Open source zcash blockchain analysis platform
(development and hosting)
ZcashFoundation / GrantProposals-2018Q2, #39

Motivation and overview

People interested in investigating the capabilities and limitations of a given cryptocurrency usually have to start their job with time-consuming preliminary steps. These begin with installing a node and downloading and keeping track of blocks. Afterwards, they have to develop the way of obtaining information they are interested in, usually by either developing a set of RPC calls or importing the data to the database that might be easier to query. These steps are actually repeated for each researcher or specialist and preforming those is a time consuming prerequisite for starting the real job.

The goal of the proposed project is to enable the specialists and researchers to focus on their main tasks instead of taking care of creating and maintaining local environment. This goal is planned to be accomplished by a software platform that will offload the investigators by providing them a set of tools for doing research. This platform will be available as a free hosted service as well as for downloading, since its code will be made open source, it will be released as a docker image too.

The project is intended to shorten the time of starting and performing the analyses. If successful, apart from letting the researchers perform their tasks easier, it should also attract new people (e.g., with the network science background) to investigate the cryptocurrencies' blockchains. Those would not need the technical knowledge on how to convert the blocks data or handle RPC calls. This is also our motivation, since we want to support the community with a reliable platform that will bring more people to the community. We believe that the more eyes look at this area, the better, as new researchers can bring new insights into how to improve various aspects of zcash (and, possibly, other cryptocurrencies as well). Another goal we also want to achieve is to support the reproducibility of research, as this is one of the key factors for a trusted and reliable science (and not only science).

Technical approach

In 2017 we started developing a platform intended for allowing the researchers of our group (BERG, see below) to query the Bitcoin blocks' database in an easy way. This task has been finished in 2018Q1 and released as a open source code (authored by Marcin Pieczka, the member of BERG). Later on, this tool was used in a number of studies performed by the members of our group, covering the research on clustering the graph of transactions, finding interesting and unusual network structures etc. Based on a positive feedback of the members of our group as well as becoming more interested in zcash, we decided to extend it in a number of directions and this is the scope of current zcash grant project proposal.

Firstly, our goal is to extend the platform in order to support zcash cryptocurrency and keep track of the changes of zcash protocol in the sustainability period of the project (during the project and at least one year after the completion of the project).

Secondly, we plan to extend the API of the platform significantly, by adding the following components:

- the option of accepting queries to be run in the background with the notification of the submitter when the results are ready alongside with the link to download them,
- each query will have its unique ID in order to be able to restore old queries' results and to share them with others in order to support the reproducibility of results,
- the implementation of selected generic network science measures and metrics (diameters, shortest paths' lengths, centrality measures, clustering algorithms, e.g., centrality: degree, betweenness, closeness, clustering algorithms: Louvain method, modularity, CPM, random walk and others),
- implementation of different temporal metrics to analyze the dynamics of the zcash blockchain,
- quantitative analyses (balance, volume of transactions over time etc.),

- nodes' types classification algorithms (if possible),
- visualization of selected from the above mentioned metrics (distributions, change over time).

Lastly, the platform will be hosted within our infrastructure during the project and a year after the end of the project (hopefully for longer).

The implementation of the project will follow our past experiences with the Bitcoin platform. Namely, it will be still using MongoDB as a reliable NoSQL database. On top of that, the API (written in Python) will be extended with the above mentioned features. Moreover, the platform will own a dedicated website with the documentation, use cases, code examples, source code links. The platform code will be published on GitHub or other public source code repository. We will also encourage the platform users to provide us with the links to the research articles, articles or blog posts they wrote about zcash with the data gathered using the platform. This is to show what is possible and has been done so far, to integrate the researchers and experts in the area and to facilitate the reproducibility of research.

What are the risks we need to take into account? So far, our experience thought us that one of the most important things is keeping the platform in sync with the protocol changes. So, for instance, if zcash will update its protocol, we should react fast in order to adjust the platform to obtain and integrate new blocks. Moreover, we've encountered the problems with the project dependencies. For instance, if the underlying package has not updated to support some functions we based on (regardless of its area), this will require intervention (contacting tools' authors, the implementation of workaround etc.). For some analyses, resulting datasets can be large, so we will have to develop an efficient strategy to compress and store old results (without deleting them). At this stage it is hard to anticipate how many API calls we will be having, but the storage has to be taken into account. The server we use for hosting the platform is a high-end one (Intel Xeon, 144 GB RAM) with fast Internet link, so we believe within the sustainability period of the project we should not be having the difficulties in handling the load.

