

Self-sovereign Identity as Trusted Root in Knowledge Based Systems

Nataliia Kulabukhova(✉)

Saint-Petersburg State University, Saint Petersburg, Russia
n.kulabukhova@spbu.ru

Abstract. In this paper we continue to speak about the concept of Self-Sovereign Identity (SSI), but not in the cases of IoT devices as it was in previous works [1]. The main purpose of this research is the usage of digital identity in two cases: a) SSI of a single person in Knowledge based system "Experts Ledger" and b) SSI of a company and candidate in HR matching systems. Though these two systems are developed for different issues, the idea of SSI in both is similar. The overview of these systems is done, and the pros and cons of using SSI with the relation of Zero-Knowledge Proof (ZKP) in each of them is made.

Keywords: Self-sovereign identity · Decentralized Identifiers · Expert systems · Distributed ledgers · Zero-Knowledge Proof

1 Introduction

Self-Sovereign Identity (SSI) become very popular for the last two-three years, because of the growing interest to the blockchain technologies, though the idea of it has its own life before the distributed ledger appears.

That is why, it is more correct to divide the SSI concepts into BlockChain and Non-Blockchain. In this paper, we will focus more on SSI using distributed ledger technologies, because the main stopping point of evaluation of Non-Blockchain approach was the usage of centralized root. The progress of distributed ledger gave the solution of this problem. In this case, the distributed ledger plays the role of a distributed trust root, excluding the possibility of falsifying data about digital identity and its capabilities in the digital world.

In the works [2–4] the survey of rules on which the SSI must be based is done. The main properties of SSI according to these papers are: Existence, Control, Access, Transparency, Persistence, Portability, Interoperability, Consent, Minimalization, Protection, Provable. All these is very important for the applications for SSI. The overview of existing applications [5] (though some of them are just the description of the concept) shown that it is very difficult to follow all these rules at once. Besides, very interesting assumption was made in [6] about the dependence from the type of blockchain: permissionless or permissioned. This makes development of SSI based applications a very comprehensive task, and the main reason for not existing widely used and fully developed open-course one yet.

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12251, pp. 14–24, 2020.
https://doi.org/10.1007/978-3-030-58808-3_2

Despite of mentioned problems the concept of digital verified identity of single person, company or even some IoT device, connecting with each other in some environment like knowledge based system is some kind state-of-the art and becoming really necessary in todays digital world.

In this work the idea of using SSI with the ZKP inside in two different knowledge based systems is described. First the overview of the "Experts Ledger" is made, the other is developing for HR purposes. But these two are only the examples of such kind of systems, the list is not limited by them. Further our conversation will show, that the idea is more general.

2 Overview of the Experts Ledger System

The main goal of the system "Experts Ledger" is to create an intellectual knowledge base about researchers in various fields of science, with the ability to rank and compare experts by different relevant criteria. The system also allows the user to browse the information about researchers, their scientific and expert works. Based on all the accumulated information, the system allows you to build the block of analytic data about researchers and their work. Types of work presented in the system are the following:

- Applications for participation in various projects;
- Grants where researchers are either leaders or participant;
- Research publications:
  1. Articles;
  2. Monographs;
  3. Books;
  4. Materials in conference proceedings;
  5. etc.
- Contracts;
- Expertises.

According to this information the system has the following features:

1. Flexible full-text search for people and their works (grants, contracts, articles, expertise);
2. Loading (adding) and uploading data about experts;
3. Graph representation (mapping expert relationships);
4. Analytics based on information available in the database.

2.1 System Architecture

The General scheme of the system interaction is shown on the Fig. 1. The system allows you to load data from several interconnected databases. After comparing and analyzing the data, the system forms its internal database, which is then used for work. In addition to uploading data from external databases, the system also allows to add external files uploaded to the database by system users.

