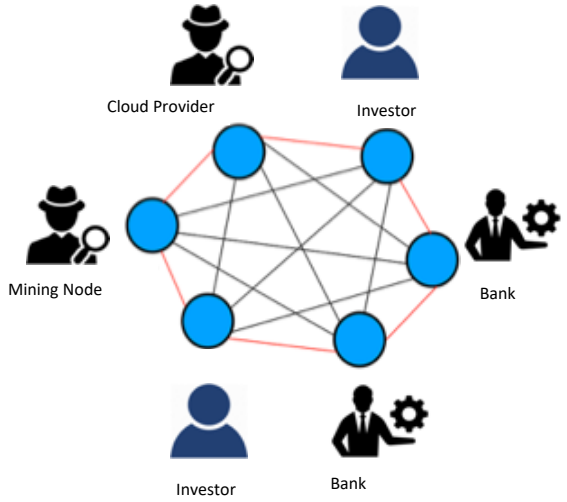


Probabilistic Blockchains: A Blockchain Paradigm for Collaborative Decision-Making

Tara Salman, Raj Jain, and Lav Gupta

{tara.salman, jain, lavgupta}@wustl.edu

Tara Salman, Raj Jain, and Lav Gupta, "**Probabilistic Blockchains: A Blockchain Paradigm for Collaborative Decision-Making**," 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON 2018), New York, NY, November 8-10, 2018, pp.

Problem Statement

- **Problem:** How to use the blockchain technology to make probabilistic, collaborative, and consensus decisions
  - Risk assessment

- Risk assessment applications
  - Networking: intrusion or malware detection where multiple inspectors inspect the same file/URL/flow. One consensus decision is returned
  - Financial: stock market prediction where investors share their predictions and the system achieve a consensus.
  - Others: recommendation system, voting system

Our Paper

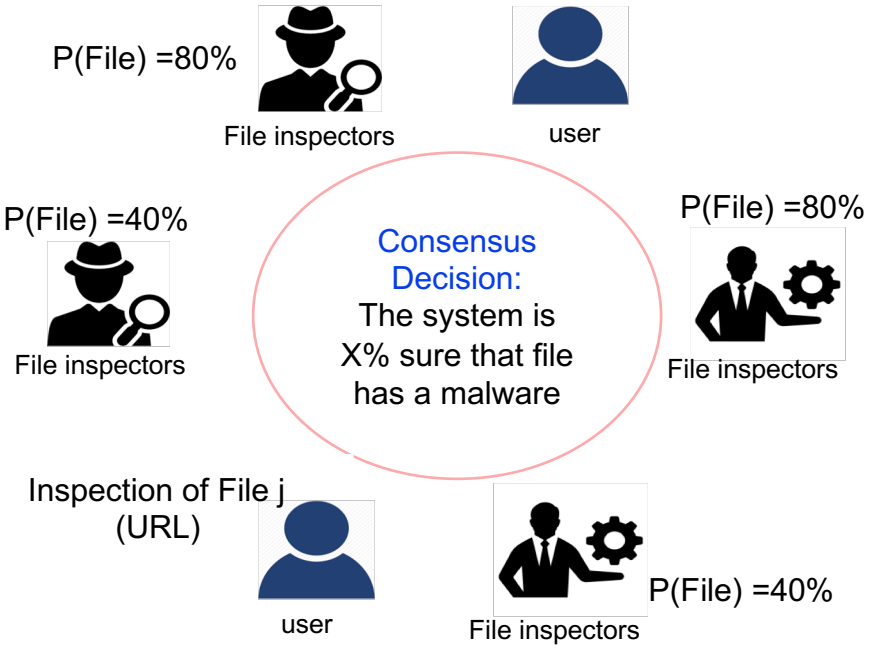
- Main contribution
  - Probabilistic blockchain: Extend the blockchain to make collaborative consensus decisions for risk assessment
- This presentation is about the probabilistic blockchain to suite consensus decision making
  - Why and how?
  - Probabilistic blockchain architecture, process
- Initial work on intrusion detection as a case study

Outline

- Consensus decision making
- How is it done today
- Blockchain-based decision making
- Probabilistic blockchain
- Case-Study: blockchain-based IDS

