Blockchain-Based IoT Digital Forensics

Review of An Enhanced Blockchain-Based IoT DF Architecture Using Fuzzy Hash

Luka Boljević, Teodor Janez Podobnik, Hana Zlobec

June 23, 2022

Outline

- Introduction
- Framework architecture
- IPFS
  - IPNS
  - Pros/Cons
- Smart Contract
  - Solidity code
- Fuzzy hashing algorithms analysis
  - Text analysis
  - Image analysis
- Conclusions

Introduction

IoT ("The **I**nternet **o**f **T**hings") is a network of physical objects with:

- sensors,
- processing ability,
- ability to exchange data with other devices without human interaction

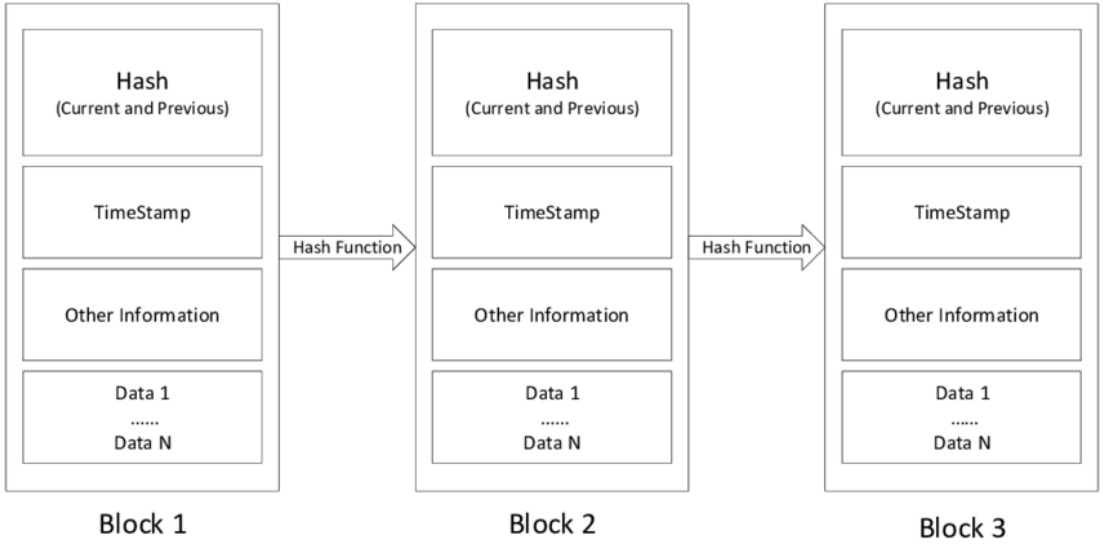
Examples: surveillance cameras, alarm systems with motion and door contact sensors, CO and smoke detectors, smart car systems...

Introduction

- Usage of IoT devices is increasing
- Data they collect can be useful for forensic investigations
- There is no standard for IoT-based forensic procedures

Blockchain

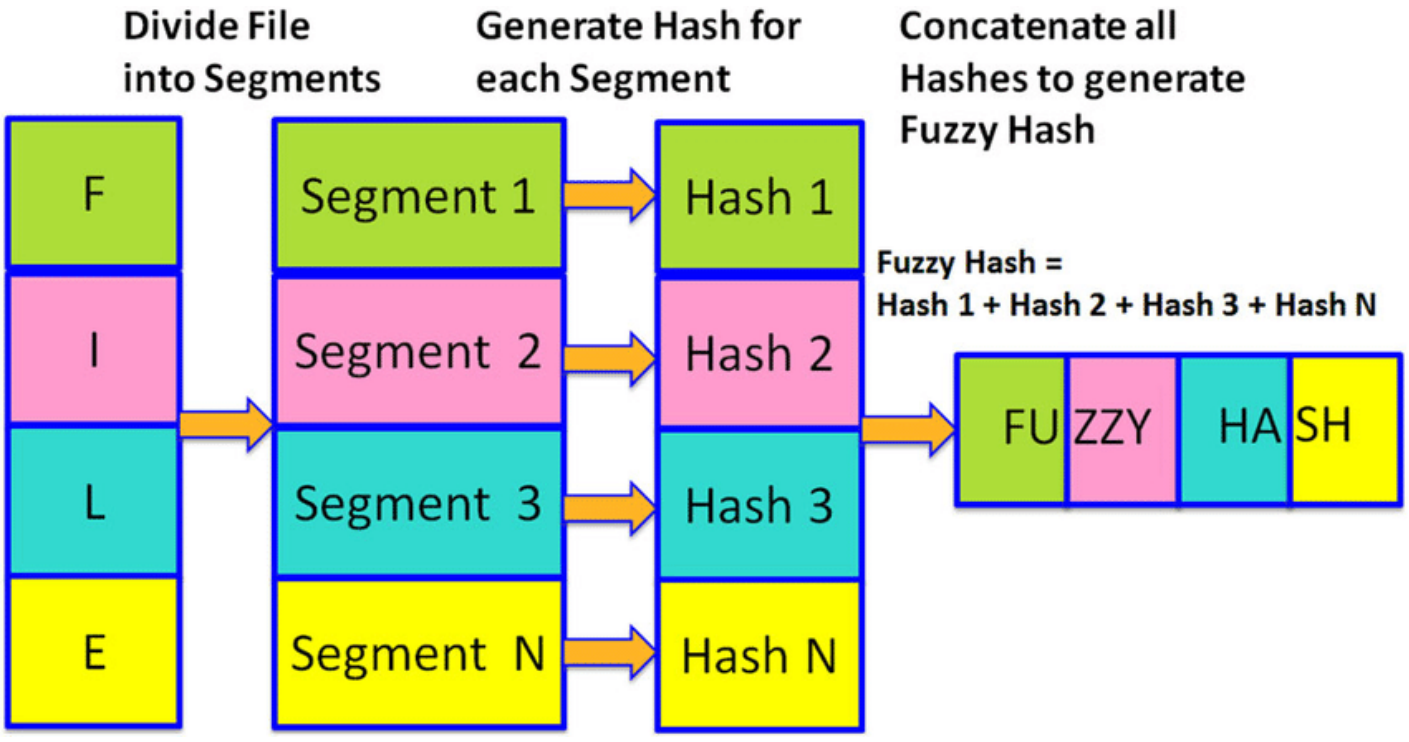
Blockchain = a chain of immutable blocks