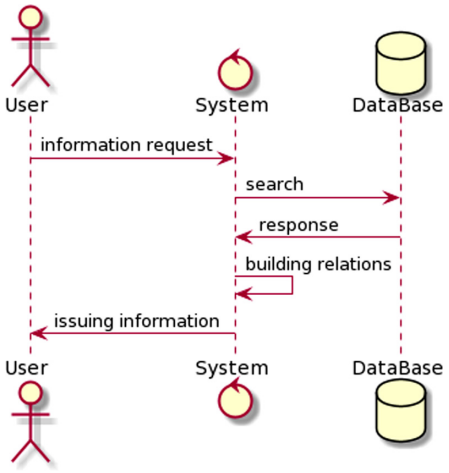

Fig. 1. Interaction diagram of system components

The user communicates with the system via the web interface. Then the system processes the user's requests and interacts with the database (figure), which stores all the information.

2.2 Database Structure

The General scheme of the database structure of the system is shown in the Fig. 2 The information is uploaded from different databases and preprocessing of the received data is made. After that, the database of the knowledge based system is formed and relationships are build.

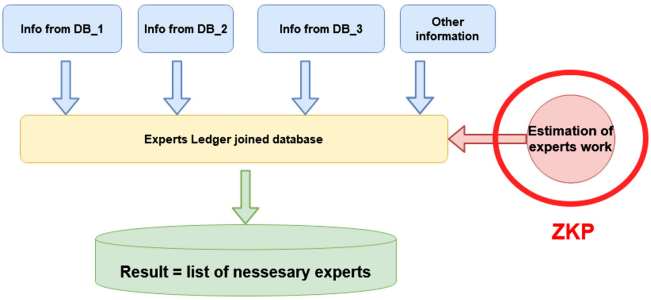

Fig. 2. General view of the system structure

For example, information is being uploaded from one of the bases on the server:

- People (persons);
- Projects (Projects);
- Grants (awards);
- Articles (scientific outputs);
- Requests (applications);
- Organizations (external Organizations);
- Dissertations (student Dissertations).

Updating occurs by adding new information to the existing database, while preserving all dependencies and ratings that were made before. This prevent the possibility of losing data. And all new relationships are made upon the previous iteration of update information, but not constructing from the beginning. This is very important for the SSI, because allows to avoid the reconstruction of Decentralized Identifiers (DID) for each entity of the system.

The database is updated 2 times a week. The system also provides a database update date display to correctly display search data on the date of database upload. Since syncing with Base1 and Base2 databases and then updating the existing database in the system takes several hours, it is not reasonable to update the data more frequently.

2.3 The Technological Stack

The technology stack for backend development includes the following components (the license types under which the software is distributed are given):

- development Language-Python 3 ([7]) (PSFL);
- a web Framework Flask ([8]) (Apache License 2.0);
- expert Database - Elasticsearch ([9]) (Apache License 2.0);
- user Management - SQLAlchemy ([10]) (MIT License).

The technology stack for frontend development includes the following components (the license types that the software is distributed under are shown in parentheses):

- development Language-javascript ES6+/typescript v3.6. 3 ([11]) (Apache License 2.0);
- JavaScript library for creating user interfaces-React V. 16. 0. 0 (React Hooks) ([12]) (MIT License);
- Style primitives-styled-components ([13]) (MIT License);
- to build the project, use parcel V. 1. 12. 3 ([14]) (MIT License).

2.4 The Use of Distributed Ledger

You can use a distributed ledger in this system in two states:

1. as a record of transparency of operations performed in the system (logging information about user actions);
2. anonymization of the expert (confirmation of the reliability of the assessment of expert activity without disclosure the private information about the expert).

In this paper we will discuss more deeply about the second case, though the first is also very interesting to research.

2.5 Estimation Settings Mechanism

In this case, experts are defined as researchers who carry out expert activities in a specific subject area. By evaluating the work of an expert, we mean confirming the expert's skills by specialists. Specialists can be other users of the system who have permission to issue an expert assessment.