Consensus Decision Making

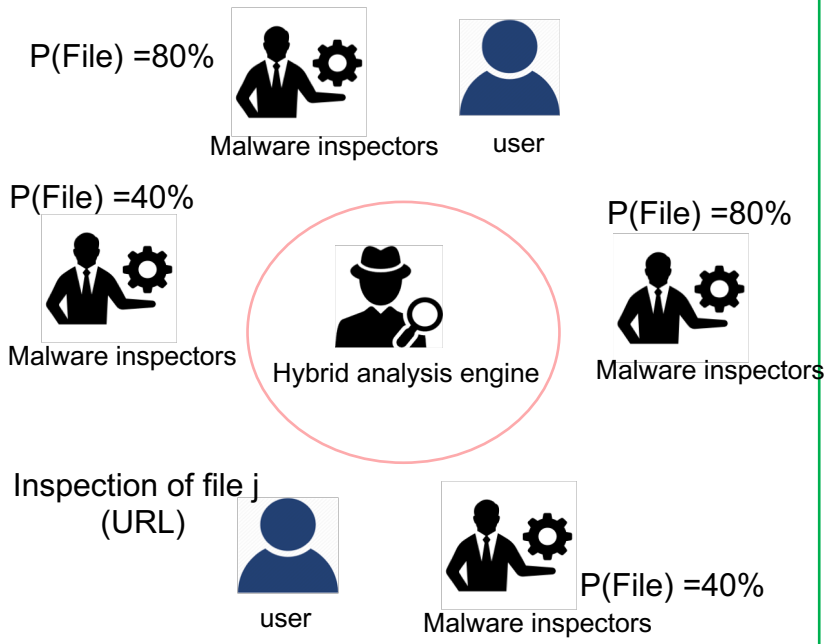
- Consensus decision-making process:
  - Achieve one decision even if group members disagree
- Example: Malware detection system
  - A network consists of distributed malware detectors and users
  - A user wants to download a file so it requests an inspection (providing a URL)
  - Multiple inspectors check the file and share their decisions
  - The system reaches a consensus decision about that file



A collaborative malware detection example

How is it Done Today

- Centralized solutions
  - One central party makes decisions
- Example Malware detection
  - Hybrid analysis engine *
  - Given a file or URL from a user, different inspectors inspect and return their results
  - Hybrid analysis analyze the results and return a consensus decision
- \+ Less communication overhead, sometimes faster



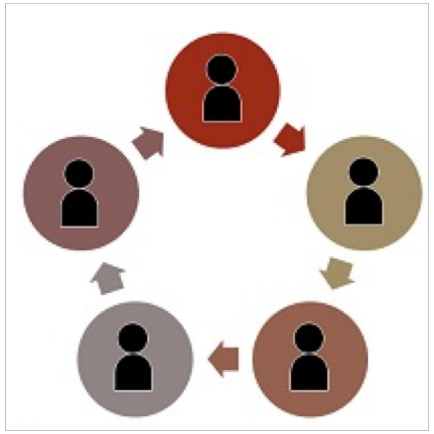
Malware detection example

* Hybrid Analysis website, https://www.hybrid-analysis.com/

Centralized vs. Distribution

Centralized solutions

1. One central decision
2. Single point of failure
3. Easier to hack



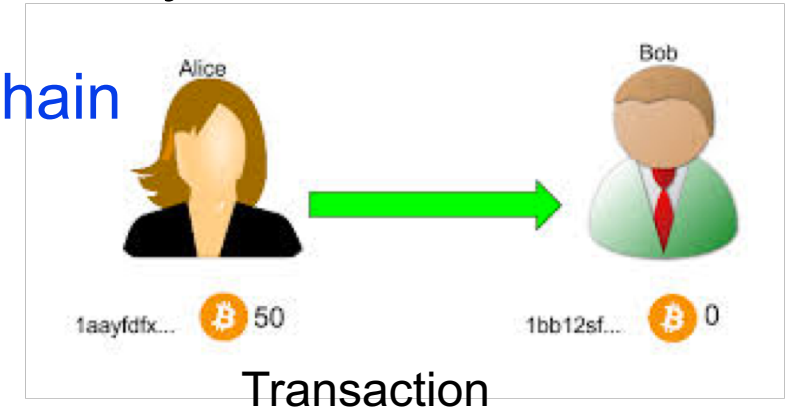
Distributed solutions (Peer to peer network)