- Increasingly popular in forensics
- No central authority needed for verification
- Resistant to alterations
- Internal information visible to all participants
- Anyone can verify the information

Fuzzy hash

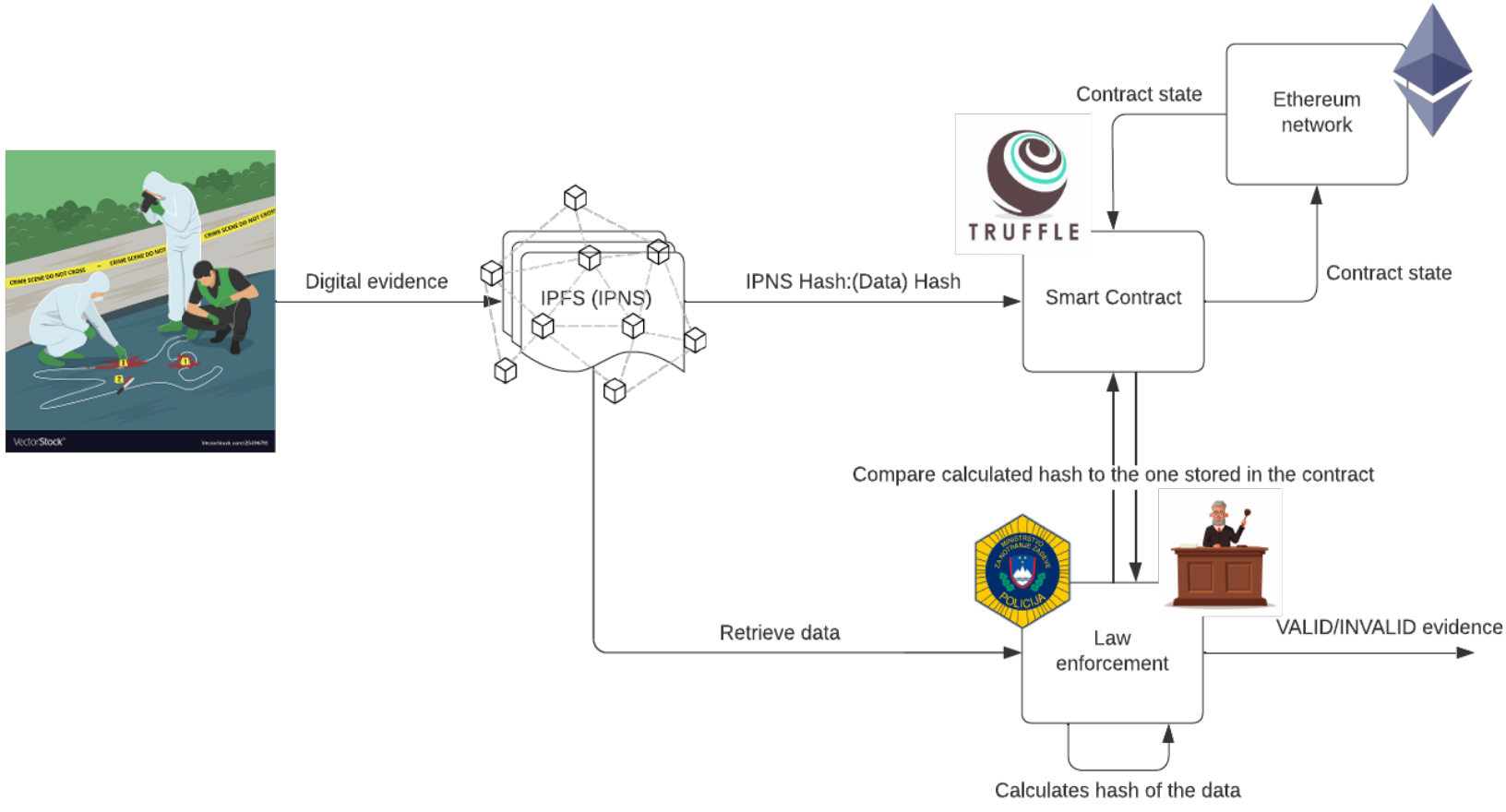
- Traditional hashing techniques ignore similarity
- Fuzzy hash does not
- Allows to look for related evidence