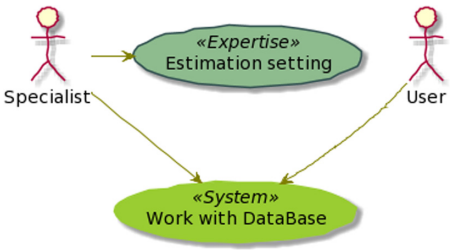

Fig. 3. Roles of system users when setting an expert assessment

There are two ways to evaluate the expert's work in the system:

1. Add a comment in the "expert profile";
2. Add a report on the conducted expertise with the results of the expert activity assessment.

How all this work i the concept of SSI and ZKP will be shown in Sect. 4.

3 Overview of Neety System

Neety is the HR platform the main purpose of which is to help employers and job applicants mach each other in an easy and lite way.

Obvious, that currently available applications for hiring employees do not perform the task as desired. The main idea of the solution described here is to

create a service that allows user to quickly and objectively find an employer or applicant in accordance with the criteria specified in the resume/vacancy. The goal of development is convenience and reliability in finding employees on the one hand, and work on the other.

The problems of existing HR systems:

1. Job descriptions do not reflect the essence of what you have to work on;
2. There is a lot of superfluous information in job descriptions, but the necessary information is missing;
3. Description summary is too long and unreadable;
4. Criteria for searching for employees and vacancies do not allow you to filter out the right people or companies;
5. The structure of the summary and vacancies does not solve your problem.

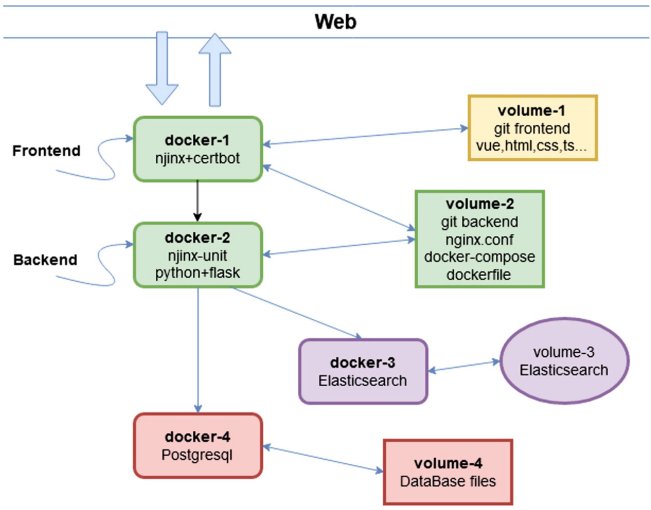

Fig. 4. The general scheme technological stack of the Neety System

Thus, the developed solution should reduce the amount of unnecessary information that hinders both the applicant and the employer when searching. To do this, the system has a built-in matching algorithm based on three main steps:

1. Morphological correspondence;
2. Matching via tags;
3. Neural network.

As you can see, this is a standard approach to natural language processing. The main feature is the method of filling out the questionnaire, which allows you to improve the output of results already at the stage of text processing by the neural network.

The main factors of influence on the results are:

- The relevance of the main skill;
- Relevance for the rest of the skills, taking into account prioritization;
- Relevance for the money (no money → to the end of the issue);
- Rating.

But this algorithm will be discussed in more detail in the next article, because the subject of this article is SSI.

3.1 The Technological Stack

The technologies we use in this case are practically the same as in "Experts Ledger", and they are listed in the Sect. 2.3 and are shown on the Fig. 4.

3.2 The Use of Distributed Ledger

In this case, the distributed ledger will act mainly as a decentralized trusted root, and the usage of ZKP is to evaluate the entities inside the system, based on a trusted, protected voting. Figure 5 shows the idea that using the Decentralized Identifiers based on the blockchain will protects the system from unverified users.

