Peer to Peer Degrees of Trust

a white paper from Rebooting the Web of Trust VII

by Harrison Stahl, Titus Capilnean, Peter Snyder, and Tyler Yasaka

ABSTRACT

Authenticity is a challenge for any identity solution. In the physical world, at least in America, it is not difficult to change one's identity[1]. In the digital world, there is the problem of bots. The botnet detection market is expected to be worth over one billion USD by 2023[2], in a landscape where most digital activity is still heavily centralized. These centralized digital solutions have the advantage of being able to track IP addresses, request phone verification, and present CAPTCHAs to users in order to authenticate them. If this problem is so difficult to solve in the centralized world, how much more challenging will it be in the decentralized world, where none of these techniques are available?

In this paper, we explore the idea of using a *web of trust* as a tool to add authenticity to decentralized identifiers (DIDs). We define a framework for deriving relative *trust degrees* using a *given trust metric*: a "trustworthiness" score for a given identity from the perspective of another identity. It is our intent that this framework may be used as a starting point for an ongoing exploration of graph-based, decentralized trust. We believe this approach may ultimately be used as a foundation for decentralized reputation.

civic Protocol Labs sovrin

EXODUS™ htc interlinked connecting intelligence PTBVENTURES

TIERION Microsoft

Sponsors for the Rebooting the Web of Trust VII Design Workshop

[1] https://www.nytimes.com/2011/07/17/sunday-review/17disappear.html

[2] https://www.prnewswire.com/news-releases/botnet-detection-market-worth-11911-million-usd-by-2023-681515921.html

RELATED WORK ON GRAPH-BASED REPUTATION SYSTEMS

PGP

PGP (Pretty Good Privacy) is a system for asymmetric encryption. It was originally designed as a system for encrypting email contents, but it is also used for a variety of other file-encryption purposes.

In addition to defining a system for asymmetric message encryption, PGP also defines a system for key discovery, based off of key servers (which are conceptually distributed, but in practice use a small number of well-known and trusted servers), a method for assigning trust to keys that have been uploaded to these servers, and an algorithm for determining how much trust to assign to a key that was uploaded by a given email address, but which belongs to an account unfamiliar to the email sender. Collectively this system is called the "Web of Trust".

This system has many similarities to the kinds of systems envisioned in this paper. The design is fundamentally decentralized, it allows parties to reason about whether, and how much, to trust unfamiliar peers, and it allows participants to model trust in a non-binary way (i.e. users can be trusted "some", but not fully or not at all).

However, the PGP "Web of Trust" model differs from the model envisioned in this paper in several ways. First, while trust can be assigned in a non-binary manner, it can only be "derived" binary (i.e. a key is either trusted or untrusted, but the "Web of Trust" algorithm isn't intended to be used to provide finer grained trust information). Additionally, the PGP Web of Trust does not include a natural way to describe different domains of trust. One might wish to say "I trust party X very much when they're discussing music, but very little when they're discussing politics". The PGP model is not a natural fit for such statements.

Social Networks

Online social networks such as Facebook, Twitter and Instagram conceptually model their systems as graphs, with participants as nodes, and connections between the participants as edges with labels like "friend", "follower" or "boyfriend". These edges can be directed or bi-directional, depending on the relationship being represented.

The maintainers of these systems use the graph representation to make decisions about relationship, event or interest recommendations: a friend of your friend is more likely to also be your friend than a randomly selected person in the graph. While these systems are not primarily intended to be reputation or trust-management systems, the same algorithms and representations used in online social networks could guide the development of decentralized trust systems.

Recommendation Systems

Graph-based recommendation systems have been deployed for many commercial purposes. Examples include the video recommendation systems used in YouTube and Netflix and the product-recommendation systems used in sites like Amazon and eBay. In such systems, videos and products can be modeled as nodes. The users and customers of such sites are also represented as nodes. Recommendations are represented as edges from the latter to the former, possibly weighed for systems that allow non-binary edges (e.g. 1-5 star ratings). The structure of such graphs are used to recommend similar, popular products and videos to other users.

As in the "Social Networks" case, the same representations and techniques used in such systems can inform the design and algorithms used in a decentralized reputation / trust system.

DEGREES OF TRUST: A FRAMEWORK

A Relative Trust Score

Google's PageRank algorithm is famously able to assign ranking scores to web pages from a directed graph, where edges represent links from one website to another. One major problem with PageRank, however, is that it is vulnerable to Sybil attacks. Indeed, after Google implemented PageRank we witnessed the phenomenon of link farms: a form of Sybil attack that takes advantage of the fact that (in the pure form of PageRank) all pages are first-class citizens with equal ability to cast "votes" in the algorithm, in the form of web links. Dummy web pages could be created cheaply to deceive the algorithm. Of course, it is now widely known that Google uses a much more sophisticated version of PageRank that is less susceptible to link farms (though the algorithm itself is kept secret).