1. Collaborative decisions
2. No single point of failure
3. Very difficult to hack

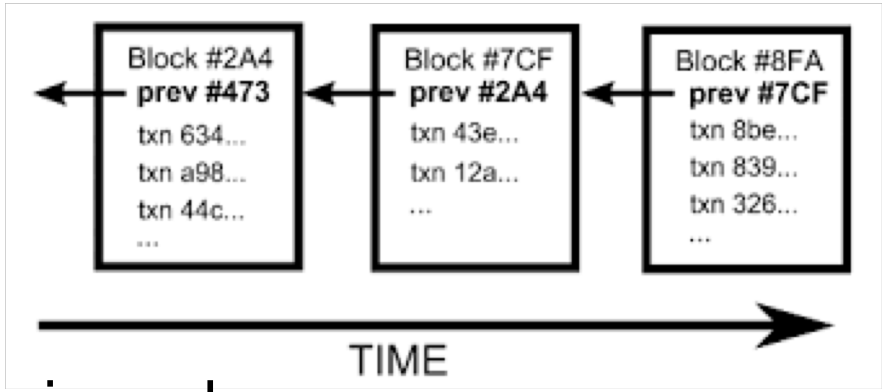
Blockchain technology is one way to distribute

Blockchain Technology

- Peer to peer, distributed network with no central authority
- Consist of transactions, blocks, and a distributed chain
- A transaction is any type of interaction
  - e.g. Alice sends money to Bob
  - Each transaction is signed by the source
- A block is a set of transactions that are signed
- A chain is a set of linked blocks
  - Block linked by the ID of the previous block
- Chain is globally distributed at different blockchain nodes



Transaction



Blockchain

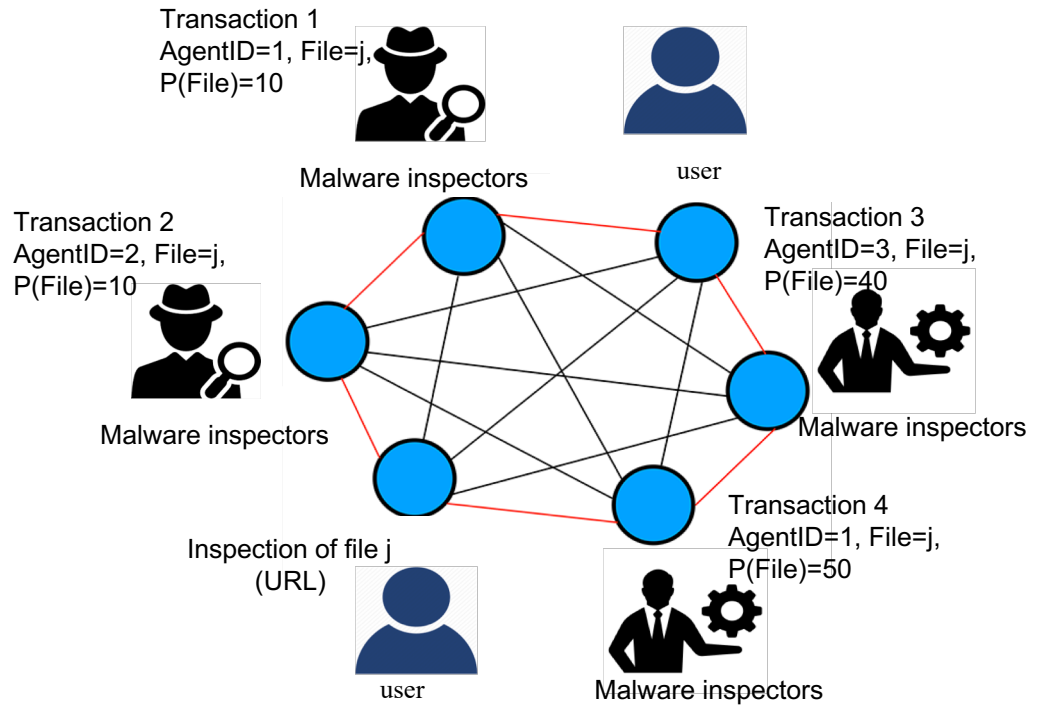
Properties of Blockchains

1. Distributed nature: Not easy to hack
2. Decentralized consensus: No single decision point
3. Cryptographically secure: Transactions and blocks signatures, can't deny interactions in the system