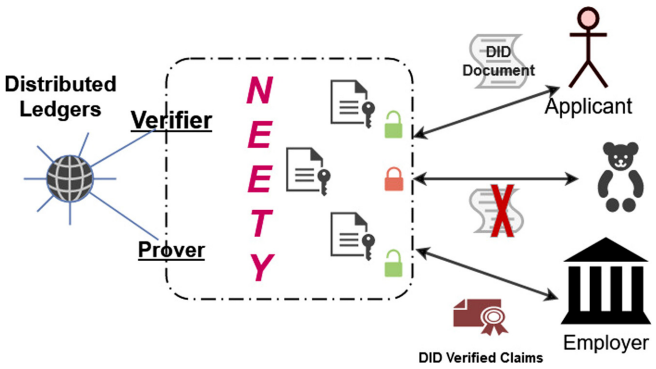

Fig. 5. The general scheme of the Neety System

3.3 Estimation Settings Mechanism

In the case of HR mechanisms ZKP concept as it was mentioned before will work generally as voting.

Factors influencing the rating:

- No confirmation-neutral (zero points added);
- there is a confirmation - plus one point;
- there is a probability that the data does not correspond to reality - minus one point.

4 The SSI Approach

Both systems discussed above are based on working with verified data from both users and organizations. In the case of the first system, we refer only to the reliability of the researchers' data contained in the system. In the second case, the data of both: the individual submitting their resume for a job, and the data of the organization (firm) that placed this vacancy - are important. Let's look at these two cases separately first, and then look at their similarities and differences.

In the first system, when we talk about researchers and their skills and competencies, at the beginning we based only on information collected from various databases. At the same time, this information can be either documented (we will call it verified information), or "written from words" (we will call it unverified information). And when we talk about the SSI concept, we mean just confirmation of data about this unverified information. To do this, the system introduces the concept of "Expert evaluation". What is meant by this concept, let's explain with an example. One of the researchers has information that he is an expert in the field of archaeology, but there is no information in our database about articles, projects and other documents in this field of research. In this case, this is unverified information. And in order to confirm or deny it, the system introduced the concept of "expert". An expert is a Trustee with the right to issue an expert assessment in a particular area.

In this case, the concept of SSI will apply both to the researcher with information that is not verified at the initial stage, and to the Expert who confirms this information. For a complete digital identity of the researcher, it is very important that all information is fully confirmed. To do this, the expert Advisor must also have a fully trusted digital identity. In the context of decentralized identifiers (DIDs), this means that the Expert has the right to rate the researcher. To do this, the expert must have a DID for rating in our example for archaeology. If the Expert has a DID for grading in mathematics, physics, or philosophy, but not in archaeology, the system will not allow the Expert to evaluate the researcher's activities in the field of archaeology.

In the case of the second system, two aspects are also important to us. The first is the confirmed data of the organization that posts vacancies. This will avoid finding "fake" vacancies in the system, and false triggering of the algorithm. In this case, SSI no longer works for one person, but for the entire organization. But on the other hand, the more applicants will also have SSI, the more reliable the information on the other side of the matching algorithm. Ideally, you can consider sending certain decentralized IDs to confirm your skills rather than filling out your competencies and creating a resume. However, this is still not possible in the near future, and it negates the operation of the matching algorithm.

5 Zero-Knowledge Proof in both Cases

First of all, it should be said that the Zero-Knowledge Proof concept (ZKP) as well as SSI existed and was used before the widespread usage of blockchain

technology began. The essence of it is that an individual can confirm that he has the right to perform some action, without revealing any information about him- or herself and who gave the right for this action. This principle was widely used in the development of distributed ledger Bitcoin. There it evolved from ZeroCash [16] first to Zcach [17], after it transformed to independent methods of zk-SNARKs [18] and zk-STARKs [19]. The comparison of SNARKs and STARKs is given in the Table 1.

Table 1. Comparison of SNARKs and STARKs main features

+/−	ZKP	
	SNARKs	**STARKs**
Pros	– Small proof size; – Fast verification; – Relatively fast proof generation;	– Resistance to hacking by quantum computers – Relatively rapid generation of evidence – Relatively fast verification of evidence – No toxic waste
Cons	– Complex procedure for setting public parameters; – Toxic waste; – Relative complexity of the technology; – Not resistant to hacking using quantum computers	– Complexity of the technology – Large proof size

Currently, these principles are mainly used to track the correct distribution of coins when making transactions in a distributed system. In the described above "Experts ledger" and "Neety" systems, the ZKP ideology is applied not only to the distributed voting, but also to the trustful interaction between the system participants. Let's look at this in more detail.

