Literature survey - literature

September 15, 2022

1 Literature

1.1 Existing mechanisms for trust

ConTrib (entity)

Url: https://pure.tudelft.nl/ws/portalfiles/portal/89353583/1_s2.0_S1389128621001705_main.pdf
Title: ConTrib: Maintaining fairness in decentralized big tech alternatives by accounting work
Author: Martijn de vos and Johan Pouwelse
Content: A universal mechanism to maintain fairness in decentralized applications by accounting the work performed by peers. There currently is no universal accounting mechanism that can be used to address fairness issues in decentralized applications, to the best of our knowledge. We refer to the illegitimate modification of a record as fraud. To detect fraud, peers continuously request random records from other peers and disseminate newly created records in the network. Peers verify the consistency of incoming records with the ones stored in their database. Everyone has a personal ledger which works like a hashchain (contains information to previous record for example), has a sequence number. Inconsistency detected in algo 3 on line 16. ConTrib can detect fraud from up to 2 conspiring nodes in the network.
Is this too experimental?

DART: A DISTRIBUTED ANALYSIS OF REPUTATION AND TRUST FRAMEWORK (Entity?)

Url: https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-8640.2012.00453.x
Title: DART: A DISTRIBUTED ANALYSIS OF REPUTATION AND TRUST FRAMEWORK
Content: This paper introduces a Distributed Analysis of Reputation and Trust (DART) framework. The environment of DART is decentralized and game-theoretic. Not only is the proposed environment model compatible with the characteristics of open distributed systems, but it also allows agents to have different types of interactions in this environment model. Besides direct, witness,

and introduction interactions, agents in our environment model can have a type of interaction called a reporting interaction, which represents a decentralized reporting mechanism in distributed environments. The proposed environment model provides various metrics at both micro and macro levels for analyzing the implemented trust and reputation models. Using DART, researchers have empirically demonstrated the vulnerability of well-known trust models against both individual and group attacks.
NOTE: This framework is used to test and analyse the trust/reputation models/algorithms of researchers by enabling one to model distributed systems and the interactions which these entail.

Previous thesis on very similar subject where trust is calculated through the public records of interactions and their verification (entity)

Url: `https://github.com/Tribler/tribler/issues/3357` Title: Creating trust through verification of interaction records Content: Trust on the internet is largely facilitated by reputation systems on centralized online platforms. However reports of data breaches and privacy issues on such platforms are getting more frequent. We argue that only a decentralized trust system can enable a privacy-driven and fair future of the online economy. This requires a scalable system to record interactions and ensure the dissemination and consistency of records. We propose a mechanism that incentivizes agents to broadcast and verify each others interaction records. The underlying architecture is TrustChain, a pairwise ledger designed for scalable recording transactions. In TrustChain each node records their transactions on a personal ledger. We extend this ledger with the recording of block exchanges. By making past information exchanges transparent to other agents the knowledge state of each agent is public. This allows to discriminate based on the exchange behavior of agents. Also, it leads agents to verify potential part- ners as transactions with knowingly malicious users leads to proof-of-fraud. We formally analyze the recording of exchanges and show that free-riding nodes that do not exchange or verify can be detected. The results are confirmed with experiments on an open-source implementation that we provide.
Thoughts: This only works if a trust measure can be calculated from the detection of fraud in one's records. How do we assign trust to information? How to say what is true? No records here.
TUD repo:
`https://repository.tudelft.nl/islandora/object/uuid\%3A4716c3f8-b9b1-4e80-8537-10b006bb7`

In tags we trust (Entity and content)

Url: `https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6153150`
Title: In tags we trust
Content: This is an article which summarises multiple approaches which have been suggested for tackling the problem of trust by tagging.

NetFlow and Temporal PageRank algorithms (entity)

Url: file:///home/thomas/Downloads/thesis\%20Pim\%20Otte\%20(1).pdf
Title: Sybil-resistant trust mechanisms in distributed systems
Content: Is about reputation of agents, not information. May however look further into it to assess reusability?

Trellis: Actually collects extensive user feedback to estimate trustworthiness of resources (content, user-feedback-based)

Url: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.9628&rep=rep1&type=pdf
Title: Trusting Information Sources One Citizen at a Time
Content: Allows users to create annotations on resources. Requires A LOT of user input/effort. As users add annotations, they can include measures of Credibility and Reliability about a statement, which are later averaged and presented to the viewer. Using the TRELLIS system, users can view information, annotations (including averages of credibility, reliability, and other ratings), and then make an analysis.