In many respects, deriving trust scores in an open web of trust is similar to the problem of deriving web page rankings in an open internet. Both the web of trust and the internet can be represented as a directed graph, where edges represent votes from one node in favor of another. Indeed, the defunct website Advogato used a web of trust with designated "seed nodes" in order to generate trust scores for members of the site[3]. Advogato's solution was Sybil resistant thanks to the seed nodes, and Google's updated ranking algorithm may very well use a similar concept of "seed" websites to protect against link farms.

There is a problem with seed nodes, however. Seed nodes are inherently more powerful than other nodes -- an inequality of power that we do not want in a decentralized system. Seed nodes seem to be necessary when these two requirements exist for a graph-based system: (1) that it must produce absolute scores for each node, and (2) that it must be Sybil resistant. We want to keep the second requirement, obviously. But what if we were to relax the first? Next we explore the concept of deriving *relative* trust scores. We define a trust score to be relative if it varies based on the observer.

Assumptions

In this framework, we assume an underlying web of trust that is represented as a directed graph. Each edge in the graph represents a *trust link*: a statement that one node trusts another. These links may be weighted or unweighted, depending on the implementation. This web of trust may be created explicitly (e.g. by explicit statements among users) or it may be derived from some other data (e.g. "follows" on a social media platform).

[3] https://web.archive.org/web/20170627230829/http://www.advogato.org/trust-metric.html

We do not make any assumptions as to what is represented by a node and what is represented by an edge. The most intuitive interpretation is that a node represents a human and that an edge represents a statement of trust made by one human about another. However, this framework may be applied to any application where nodes have use for decentralized reputation, and where there is some data available regarding which nodes are considered reliable by other nodes. These nodes could just as well represent devices, organizations, or something else entirely.

Trust Metrics

We define a trust metric as a function that takes as input a web of trust, a source node, and a target node, and outputs a value that represents the degree of trust from the source to the target. In other words, this function yields a degree of trust (how trustworthy one node is from the perspective of another node), given a web of trust. A simple example of a trust metric would be a function that counts the number of hops along the shortest path from a source node s to a target node t. The inverse of this number could be returned as the result of this function. The effect of such a metric would be that the degree of trust is inversely proportional to the number of hops required to get from s to t.

However, other metrics could also be used. Another similar metric might be to take the inverse of two raised to the power of the number of hops. In this metric, the degree of trust would decrease *logarithmically* as the distance from s to t increased. More complex metrics might do other interesting things, such as taking into account multiple paths from s to t.

Ultimately then, these metrics are all performing some sort of graph analysis, where the graph is centered around the source (i.e. from the perspective of the source). There is certainly a limit to what can be known purely based on a graph analysis, and intuitively one would expect this limit to depend on the connectedness of the graph as well as the error rate of trust links. It is assumed that there would be at least *some* error rate (i.e. some percentage of trust links that are directed to untrustworthy nodes). With too high of an error rate and/or too sparse of a graph, it might be difficult to distinguish with much certainty a trustworthy node from a malicious node. Despite these limitations, we believe that in some systems there may exist some useful information related to trust. It is the intent of this framework is to allow as much of this information to be utilized as possible.

Sybil Resistance

So far we have been using the term *Sybil resistance* rather vaguely. Here we will provide a precise definition that applies to the manipulation of trust scores in a web of trust.

To do this, we will borrow the concept of good nodes, confused nodes, and bad nodes from Advogato. Good nodes in a web of trust are honest and will not attempt to manipulate trust scores. Confused nodes are also honest and will not attempt to manipulate trust scores, but may mistakenly trust one or more bad nodes. Bad nodes, then, are those that will create fake nodes in order to game the system in some way.

The definition of a confused node is rather open-ended, however. Does a confused node trust a single bad node, or multiple bad nodes, or an unlimited number of bad nodes? For the sake of clarity, we will define an edge from a confused node to a bad node as a confused edge. We will also define a puppet node as a fake identity in a web of trust created by an attacker for the purpose of gaming the system. More precisely: puppet nodes are the subset of bad nodes that are not adjacent to a confused edge.

We propose the following definition: a trust metric is Sybil resistant if and only if the upper limit of the combined trust scores that can belong to bad nodes is bounded by the number of confused edges. Stated another way, a trust metric is Sybil resistant if and only if there is a finite amount of combined trust an attacker can gain in a system merely by creating puppet nodes.

This definition alone is not sufficient when considering Sybil attacks. While it helps to know that the impact that a Sybil attack can have is bounded, we might also wish to know the *degree* of Sybil resistance that a metric has. We will define the degree of Sybil resistance as the maximum trust that an attacker can have with a single bad node, divided by the maximum combined trust that can be accumulated by an attacker with an unlimited number of puppet nodes. A metric where puppet nodes have no effect has a Sybil-resistance degree of one; a metric where puppet nodes can increase the combined trust score of an attacker has a Sybil-resistance degree less than one; and a metric where puppet nodes can increase the combined trust score of an attacker without bound has a Sybil-resistance degree of zero (i.e. is not Sybil resistant).

Inclusion Status

Here we want to introduce the concept of *inclusion status*. We envision that many applications that use webs of trust will need to determine whether some identity is *included* in a set of valid identities. This inclusion status is subjective, just like the trust degree; the set of included identities will vary based on the source identity.