In this case, in both systems, we can use not only the SSI concept, but also the ZKP concept. When we talk about setting an expert rating for a researcher in the system of the "Experts Ledger", and when the rating of the applicant and the organization is calculated, it is important to correctly calculate the points assigned to all objects participating in both systems. It is important to maintain the reliability and independence of this rating. Therefore, it was decided to perform this calculation based on the blockchain technology and the ZKP concept. The main feature of using these tools allow the corrent user to be sure that this rating can not be forged or changed in anyone's direction. Thus, we get an independent variable rating of all objects in both systems, which on the one hand is based on the opinions of specific people, but on the other hand is independent, devoid of subjectivity.

6 Conclusion

Though, all described above has strong mathematical background, there are still some questions about the security and usability of SSI application itself and those with ZKP either.

Despite this, we are planing to integrate the technologies of ZKP concept in the future work on the projects. The existence of another ZKP application - Bulletproff [20] - should solve some of the problems of SNARKs and STARKs. And the future development of the project will be aimed to research and implement this technology.

Acknowledgment. The author would like to thank the whole team of the project "Experts Ledger": Oleg Jakushkin, Vladimir Korkhov, Artem Shurov, Ivan Marakhovskiy, Alexander Voskresenskiy. The author would like to express gratitude to Evgeniy Abramenko for the opportunity to work with the Neety project.

References

1. Kulabukhova, N.V.: Zero-knowledge proof in self-sovereign identity. CEUR Workshop Proc. **2507**, 381–385 (2019)
2. van Bokkem, D., Hageman, R., Koning, G., Nguyen, L., Zarin, N.: Self-Sovereign Identity Solutions: The Necessity of Blockchain Technology (2019). https://arxiv.org/abs/1904.12816v1
3. Stokkink, Q., Pouwelse, J.: Deployment of a Blockchain-Based Self-Sovereign Identity (2018). https://arxiv.org/pdf/1806.01926.pdf
4. Cameron, K.: The Laws of Identity (2005). https://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf
5. Kulabukhova, N., Ivashchenko, A., Tipikin, I., Minin, I.: Self-Sovereign Identity for IoT Devices. In: Misra S. et al. (eds.) Computational Science and Its Applications – ICCSA 2019. ICCSA 2019. Lecture Notes in Computer Science, vol. 11620. Springer, Cham (2019)
6. Panait, A.-E., Olimid, R.F., Stefanescu, A.: Identity Management on Blockchain – Privacy and Security Aspects (2020)
7. https://www.python.org/
8. https://flask.palletsprojects.com/en/1.1.x/
9. https://www.elastic.co/
10. https://www.sqlalchemy.org/
11. https://github.com/microsoft/TypeScript
12. https://ru.reactjs.org/
13. https://www.styled-components.com/
14. https://www.npmjs.com/package/parcel/v/1.12.3
15. Kilian, J.: A note on efficient zero-knowledge proofs and arguments. https://people.csail.mit.edu/vinodv/6892-Fall2013/efficientargs.pdf
16. Bitansky, N., Chiesa, A., Ishai, Y., Ostrovsky, R., Paneth, O.: Succinct non-interactive arguments via linear interactive proofs (2013)
17. Ben-Sasson, E., Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E., Virza, M.: Zerocash: Decentralized Anonymous Payments from Bitcoin (extended version) (2014)

18. Bunz, B., Fisch, B., Szepieniec, A.: Transparent SNARKs from DARK Compilers (2019)
19. Ben-Sasson, E., Bentov, I., Horesh, Y., Riabzev, M.: Scalable, transparent, and post-quantum secure computational integrity (2018)
20. Bunz, B., Bootle, J., Boneh, D., Poelstra, A., Wuille, P., Maxwell, G.: Bulletproofs: Short Proofs for Confidential Transactions and More. https://eprint.iacr.org/2017/1066.pdf