Trust metric (entity)

Url: https://www.usenix.org/legacy/publications/library/proceedings/sec98/full_papers/levien/levien.pdf
Title: Attack-resistant trust metrics for public key certification
Content: trust metric uses group assertions for determining membership within a group. The Advogato website, for example, certifies users at three levels. Access to post and edit website information is controlled by these certifications. On any network, the Advogato trust metric is extremely attack resistant. By identifying individual nodes as "bad" and finding any nodes that certify the "bad" nodes, the metric cuts out an unreliable portion of the network. Calculations are based primarily on the good nodes, so the network as a whole remains secure.

1.2 Sybil attacks defense mechanisms

Defensive method against sybil attacks (nodes required to perform puzzles periodically)

Url: https://dl.acm.org/doi/10.1145/2382536.2382548
Title: SybilControl: practical sybil defense with computational puzzles
Content: TODO

Defensive method agains sybil attacks (log analysis)

Url: https://ieeexplore.ieee.org/document/6257993
Title:
Content:
TODO

Defensive method agains sybil attacks (log analysis) part 2

Url: https://ieeexplore.ieee.org/document/4531141
Title:
Content:
TODO

Sybil attacks

Url: https://link.springer.com/chapter/10.1007/3-540-45748-8_24
Title: The Sybil Attack
Content: Large-scale peer-to-peer systems face security threats from faulty or hostile remote computing elements. To resist these threats, many such systems employ redundancy. However, if a single faulty entity can present multiple identities, it can control a substantial fraction of the system, thereby undermining this redundancy. One approach to preventing these ìSybil attacksî is to have a trusted agency certify identities. This paper shows that, without a logically centralized authority, Sybil attacks are always possible except under extreme and unrealistic assumptions of resource parity and coordination among entities.

1.3 Trust

Pioneering paper on trust and reputation (2001)

Url: https://sci-hub.se/10.1109/HICSS.2002.994181
Title: A Computational Model of Trust and Reputation
Content:

More blockchain-based reputation model fro ensuring trust

Url: https://ieeexplore.ieee.org/abstract/document/9312998
Title: Blockchain-based distributed reputation model for ensuring trust in mobile adhoc networks
Content:

A general overview of trust in distributed systems

Url: https://sci-hub.se/10.1007/s42452-019-1598-6
Title: A glimpse on Semantic web trust
Content: Explains nicely why trust is needed and how it could be modeled.

First way of modelling trust which also incorporates distrust

Url: https://dl.acm.org/doi/abs/10.1145/988672.988727 Title:
Content:

1.4 Spread of misinformation

Detecting fake news method

Url: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8768083
Title: Detecting Fake News Over Online Social Media via Domain Reputations and Content Understanding
Content:

Using blockchain to monitor the spread of fake news

Url: https://ieeexplore.ieee.org/abstract/document/8764081
Title: Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News
Content:

how do users identify fake news?

Url: https://www.emerald.com/insight/content/doi/10.1108/OIR-08-2020-0333/full/html
Title: An empirical approach to understanding users' fake news identification on social media
Content:

1.5 Content to trustworthiness

A method to model the content of information to trustworthiness

Url: https://sci-hub.se/10.1145/1135777.1135861
Title: Towards Content Trust of Web Resources
Content: States the factors which influence user's trust. Created several models and calculations for user trust in web resources for specific queries. Makes a lot of assumptions. Prone to sybil-like attacks. Can't be used in it's current form

as it requires users to provide a function.

System which helps users to determine trustworthiness for themselves

Url: https://link.springer.com/chapter/10.1007/978-3-540-74851-9_4
Title: Trustworthiness Analysis of Web Search Results
Content: When searching for information, it will show how much a certain page relates to other pages by detecting their topics.
Note: This requires a central authority to a topic's presence on "the web". OR can be executed by a distributed client in an infeasible amount of time.

A way to propagate content-based trust scores

Url: https://sci-hub.se/10.1145/2020408.2020567
Title:
Content:

TODO

Url: https://www.sciencedirect.com/science/article/pii/S1877050916305476
Title: Trust necessitated through Metrics: Estimating the trustworthiness of websites
Content:

Web 2.0 method to assess trustworthiness

Url: https://sci-hub.se/10.1186/2196-064X-1-5
Title: Two sides of the coin: measuring and communicating the trustworthiness of online information
Content: assess information trustworthiness in three domains: the provenance of information, its quality, and the integrity of the information infrastructure used to communicate the content from author to final consumer. Calculate value for each of these and then apply a formula to calculate the end trustworthiness value.