We define an *inclusion function* to be a function that takes as input a *trust degree* and an *identity cost* and returns a boolean (true or false). This trust degree and identity cost should pertain to a target identity in a web of trust, and the returned value should indicate whether or not the target identity is included in the source's trusted network.

We of course have not yet defined identity cost. We will define an *identity cost* to be some cost required for an identity to be considered valid. In the physical world this might be something such as an IP address, but this could also be a monetary deposit (e.g. burning a token of a cryptocurrency) or a proof-of-work based deposit (where proof of some computational work must be provided for validation). This measure is tied to that identity for its lifetime. The identity cost is assumed to be an objective measure that is not disputable. The intent behind the identity cost is to mitigate abuse through Sybil attacks. It can act in a similar matter to rate limiting, significantly reducing the damage that might be done by creating fake identities to manipulate or simply overwhelm the system. By attaching a cost to identities, sybil attacks become less economically viable.

DISCUSSION

The Intersubjectivity of Trust

Intersubjectivity has been defined as "mutual awareness of agreement or disagreement and even the realisation of such understanding or misunderstanding"[4]. Trust cannot be connected to objective realities or pure mathematical algorithms as it is highly connected to the concept of intersubjectivity. Trust is relative based on our perceptions of other human actors in a web of trust.

In the context of trust, this means that an actor's trust perception within a web of trust evolves continuously and subjectively based on their interactions and their observation of other interactions.

Isolation Prevention in Intersubjective Systems

In the section on Sybil resistance, we introduced the concept of good nodes, bad nodes, and confused nodes. While this is a useful framework for considering Sybil attacks, it is important to keep in mind that these labels are subjective and do not exist in the real world. What one node considers bad might be considered good by another. Graphs and groups that routinely exclude perceived "bad" nodes based on fixed rules will have a tendency to create isolated subgraphs and groups. Nodes will cluster around shared beliefs, perspectives, and interests. In doing so, they block communication channels and as a consequence continue to separate and draw groups apart. The result is a feedback loop that creates factions which are highly *intra*-connected but not well *inter*-connected.

We can describe this principle more generally: clusters of nodes in a system are bound to make contact, either continuously or intermittently. If they are continuously integrated, this process will be smooth and manageable. If they are allowed to drift apart for too long, they will eventually collide, and the collision may be explosive and unpredictable.

The challenge we face is thus: how can we keep clusters continuously integrated? Interestingly, the *confused nodes* that we described in the section on Sybil attacks may best be equipped to perform this role. Though they were considered an undesirable aspect previously, they may serve an important function in keeping disparate groups from drifting too far apart. We present the following questions as food for thought: *how can we create confused nodes that connect the groups that tend to move further apart?* And at the same time: *how do we design webs of trust where while encouraging confused nodes, we don't turn everyone into one?*

FUTURE WORK

Here we have attempted to introduce and define a framework for measuring degrees of trust. There is much follow-up work to be done. First, we welcome criticisms, extensions, and modifications of the framework. Second, we are interested in seeing various trust metrics proposed, in addition to analyses of the strengths and weaknesses of each metric. Game-theoretic analyses, models, and simulations are welcomed. We are interested in how this

[4] http://users.utu.fi/freder/gillespie.pdf

framework might be combined with other techniques to create real-world reputation systems. Finally, we are interested in further consideration of the human, social implications of the ideas presented.

ADDITIONAL CREDITS

Lead Author: Tyler Yasaka

Authors: Harrison Stahl, Titus Capilnean, Peter Snyder

About Rebooting the Web of Trust

This paper was produced as part of the Rebooting the Web of Trust VII design workshop. On September 26th through 28th, 2018, over 40 tech visionaries came together in Mississauga, Ontario to talk about the future of decentralized trust on the internet with the goal of writing 3-5 white papers and specs. This is one of them.

Leadership Team: Christopher Allen, Joe Andrieu, Kim Hamilton Duffy, Manu Sporny, and Heather Vescent

Gold Sponsors: Civic, Protocol Labs, Sovrin, an anonymous donor

Silver Sponsors: HTC, Microsoft, PTB Ventures, Spherity, Tierion

Community Sponsors: Blockchain Commons, Learning Machine, Legendary Requirements, Purple Tornado, Veres One

Workshop Credits: Christopher Allen (Founder), Joe Andrieu (Producer and Facilitator), Shannon Appelcline (Editor-in-chief), and Claire Rumore (Graphical Recorder)

Thanks to our other contributors and sponsors!

What's Next?

The design workshop and this paper are just starting points for Rebooting the Web of Trust. If you have any comments, thoughts, or expansions on this paper, please post them to our GitHub issues page:

https://github.com/WebOfTrustInfo/rwot7/issues

The next Rebooting the Web of Trust design workshop is scheduled for the week of February 27th to March 1st, , 2019. If you'd like to be involved or would like to help sponsor the event, email:

rwot-leadership@googlegroups.com