Stranger Danger

"Using Distributed Ledgers to Define Trust for Conversational Bots"

By: Michael Queralt
Michael@caumike.com
May 25, 2017

Trust = "firm belief in the reliability, truth, ability, or strength of someone or something"[1]

Hypothesis:

Is it possible to develop a model that provides users with the ability evaluate and trust a conversational smart bot before engaging with them?

Can Distributed Ledger technology be utilized to develop a data driven trust model that is rooted in the amount of transactions that the bot has completed?

To start, let's clarify some of the terminology in this document and describe the role that each one plays within the transaction.

- **Commercial Organization** – Any organization and/or brand that currently utilizes Automated Assistants or Conversational Interfaces to interact with the consumer either for support, knowledge transfer, commercial or banking transactions.
- **Users** – An individual or a surrogated assistant that interacts with a conversational bot for a commercial interaction – that may be a transaction, information or other activities.
- **Conversational bot, Conversational Interfaces or Automated Assistants**[2] – An automatic tool, usually with some Artificial Intelligence (AI) and Natural language (NL) capabilities to perform an action and interact with the end users in a conversational manner.
- **Platform** – This is the ecosystem where the conversational bot operates from – for example there are Alexa bots (Amazon) or the ones that work within Messenger (Facebook) or Slack and many others.
- **Distributed Ledgers (DLT)**[3] - A distributed ledger is a database that is consensually shared and synchronized across network spread across multiple sites, institutions or geographies
- **Blockchain -** A digital ledger in which transactions made in bitcoin or another cryptocurrency are recorded chronologically and publicly
- **Trust Score –** A number representing the aggregate number of transactions in a ledger. Both the ledger and trust score are immutable.

The Problem:

The unmitigated growth of bots and other conversational automated tools that are infused with AI and NL continues to increase and with it so does the risk that such interactions can become malicious in nature. The intimate nature and simplicity of the conversational interfaces make this a security issue that can be exploited for; behavior driven phishing, brand misrepresentation, transactional attacks and misguiding of funds based on misrepresentation. The results of such negative interactions are numerous; including the negative impact on the affected brands, breaks down the trust relationship between the user and the technology, and lastly it opens up another security attack vector where further attacks can be perpetrated.

[1] https://en.oxforddictionaries.com/definition/trust
[2] https://chatbotslife.com/chatbots-the-rise-of-conversational-ui-8a59078e2f95
[3] https://bitsonblocks.net/2017/02/20/whats-the-difference-between-a-distributed-ledger-and-a-blockchain/

An example of the risk of such interactions and the impact on a brand is made by John Tolbert.[4] in his article about "Fake Tech Support Spiders".

So how does a user determine that an automated assistant is trustworthy?

Let's use the example of a consumer entering a restaurant for the first time to – they are entering the restaurant based on a few things;

a. they have a need;
b. the fact that the restaurant has a physical presence creates a perception of validation and trust; and
c. the ability to evaluate the trustworthiness of the restaurant by its product displays, brands and the amount of people eating or had a meal at the restaurant, provides a level of comfort. That it is a restaurant and it provides food.

I believe that by using a distributed ledger model ; we can replicate such activities within the cyber world.

Here is the same example, interacting with a conversational bot.

1. the user has a need;
2. they reach out to a recognizable brand; and
3. today, the user has no ability to know how many (if any) transactions the bot has performed? What if – the user has the ability to know how many "validated" transactions the bot has performed? They can now *make an informed* decision before interacting with it.

DLT provides a mechanism to assess the trustworthiness of a bot because at its core, it is a transactional ledger where all parties validate the transaction before entering into a public ledger. This validation by independent parties, coming together for a given transaction provides a strong level of reliability, as they all must "vote" that the transaction was performed, before it is entered into a public ledger. Otherwise it is not entered into the ledger. Being independent parties, they can not be influenced or persuaded to validate the transaction by any individual party, making the DLT a mechanism that can provide the transactional information regarding bot activity and delivering a basic level of trust that the bot is performing the activities that it represents.