Url: https://sci-hub.se/10.1007/11890850_12
Title: Combining Provenance with Trust in Social Networks for Semantic Web Content Filtering
Content: we present a two level approach to integrating trust, provenance, and annotations in Semantic Web systems. We describe an algorithm for inferring

trust relationships using provenance information and trust annotations in Semantic Web-based social networks. Then, we present an application, FilmTrust, that combines the computed trust values with the provenance of other annotations to personalize the website. The FilmTrust system uses trust to compute personalized recommended movie ratings and to order reviews.
Note: this article contains a very interesting method of inferring trust based on paths

1.6 Trust graphs

Trust networks in semantic web

Url: `10.1007/978-3-540-45217-1_18`
Title: Trust Networks on the Semantic Web
Content: Basic explanation/implementation of trust graphs on the semantic web
Interesting: At runtime, and before joining an IRC network, TrustBot builds an internal representation of the trust network from a collection of distributed sources. Users can add their own URIs to the bot at any time, and incorporate the data into the graph. The bot keeps a collection of these URIs that are spidered when the bot is launched or called upon to reload the graph. From an IRC channel, the bot can be queried to provide the weighted average, as well as maximum and minimum path lengths, and maximum and minimum capacity paths. The TrustBot is currently running on icr.freenode.net, and can be queried under the nick 'TrustBot'.

Using trust graph to assign trustworthiness values to all users in graph

Url: `https://sci-hub.se/10.1007/978-3-540-39718-2_23`
Title:Trust Management for the Semantic Web
Content: Very mathematical approach to calculate trustworthiness of all users based on a very limited set of directed trust edges provided by the user. Each user has their own set of trustworthiness values for each other entity.

1.7 Other

Economic point of view

Url: `http://users.eecs.northwestern.edu/~hxb0652/HaitaoXu_files/TWEB2017.pdf`
Title: An Empirical Investigation of Ecommerce-Reputation-Escalation-as-a-Service
Content:

What is spam to one person, may be interesting to another

Url: `http://airweb.cse.lehigh.edu/2009/papers/p41-markines.pdf`
Title: Social spam detection.
Content:

Peer2peer has been a projected issue since at least 2001

Url: `http://robotics.stanford.edu/~kevinlb/ec01-short.pdf`
Title: Incentives for Sharing in Peer-to-Peer Networks
Content:
TODO

2 Still to read

Methods of assigning trust values to content:
https://sci-hub.se/10.1007/978-81-322-3592-7_18 (not trustworthy paper?)
https://dl.acm.org/doi/abs/10.1145/1013367.1013409 More needed?
TrustGraph https://dl.acm.org/doi/abs/10.1145/988672.988727 https://www.sciencedirect.com/science
Probably not: https://www.sciencedirect.com/science/article/pii/S1568494622000357?
Read articles from alexander stannat: https://github.com/Tribler/tribler/issues/4481#issuecomment-493429179

3 Ideas/brainstorming

logging resources access

When users access a resource, they connect with the server and both log that they have accessed this resource.

- Popular resources more trustworthy?
- Would be very subsceptible to sybil attacks.
- People that are close to you in the trustgraph more trustworthy?

Trusted nodes

Similarly to XRP Ledger, there exist a number of nodes which are trusted. These nodes can be accessed by individuals to check resource trustworthiness for a given query.

- How do trusted nodes know who to trust?
- What's in it for the trusted nodes? There must be an incentive for individuals to run trusted nodes without a powerful third party.

- What if a user is deliberately searching for false information? Will those results be censored from the result set? Or will they still appear below the trusted information.
- What if a user is searching for information not contained by the trusted nodes?

Timing access

One method (find source) to measure trustworthiness, is by measuring time spent accessing certain resources. For a given query, trustworthy resources will be accessed for a longer amount of time than untrustworthy resources.

- Connect trustworthiness value to search query, such that people looking for untrustworthy information will still be able to find it, as it may be considered trustworthy for a given query.
- How to prevent automated attacks? We need to be able to distinguish between user and bot, which is an entirely different problem altogether. Combine thesis with this?
- it may be different to store logs of access times than to store logs of interactions
- may be used to build a trust graph?

Calculating a trust score for content based on a formula

The formula could consist of:

- Distance in trust graph
- Trustworthiness of the author (out of scope/other mechanisms)
- Content similarity with other results for a given query?
- Average time spent accessing that content (need to find a way to indicate end time)

Other points:

- Could we dynamically change one's position in the trust graph or one trustworthiness by slightly increasing their reputation when accessing the given resource.

Using links between contents

This method has been explored where trustworthy articles only point to other trustworthy articles.