- IPFS - The Interplanetary File System
- A Blockchain based Privacy-Preserving Incentive Mechanism in Crowdsensing Applications

WitnessChain

4/16/18

Dhruv Gupta, Neel Mehta, Kevin Sadhu, Harshal Singh

{dgupta, neelmehta, kevin_sadhu, hsingh}
@college.harvard.edu

What is WitnessChain?

Problem

How can we build a way to incentivize crowdsourced collection of parking ticket evidence for metropolitan police departments?

Updates

- Ethereum server built
- iOS front end built
- Image storage on Google Cloud

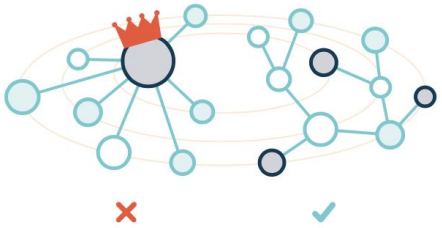
Intellectual Challenges

- Incentive system assuming police & citizens don't trust each other
- Preventing unauthorized viewing of preview images
- Preventing screenshots

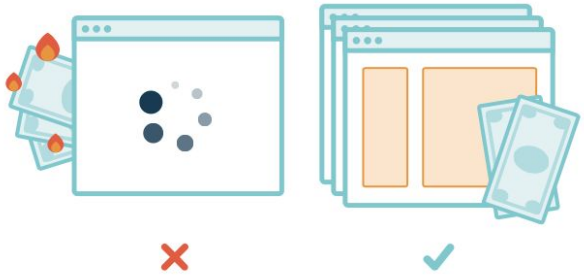
Paper 1: IPFS (InterPlanetary File System)

Problem Statement

- Centralization



- Cost of HTTP



- Record-Keeping
- Resilient Networking

Proposed approach

- Replace HTTP standard with IPFS
- Files stored by interesting nodes
- Duplicate removal, version history, immutable files
- Look up files based on hash (easier!)
- Decentralized naming system

Summary of Results

- Built a robust, Peer2Peer file system
  - Data is shared, decentralized
  - No node is privileged
- Network stack builds on networking technological advancements
- Distributed Hash Tables are used to maintain coherence among the nodes

This idea is built on the crowdsourcing of storage - where each local node provides a means of storing the file, and contributes to routing access requests. Data of the chain is stored and used to increase functionality

Key Enabling Points

- Large dataset accessibility
- Archiving
- No need for a centralized backbone - low latency

Takeaways

Distribution

Distributing the load of computation can increase efficiency. It adds a sense of crowdsourcing to the product - works well with public evidence reporting

Simplicity

Blockchain aspect is abstracted away for the end user - they only see the benefits of it

Crowdsourcing

IPFS relies on the established community of file users. WitnessChain would harness this to store its data (images and timestamps)

Paper 2: Blockchain Incentives (Privacy Preserving Crowdsensing Incentives)

Problem Statement

- How do we eliminate the privacy concerns of centrally stored user information, while also incentivizing participation in crowdsensing applications?
- Transaction info can disclose user privacy, how can we ensure anonymity?

Proposed Approach

- Blockchain-based incentive mechanism.

  Miners evaluate quality of data. Eliminates need for central authority.
- Node cooperation privacy method for all users to achieve k-anonymity

Summary of Results & Key Enabling Points

1) Traditional BlockChain Based Structure

User uploads sensing data and Server pays user - transaction is verified by miners and stored on P2P network

2) Novel incentive mechanism

Server releases sensing task with evaluation criteria for data quality, and prepays a deposit. Miners verify quality of the sensing data that users submit using an evaluation function provided by the server.

3) K-Anonymity Privacy

Makes users' privacy to be hidden in group to deal with the impersonation attacks in the open and transparent blockchain.

Takeaways

Decentralized Crowdsensing

A distributed incentive mechanism such as Blockchain, is a viable option for data sensitive crowdsensing applications that typically utilize a central authority.

Privacy Preservation

K-anonymity in blockchain crowdsensing guarantees that Individuals who are the subjects of released data cannot be re-identified while the data remains practically useful.

Data Evaluation

Using an evaluation function, Miners can verify transactions based on predetermined criteria to pay users accordingly and to also promote high quality data submission.

CONCLUSION

- **IPFS:** version history & immutability for web files, decentralized file storage, decentralized file naming system, great for “blob storage”
- **Blockchain Incentives:** miners run “evaluation function” on crowdsourced data to determine data’s quality; server pays amount proportional to data quality
- How WitnessChain could benefit
  - **IPFS**: potential extension to project to enable full decentralization
  - **Blockchain Incentives**: could adopt “evaluation function” to ensure citizen payment is proportional to quality of submitted evidence

Paper 1 Questions

1. Edge cases where IPFS is problematic?
    a. Nontrivial price for uploaded file; expensive! Also, all IPFS files are public; we need some to stay private.
2. IPFS vs BitTorrent? Latency comparison?
    a. BitTorrent: file streaming. IPFS: file storage. IPFS doesn't need special software, so less latency.
3. What if child porn is stored on IPFS? How does BitTorrent handle it?
    a. IPFS will implement central authority with power to take down bad content. BitTorrent is a protocol, so cops just take down server hosting content.

Paper 1 Questions, Continued

1. IPFS vs Filecoin? How do you incentivize IPFS?
   a. Filecoin is built on IPFS. Filecoin pays nodes who host files, incentivizing nodes to participate in IPFS.
2. How is multihash format used?
   a. Each node can define & use its own hash function, increasing versatility. E.g. you can choose between higher security or faster performance.
3. Does anyone else use IPFS's multiaddr format?
   a. BitTorrent uses similar format. Multiaddr combines addresses at each layer (TCP, IPv4, IPv6, etc) in one standardized package.
   b. Multiaddr format = (/<addr str code>/<addr str rep>)+
      /ip4/<ipv4 str addr>/udp/<udp int port>
      /ip6/<ipv6 str addr>/tcp/<tcp int port>

Paper 2 Questions

1. How do you judge content quality?
    a. We'll judge this based on the police response, whether the evidence leads them to issue a ticket or not.
2. Can the privacy maintaining algorithm be simplified for implementation?
    a. We only ever store the user's wallet address on our blockchain. We do store the user email and keys, but these are hashed on our internal server so that we can't see that information.
3. Provide an example of low/high quality data. How do the miners judge data quality and pay accordingly?
    a. See "Blockchain Incentives" paper. Server publicizes criteria for high-quality images (e.g. license plate number shown, well-lit scene) and gives an evaluation function that miners can call to estimate the quality of the image. Miners can run the evaluation function to decide if the image is high quality.

Paper 2 Questions, Continued

1. Why is K-anonymity relevant in crowdsensing?
   a. There are K people in a group that could be identified with a certain identifier. Dhruv is in Latanya Sweeney's class right now who invented this. In this app, users are lumped into groups with similar characteristics, so you can't identify which specific user you are referring to.
2. How does storing data in a Merkle tree make verification of the data easier?
   a. Merkle trees are great for verification because they store hashes, not original files (which are much bigger). That's especially important on decentralized systems. They allow for data synchronization and are consistent.
3. Why is it beneficial to interact with a group of users instead of a single user?
   a. Again, this is useful for K-anonymity so that you can use the group of users to hide the individual user.

THANK YOU