. The voters must collectively approve the asset at a percentage above a set threshold for authentication. Each human verifier’s voting result will have either a positive or a negative effect on her accuracy score based on whether or not she was among the majority. The higher a user’s accuracy, the faster she is likely to receive another task delegated to her from the routing nodes.

An incentive system is built in to reward the human nodes for their time and participation. This work utility will add another dimension to the gig economy made prevalent by companies like Uber and AirBnB. The goal is to have a network of over 500,000 people collectively interacting with and teaching the AIs on our platform by 2020. We feel that a large human network of intelligence to learn from will be a game changer for the progress of Artificial Intelligence.

The Marketplace

At its core, WhistleAI is a marketplace for content that connects buyers with sellers. The key feature to enable buying is our auction system. The state and values of each auction item are stored directly on the blockchain with larger items storing references that points to data stored within SecureMessaging. The associated currency and unit for bidding is WISL.

The Content

When content is uploaded to the system, it includes a digital file, description, initial list price for the sale and any other supporting documents required for authentication.

The description is used to make sure the content follows the intentions of the network per the founding principles to help govern the network. It can also be used as the barometer for the human analyzers to measure the actual asset against for trustworthiness. This basic piece of information supplied with the authentication request should allow them to make a more accurate snap judgement. If more than 50% of crowdsourced participants flag a piece of information, then it will not pass authentication and not be sellable through the marketplace.

The Buyer

Buyers within WhistleAI have the option to disclose themselves and have their identity verified or remain anonymous. Verified buyers have a distinct advantage because sellers will be able to seek them out with offers for relevant information. Anonymous buyers must search through the network for interesting information themselves. Making it even harder is sellers may requiring the buyer of their information to be verified, making the possible pool of information to buy smaller.

The Seller

The only information stored about the seller through the network is a unique ID, that is just a unique number. This ID must be different from the wallet address to help preserve a sellers anonymity. All other information about the seller should never leave their own peer node and will permanently delete itself after a transaction has been completed.

The seller is ultimately in control of what happens to the information. For instance they can restrict to only sell to buyers that are known and verified so that they can maintain an understanding of what will happen with the formation after the sale.

Asset Exchange

The asset shall remain encrypted and only stored on the sellers node until purchases is complete. Upon agreement and execution of sale the asset shall be unencrypted by the owner and then re-encrypted and exchanged through the SecureMessaging protocol such that only the buyer can decrypt the asset.
Once the transaction takes place and the buyer is satisfied then the funds are released from escrow. For extra protection for both the buyer and seller, a merit officer can be added as an escrow signatory as well for an added fee, requiring only 2 of the 3 signatures to release the funds.

Internationalization of The Application

WhistleAi is a global marketplace, and therefore the mobile application needs to be available in multiple languages so as to include all users with important information to sell. Because adopting languages is a challenging task, however, only English and Spanish will be supported in the first version. Future flexibility will not be ignored, however. Languages are to be installed as swappable plugins and the code open-sourced so that anyone in the community can have the ability to submit new language capabilities.

Gig Economy

Human beings are an essential aspect of machine learning and the AI Network. A three-part incentive system is built in to reward the human nodes for their time and participation. This work utility adds another dimension to the gig economy made prevalent by companies like Uber and AirBnB.

Three-Part Human Incentive System

Money alone can't motivate every type of person. Recognizing this, we aim to deliver a three-part reward system based on merit, to both incentivize performance and to add the elements of accomplishment and progress to the human verifier experience. As a person participating in the WhistleAI gig economy, there are three incentives for performing. WISL coins, rank, and badges. When a person joins the AI Network, they begin earning WISLs for accurate verifications. As they verify more and more tasks accurately, they begin to earn badges and ranks within the system. The more badges and the higher one's rank, the more important her role becomes in the WhistleAI ecosystem, and subsequently, she gets faster access to more verifications, which equates to higher earnings over time. In this model, long-term loyal workers can get priority access to work, but new participants can hang out on the network and built up their Placement Scores in order to get a higher rank in each routing node's placement queue. The participation streak element of the calculation allows new participants to compete with longstanding participants by building up current active participation by waiting on the network.

Badges and Rankings

A user's rank score is designed to grow slowly if a user consistently works every day with a high degree of accuracy. It will begin to decrease, however, if a user's accuracy score decreases or the hours worked do not continue to grow faster than the total days the network has existed. In the latter case, a user should always see growth as long as she participates for at least two hours a day on the network.

$$Rank\ Score\ = (Hours\ Worked/Days\ Network\ has\ existed) \cdot Accuracy\ Score$$

Badges are earned and lost every time a rank score moves above or below a new number in the Fibonacci sequence. The first ten unique Fibonacci numbers are 1,2,3,5,8,21,34,55,89 and 144. These are also the first ten badge levels on WhistleAI.