Problems

- Problems with IoT:
  - Variety of IoT devices
  - IoT resource constraints
- Problems with blockchain:
  - Lightweight PoW algorithm needed

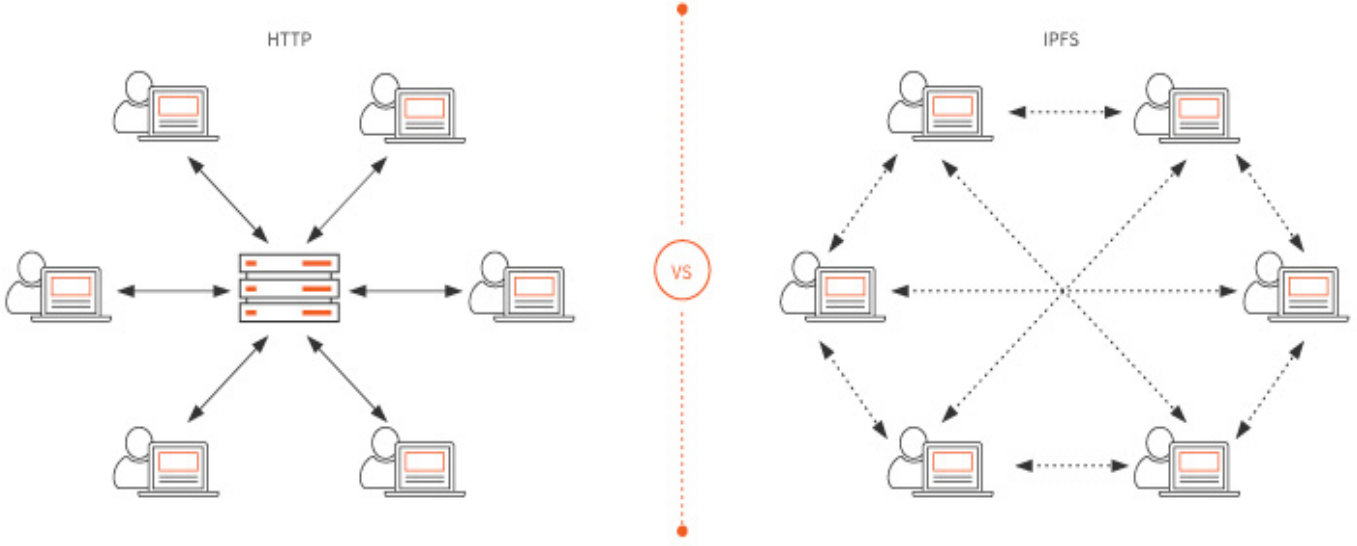
Proposed approach architecture



Description of the IPFS

IPFS stands for Interplanetary File System

- Decentralized database
- Content based addressing (CID "hash")



Description of the IPNS

IPNS stands for InterPlanetary Name System and it allows us to reference various CIDs using only one hash/address. In our particular work we use it to:

- uniquely identify either original or last relevant evidence record modification
- group each separate evidence records in the Smart contract

Pros/Cons of IPFS

Pros:

- No central authority
- Transparency
- No storage capacity limit
- Higher upload bandwidth (several sources)

Cons:

- Pinning problem (immutability)
- Requires separate PK network for data encryption

Local network, yes/no? (Power distribution in Blockchain, 51 percent attack, DDOS)

Description of the Smart Contract

Smart contract is sort of like a VM running on Ethereum network with features like:

- Records stored in a decentralized fashion
- Transaction transparency
- Immutable past records states
- Changes state using transactions
- Cons: slow and costly (data digest)