To clarify, this ledger is not to be confused for a feedback rating system, as the objective is to establish trustworthiness based on real interaction, which cannot be influenced or altered by outside parties. ***The DLT is focused on capturing how many transactions have been performed and validated by the members of the DLT, with the goal being to capture immutable data regarding the activity of the bot, which is validated by the members of the bot's ecosystem.***

[4] http://www.computerweekly.com/opinion/Fake-tech-support-spiders-on-the-world-wide-web

The solution:

Utilizing the power of the Distributed Ledger Technology to develop an algorithm that is then represented by a number that identifies the bot's activities and allows individuals to understand the risk of engagement with the bot, a number that we will refer to as the Trust Score (TS).

The Trust Score, which is derived from the entries in the ledger, may include the following information:

Description	Weight	Reason
Number of transactions	High	A higher number of transaction = higher completed interactions identifying the true activity of the bot.
Days from the first transaction	Low	Provides context on how long has the bot been active.
Days from the last transaction	Middle	Provides an understanding on the use of the bot, if the number is high, it may represent a non-valuable interaction.
Location of the most transactions	Low	If the location of the largest amount of the transactions are based in common geographical area it may represent that most of the activities are performed to influence its number via automation
Average transaction length	High	Understanding the average length of the transaction vs. the measured interaction, allows to understand the interaction value and if any automation is affecting its value

The Value:

By using DLT technology as the foundation for a rating system to be used by conversational interfaces or bots, we can create a framework that provides a trust value to the bot and gives the user or entity that is reaching out an ability to asses risk before an interaction occurs. Just to be clear this is not focused on the identity & authentication process of the bots (that issue will be addressed in the future) but it is focused on leveraging the transactional power that distributed ledgers have in validating transactions. The creation of a Trust score, derived from the information within a distributed ledgers allows users to understand, evaluate and minimize the risk of the interaction before sharing any information with a malicious actor, or gathering misguided information under a misconception due to spoofing or other malicious events.

The ability to use information derived from real data (based on completed transactions), in an ecosystem that all parties have a vote, creates a transparent result for the user. Such transparency can provide the basis for the development of a dynamic trust algorithm that can be utilized to assess the risk of interaction with an automated assistant. This type of tool , does not exist in today's environment and consumers are left to their own devices to make uninformed decisions when it comes to transactions and interactions. The creation of this type of framework would allow consumers/ users to make more informed decisions and have increased confidence during the transaction process.

As an example, application would be able to use the Trust Score.as a way of understanding risk and incorporate it within the authentication or conversational process, providing a higher level of awareness to the user, as they interact with the bot. Again, the value being provided by the ability of the DLT to deliver immutable information to determine activity.

How does it work:

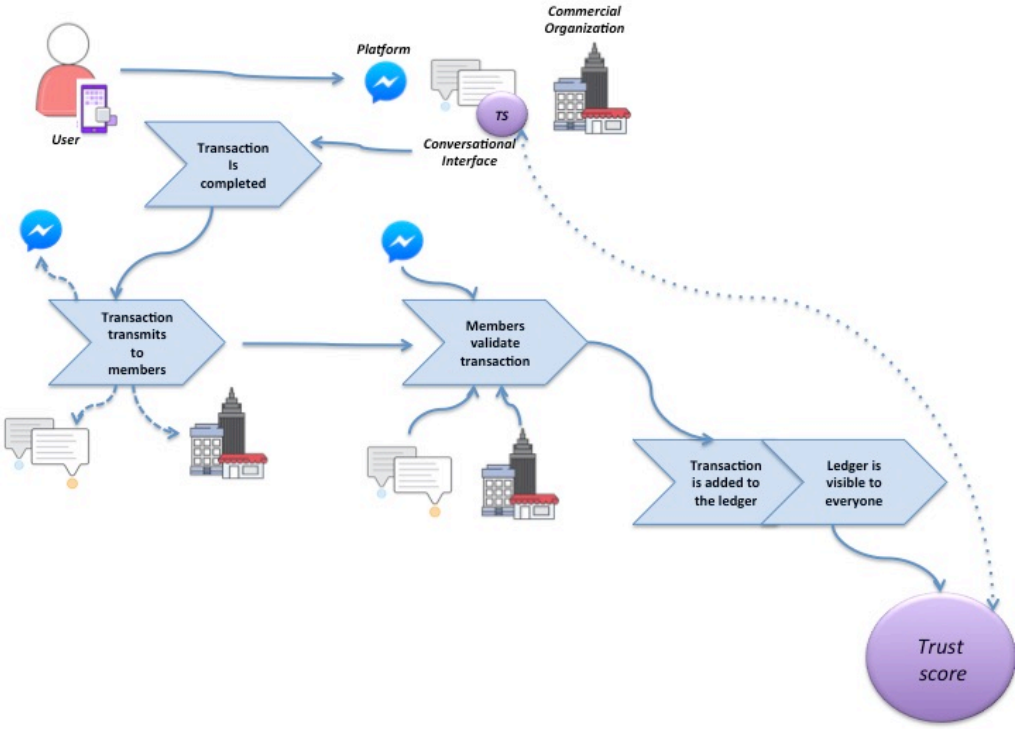

Figure 1- Description of the transactional flow & ledger inclusion

1. User has a need (i.e. purchasing movie tickets).

2. User is utilizing a messaging service (i.e. Messenger) and engages a conversational bot for the service.

3. The response comes back to start a transaction (i.e. purchasing movie tickets) via the conversational interface of the movie theater bot.

4. How does the user know if the bot is good or not? The bot presents itself as the suitable one, with all of the right logos and interactions, but can the user really trust it?

5. The user or its surrogate, can now evaluate the trustworthiness of the bot based on their Trust Score, a number that is composed of the sum of completed and validated transactions performed by the bot within the current ecosystem. This information is available to the user for viewing.

6. Once the transaction is complete, the only information that is captured into the distributed ledger is the date, time & transaction id. Since the focus is creating a trust framework for bots that is reliable based on real time data, the ledger does not capture any Personal Identifiable Information (PII) or value and content of the transaction.

7. At the end of the transaction all members involved, validate the transaction and then the sum of that score is entered into the public ledger– updating the bot's Trust Score.

Summary:

Is this approach valuable?: Trust in the cyber world is hard to achieve, but by using valid transactions we can create a starting point where users can have some guidance during their decision process – as they can now enter a conversation with the full knowledge of how many transactions have been performed by the bot – this immutable information provides a rudimentary level of trust, which is more information that is currently available to the user.

It is such independent validation that provides the strength to create a level of trust and a reference point for a user to make an informed decision, because their decisions are not based on inflated rankings or numbers that have been manipulated, but instead the bot Trust Score is based on transactional, factual numbers that have not been altered. This type of immutability provides a level of trust that is transparent to the user and that can scale as the bot ecosystem grows.

Future:

The convergence of conversational interfaces with the underlying power of the DLT, can bring a higher level of interactivity between users and technology. Developing an underlying trust model can be the baseline that bots can use to enter into a conversation and transaction, provides a framework that can also be used for bot to bot interaction and other new automated conversations that will flourish in the coming years.

References:

The Truth about blockchain : https://hbr.org/2017/01/the-truth-about-blockchain
IBM Blockchain information: https://www.ibm.com/blockchain/
Segregated Witness Benefits (SegWit) : https://bitcoincore.org/en/2016/01/26/segwit-benefits/
Litecoin; https://litecoin.org/
Ethereum : https://blockgeeks.com/guides/what-is-ethereum/
Ethereum: https://ethereum.org/