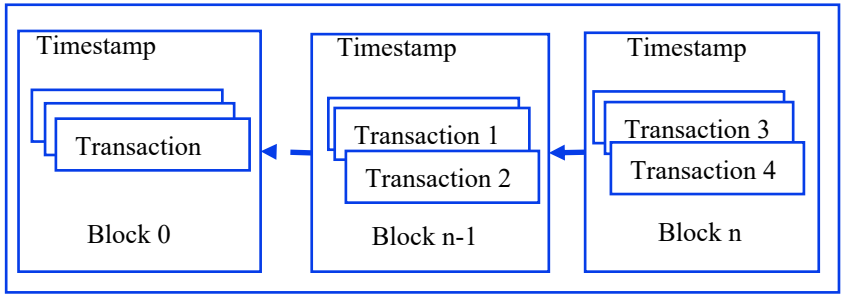
Blockchain-based Decision Making

- Use transactions to store probabilistic decisions

1. A user consults system for a decision (e.g. URL of a file)
2. Decision makers put there decisions in transactions



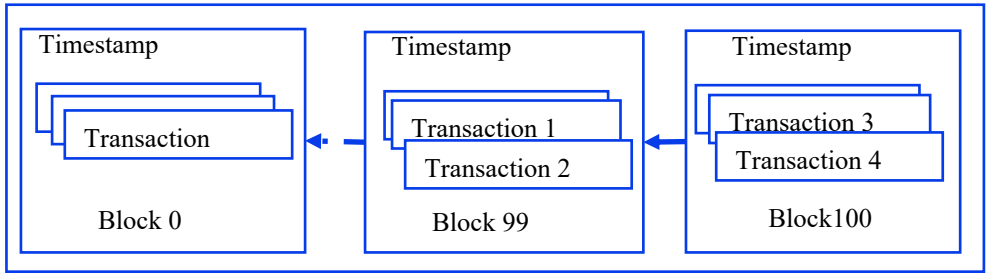
- Decisions are distributed and available to everyone
- No consensus decision available



The Challenge

- **Consensus is not yet distributed**

- The decision making process is still to be processed
  - Off chain → Third party trust
  - Made locally → no global decision
- Further, system decision makers
  - Look for decisions of same events
  - Need to process the whole chain → computationally expensive