Description of the Smart Contract


```
// SPDX-License-Identifier: MIT
pragma solidity >=0.4.22 <0.9.0;

contract HashRecord {
   // IPNS hash: string
   // Fuzzy hashes: string[]
   mapping(string => string[]) public fuzzyHashes;

   // Add modifier for access restriction

   // Store fuzzy hash mapped to IPNS hash
   function setFuzzyHash(string memory fuzzyHash, string memory ipnsHash) public {
      fuzzyHashes[ipnsHash].push(fuzzyHash);
   }

   // Get fuzzy hash using IPNS hash - user should make sure matching fuzzy hash exists for the given IPNS hash
   function getFuzzyHashes(string memory ipnsHash) public view returns (string[] memory) {
      return fuzzyHashes[ipnsHash];
   }

   // Delete hashes stored on particular IPNS hash
   function deleteHashes(string memory ipnsHash) public {
      delete fuzzyHashes[ipnsHash];
   }
}
```

Fuzzy hashing algorithms analysis

3 hashing algorithms to compare:

- `ssdeep`
- `LZJD` (Lempel-Ziv Jaccard Distance)
- `TLSH` (Trend Micro Locality Sensitive Hash)

Text analysis

Two pieces of text with only one difference (here shortened to fit):

- ... we have already given them a special name: **building** blocks. Just as a child ...
- ... we have already given them a special name: **Building** blocks. Just as a child ...

Text analysis results


```
TEXT COMPARISON:
TLSH hash comparison(the lower the number, the better):
difference = 84

SSDeep hash comparison(the higher the number the better):
similarity = 97

LZJD hash comparison(the higher the number the better):
similarity = 0
```

Figure: Results return by script

Image analysis

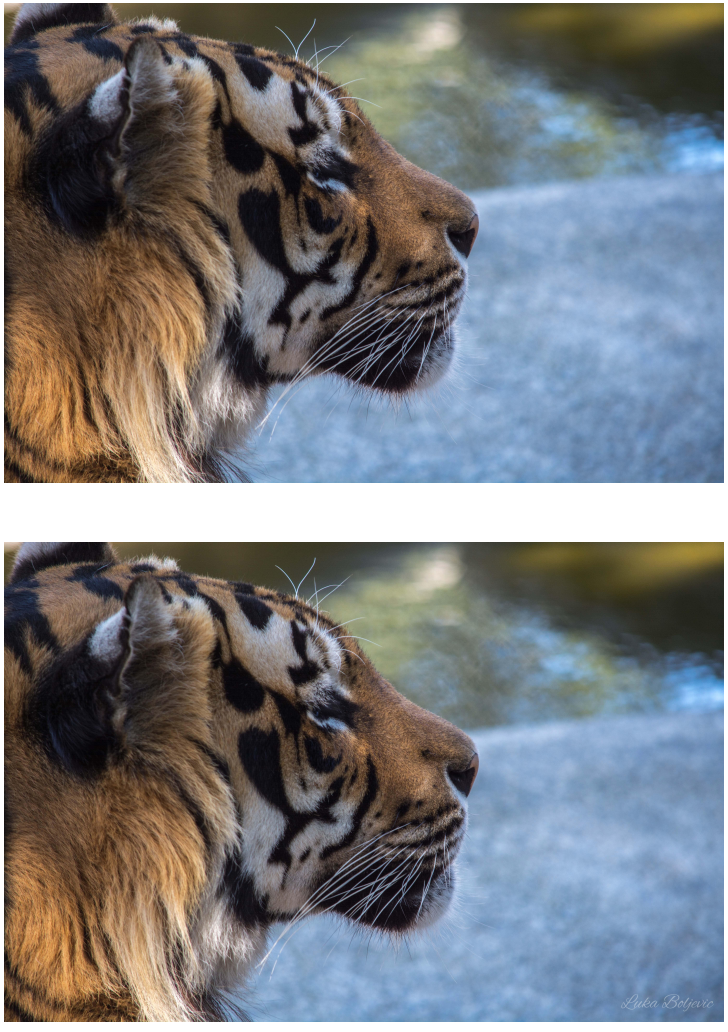


Figure: (Top) Non watermarked/Original; (Bottom) Watermarked

Image analysis results


```
IMAGE COMPARISON:
TLSH hash comparison(the lower the number, the better):
difference = 199

SSDeep hash comparison(the higher the number the better):
similarity = 0

LZJD hash comparison(the higher the number the better):
similarity = 0
```

Figure: Results return by script

Conclusions

- The proposed approach is very adequate for preservation of data integrity
- Fuzzy hashing is much more flexible than regular hashing
- Blockchain-based digital forensics architecture - the future?

The end

Thank you for your attention!

Questions?