Merit-based Governance of Participants

Moderation positions help govern both the crowdsourcing participants and the quality and relevancy of the information being submitted to the network. These moderation positions are called *Merit Officers* and are paid positions in the WhistleAI network. Merit Officers are the decision-makers for WhistleAI and are responsible for keeping the network “White Hat.” They are compensated by a combination of proprtional fees from each network transaction as well as portion of each block reward. To become a Merit Officer, a human node must reach the top 10% of total network earnings, accuracy score and rank score. Once all three are achieved, a user becomes eligible to become a Merit Officer. Merit Officers govern each other: if a Merit Officer is removed from her post, she is never again given the chance to regain that status.

Evolution From a Private to Public Network

The first version of WhistleAI is a private network with trusted computational nodes operated completely by the developers of WhistleAI. In the interim, the Computational Node client code shall remain private and connection access for these nodes must be whitelisted and approved by the developers. This private controlled environment will allow us to solve the hard information authentication problems and improve upon our product with real users. Further down the road, we will ultimately add in trustless algorithms for the artificial intelligence based nodes to achieve a byzantine fault-tolerant decentralized public network.

Importance of Human Analysis

Human beings will interact with an interface that allows “swipe right” for a positive verification and “swipe left” for negative votes against the piece of information provided. For each human attempt at authentication, there will be a set number of voters assigned a given task. The voters must collectively approve the asset at a percentage above a set threshold for authentication. Each human user’s voting result will have either a positive or a negative effect on her accuracy score based on whether she was among the majority. The higher a user’s accuracy, the faster she is likely to receive another task delegated to her from the routing nodes and the higher rank she will be able to receive.

A White Hat Network

WhistleAI is fanatical about being a force for good, and will be designed that way. We are acutely aware that sensitive information can be used for good or for evil. Blackmail is a crime for a reason: society doesn't benefit when one citizen threatens to expose another's embarrassing activities - especially activities that harm no one. With that in mind, the WhistleAI network contains ground rules that prevent it from becoming a breeding ground for "Black Hat" uses.

The WhistleAI constitution is a small set of founding principles that set present and future moderating guidelines for Merit Officers to enforce and modify. Like any good constitution, the guidelines can be amended by a two-thirds majority vote of special governing nodes called Masternodes. The voters are instructed to make sure the guidelines always follow and maintain the ethos of the founding principles.

Founding Principles of The WhistleAI Constitution

1. WhistleAI exists to expose harm being inflicted on others
2. WhistleAI does not exist to inflict harm on others

Founding Set of Amendable Guidelines for Merit Officers Enforcement

1. Abuse of power is always fair game
2. WhistleAI is not a tool for individual blackmail. Disclosures of affairs, as long as it is not an abuse of power, are forbidden.
3. Disclosure of a person's sexual orientation or gender identity are private and not harmful to others. Such disclosures are forbidden.
4. Pedophilia is fair game, but it is forbidden to sell this type of information to the intended target as a form of blackmail.

When content is uploaded to WhistleAI, it must be accompanied by a description. The network identifies and flags descriptions that contain words and phrases that could potentially represent "Black Hat" material. At this point, that content goes to a Merit Officer for review.

Additionally, as part of the verification process, WhistleAI AI compares the content to the description, to ensure that there is a match. If the AI cannot confirm, then the file is flagged and sent for review by a Merit Officer.

Governance

Governance plays a key role in the success of any decentralized network. If changes and adaptations cannot be easily made, progress will most likely stall, leading to the demise of the network. Our selection of PIVX as our key building block was due not only to its anonymity, but also because of its well-designed masternode voting system of governance. PIVX masternodes vote to approve new features onto the network, which then rewards the developers in a purely democratic manner.

Governance on WhistleAI works more like a republic than a democracy. Routing Nodes, like the masternodes, function as the voters on the system. Every 180 days, a special election is held to elect five Merit Officers and three lead developers for the project. Designated block rewards over that period are distributed to the wallets of those elected for the duration of that cycle.

The Merit Officers are the enforcers of the “White Hat” constitution, which carries the moral compass for the project to follow. Merit Officers can attempt to amend the constitution by submitting proposals. These proposals are voted on during a thirty-day network cycle, and approved if a two-thirds majority is reached. Further, Merit Officers must oversee each other’s actions. If a merit officer finds that another Merit Officer is acting contrary to the Constitution, she can start an impeachment process against that other Merit Officer. If the routing nodes vote to remove the Merit Officer, the other Merit Officer who proposed their impeachment shall receive the terminated officer’s block reward share for the rest of that term.

The elected developers are charged with maintaining and improving the network. Each new feature released must be approved. To incentivize productivity, the developer portion of the rewards for the cycle are chunked into 6 buckets, marking 30-day increments to match the cycle on the network. If a lead developer has not gotten a new feature approved on the network during a thirty-day window, then she will not receive her share of the reward during that thirty-day cycle.

Financial Model

Market Size

Corruption never seems to be in short supply; in fact, it seems to be one of the worlds most abundant, renewable resources. And while there has always been an interest in exposing corruption, a true market for it has never been able to form. Legacy media, the traditional buyer for such information, always represented a centralized “bottleneck” for information, which limited the size of the potential buying market for information. In the mid 20th century, Journalist A.J.