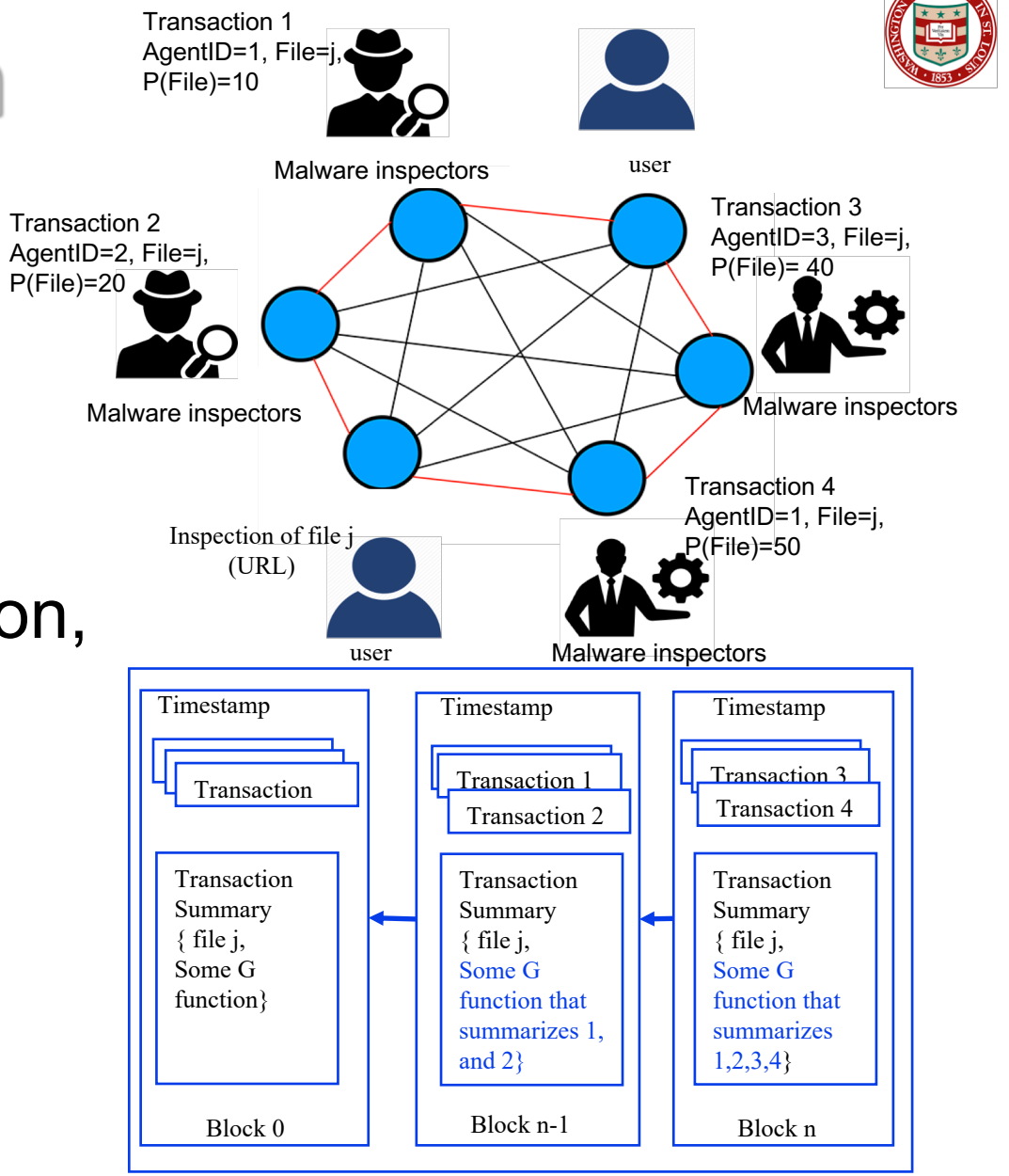


Probabilistic Blockchain

- Idea: Achieve consensus within the blockchain network

- Consensus Function (G function)
  - Summarizes events
  - $G(P_j(event_i))$, P_j individual decision,
  - Calculated within the block
  - Include enclosed events

- Decisions made in blocks' header
  - Consensus decisions available
  - Easy access