Liebling stated that "freedom of the press is guaranteed only to those who own one." Meaning that only those voices powerful enough to produce a book or a print a newspaper could be heard. Today, the power of the press is being disrupted and redistributed among the people. This represents a massive shift in the economy of information. We are witnessing the creation of a worldwide market for information, this likes of which this world has never seen. The potential size of this market is unknown yet, but we estimate it to be in the billions of dollars, with exponential growth following.

There are, however, challenges still facing this new media. “Fake News” as it’s been labeled, has plagued New Media and now Legacy Media is even coming under fire. The problem is that *trust* in the information being provided by media has traditionally been granted through the authority of the oligarchies that control it. As New Media is coming online, there is no authority to vouch for the information being provided, as the world has had in the 20th century. And as New Media and Legacy Media create conflicting content, both are feeling the pinch of an erosion in trust by the consumer.

We have seen a similar phenomenon with the evolution of e-commerce. A shift in trust from the big box retailer to the individual seller of goods had to take place. In the Ecommerce world, it happened through a combination of mechanisms such as online reviews and frameworks, like Ebay, which lent legitimacy to the transaction, without actually being the buyer or the seller in the transaction. Buying something before physically inspecting it is a very risky proposition. Allowing potential buyers to sample trusted customer reviews before purchasing created enough of a comfort level to allow consumers to buy things online. Buying something online will always be more a convenient proposition than having to go to a brick and mortar store, as long as you can trust the seller. Customer reviews of products supplied that trust, and as the public trust in e-commerce has grown, it has become over a trillion dollar per year market.

We feel that today there are two massive markets, information sellers and New Media, that need a conduit in which to transact. A framework that allows buyers and sellers to transact without being directly involved. Just as Ebay first provided the validation and trust needed for online transactions between individuals to take place on a massive scale, WhistleAI provides the framework and verification for information to be transacted between New Media / Legacy Media buyers and the world of everyday people who witness wrongdoing and corruption.

While it’s difficult to put an exact number on this emerging market for information, we feel that many similarities hold true between it and the birth of Ecommerce. We are injecting trust for online transactions when it only existed previously face-to-face. We are also creating buyer convenience and saving content creators and other buyers days or weeks of research.

Transaction Breakdown

Before a piece of information is sellable on the network, it must first be authenticated. The cost to get a piece of information verified by the network is a flat up-front fee plus a portion of the sale price going to the network profit pool.

As mentioned previously, the flat fee will be designated in WISL and can be adjust for every new cycle based on the market volatility of WISL as well as the demand for authentication. The goal is to allow the fee to increase based on demand, but to remain stable despite the volatility of the price of the coin. The flat fee proceeds are split by granting 60% to the human validators and 40% going toward the other nodes as broken down in the Incentive and Fees section above.

Once a buyer is found and a piece of information is sold, then the proceeds of the sale are split into multiple initial buckets. The first is the seller, who receives 93% of the buy price. Of the remaining 7%, 1% goes to the merit officer pool, and the rest goes into the Participation Profit Pool. The allocation of funds from the Participation Profit Pool are listed above under the Network Incentives and Fees section. The Participation Profit Pool distributes once every thirty days for the work done in the prior cycle.

This pool exists to prevent human authenticators from receiving a direct kickback from the deals they approve. This mechanism is to help remove the temptation for participants to falsely approve bad information in hope of receiving an allocated share of the profits. If all profits are shared in a way that is directly correlated toward positive accuracy and time spent on the network then the end result should be a more truthful authentication and oversight process.

Development Roadmap

Phase I: Launch our Privacy Blockchain and WISL

Phase II: Connecting Facial Recognition Neural Nets to the Peer-to-Peer Network

Phase III: Beta Information Marketplace

Phase IV: Launch Mobile App for Human Nodes and Enable Minimum Disclosure Protocol Authentication

Phase V: Add Audio & Video Asset Types

Phase VI: Launch Full Governance and Voting System for Decentralized Development

Phase VI: Selection of Initial Public Merit Officers

Phase VII: Incorporate Trustless Computational Nodes

Phase VII: Launch of Completely Decentralized Whistleblower Marketplace with Anonymity

Conclusion

The WhistleAI network has the potential to change the world for the better. At the same time, it will create a massive new market for information that has never had an outlet. From multimillion dollar pieces of information on the wrongdoings of the powerful to the $400 piece of evidence that the construction company is cutting corners.   Until recent technology advancements, there was never much hope for this kind of change, but today with the decentralization of blockchain, the advent of cryptocurrency, and its ability to enable secure, private payments as well as the progression of Artificial Intelligence, we are finally reaching a point where technology can provide the anonymity, incentives and verification to allow the truth to come to light. Louis D. Brandeis once said “sunlight is said to be the best of disinfectants.” We believe that WhistleAI can bring sunlight to the darkest corners of the world, by incentivizing whistleblowers and simultaneously protecting them from the criminals they expose.