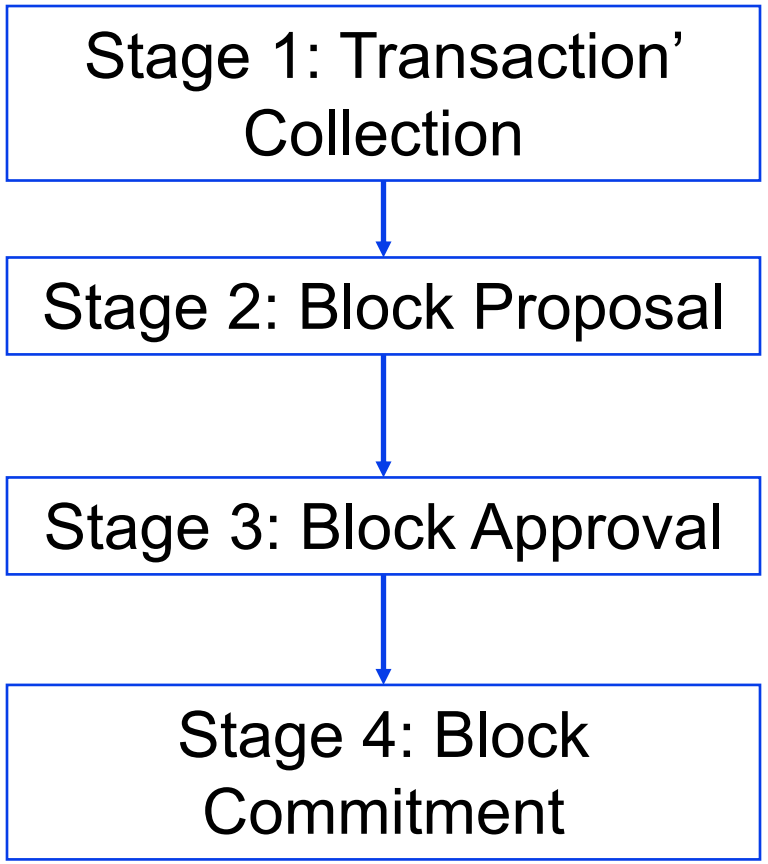
Consensus Function Conditions

- Easy to calculate and validate
- Not easy to manipulate
  - Manipulate few decisions does not change the consensus interpretation (Bad functions: min or max)
- Incremental
  - Will be needed if transactions are coming over different blocks
  - Can be calculated from current transactions and prior summary
  - Bad functions: median or mode percentiles (unless you use a incremental algorithm like P-square*)
- Satisfying function: [1st moment, number of decisions]

* Raj Jain and I. Chlamtac, "**The P-Square Algorithm for Dynamic Calculation of Percentiles and Histograms without Storing Observations**," Communications of the ACM, October 1985, pp. 1076-1085, http://www.cse.wustl.edu/~jain/papers/psqr.htm

Blockchain Workflow

- Transaction' collection: Collect decisions (transactions) from different decision makers
- Block proposal: Build a block that summarizes the collected decisions (multiple events)
- Block approval: Check that the block is correct
- Block commitment: Attach the block to the chain



Case-Study: Blockchain-based IDS

- Different machine learning trained intrusion detectors make decisions about a certain flow
- The consensus function used is the mean

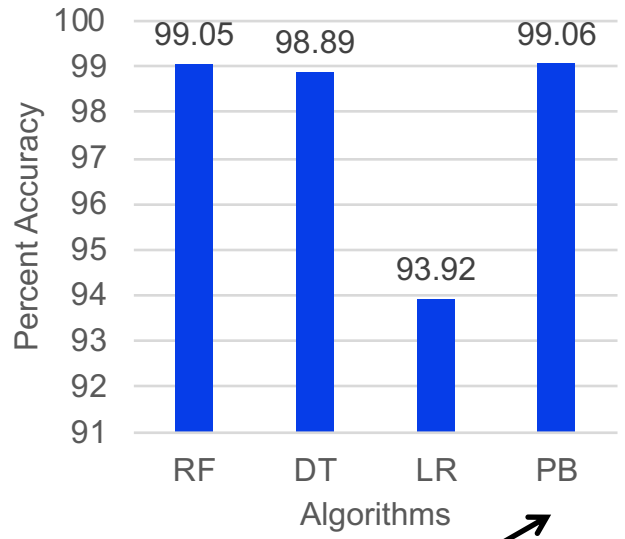
$$G = P\ (Flow_i\ is\ malicious) = \frac{1}{m}\sum P_j\ (Flow_i\ is\ malicious)$$

P_j individual decision, P consensus decision

- Two experimental settings
  - Different training machine learning algorithms
  - Different training datasets
- Accuracy is the evaluation metric used in both cases

Different Machine Learning Algorithms

Attack type	DoS attack taken from UNSW-NB15 dataset
Number of algorithms	Random Forest (RF) Decision Tree (DT) Linear Regression (LR)
Number of detection agents	1000
Training sample size	~ 57000
Test sample size	~ 46000



Probabilistic Blockchain
(collaborative decision instead of one)

Conclusion and Future Work

- Current blockchain technology cannot be used to decision making in risk assessment application
- We propose probabilistic blockchains, a blockchain extension to suite consensus decision making
- A case study of blockchain-based showed the feasibility of the proposed approach
- Future work:
  - Better consensus functions
  - More applications
  - Practical evaluations

Questions ?