Team background and qualifications

The team consists of three members of the Blockchain Exploration Research Group (BERG, in short) affiliated at Wrocław University of Science and Technology, Wrocław, Poland. Its goal is to perform the research on the blockchain, mainly understood as a complex network. The group has been established in 2017 and consists from about ten people (varying over time, but including six core members). These are the researchers and enthusiasts that performed a number of analyses of the blockchain (for Bitcoin, as for now). The analyses covered investigation of clusters' structures, blockchain evolution evaluation and now the efforts are being done to focus on the classification of nodes' types based on their structural features and behavior in the network (assuming the knowledge on the identity of some nodes). Recently, BERG was presenting their research results at the BlockNet 2018 Sattelite meeting[1] in Paris. Moreover, the paper on the blockchain processing platform was submitted to the 1st Workshop on Blockchain and Smart Contract Technologies (decision pending)[2].

The PI of the project would be Radosław Michalski[3], PhD in Computer Science, the initiator and the leader of BERG. He works as an Assistant Professor at Wrocław University of Science and Technology. Before joining the University he has been working in a large Polish manufacturing company (listed on the stock exchange), firstly as critical systems administrator (IBM AIX, Linux, Oracle Database, data processing and HA clusters, ERP, security) and later on as the Director of the IT Department in the same company. His duties when managing IT included the responsibility for business continuity of IT services, development and implementation information security policies and managing the process of extending the ERP system used by the company to fit the business needs, managing projects and IT specialists. As a researcher, he is studying social and complex networks, and recently started to look at blockchain from network science perspective. He has also the understanding of basic and more advanced security concepts (both theoretical and practical), his personal interest in cryptocurrencies started in 2013. He has strong coding skills in

[1] Michalski, R., Zychal, B.: *Blockchain as a Complex Network - the Analysis of Trends of Bitcoin Blockchain*
[2] Michalski, R., Pieczka, M.: *Analysing Blockchain as a Complex Network - a Platform for Scientific Research*
[3] https://www.ii.pwr.edu.pl/~michalski

R, codes also in Python and writes advanced scripts in bash/ksh. His research also requires him to write efficient and distributed code.

The second member involved in the project is Mr. Marcin Pieczka, soon to graduate Computer Science at the same university. Marcin developed the first version of the blockchain processing and analysis platform[4] and took care of it for the first half of 2018 to provide reliable data for other researchers from the BERG group. He is a Python coder by passion, can work independently and as a team player. Apart from the experience in coding the platform, prior to that, Marcin worked as a Python developer for Nokia for a year. He has the knowledge on social network analysis that will also be needed in the scope of the project.

The last member to join the project will be another member of BERG with Python programming and API visualization coding skills, as well as the experience with testing. We have two people to chose from, this is why so far we can't provide personal details on them other that both fulfill the requirements.

Evaluation plan

The outcome of the project will be a working platform for processing zcash blockchain, including the API services providing the functions enumerated above. Apart from that, another criteria of the success of the project will include the availability of the platform online (using our resources), the publication of the source code and docker image as well as the existence of the webpage devoted to the platform with all the documentation.

At the beginning of the project, a document will be prepared with the detailed description of what has to be done. This document will be also used for progress monitoring and evaluation of the project. Apart from project documentation, the technical report will conclude the project.

Security considerations

The goal of the project is to enable other researchers to investigate (in many ways) zcash more easily. This should bring new researchers and experts into the field. It is expected that the outcome of their research that could impose some risk to privacy, integrity, availability and decentralization of zcash will be handled maturely by them (i.e., they will disclose the details to appropriate parties first). This will also apply to us during the implementation and hosting of the project, as we will be also performing scientific analyses. If any high level risk regarding the above factors will be discovered by us during the development of the platform, we will consult it with Zcash Foundation first before proceeding further.

Regarding the technical aspects of the platform itself, the risks we believe might occur is the extensive use of the API by malicious users (or even DDoS). The first will be handled by implementing limitations on API calls (but not enabling it first, but only if such a situation occurs). The latter will require special procedures. Yet, we believe the platform would not be the goal of DDoS attacks, since it not host critical or valuable data (in terms of direct monetization).

Schedule

The project is intended to last for five months in total (plus already described at least one year sustainability period when the platform will be hosted). The schedule for the project is the following:

- month 1 - creating an abstraction layer in the current platform to support multiple cryptocurrencies (i.e., extending MongoDB schema, creating structure for new full nodes, extending current API calls with cryptocurrency indication), creating a webpage of the project (*milestone 1*: having the platform with an abstraction layer for supporting multiple underlying cryptocurrencies)
- month 2 - integrating with zcash blockchain, testing the current version of the platform with zcash blockchain, developing the background job scheduler and notifier (*milestone 2*: platform integrated with zcash and supporting it with already implemented set of functions)

[4] https://github.com/MarcinPieczka/Mongo-BTC-Blocks-Database/

- month 3 and month 4 - extending the API with calls as defined above and implementing the visualization layer (*milestone 3*: passing functional tests for all the API calls and requirements as defined in the project document)
- month 4 - month 5 - testing the platform, creating documentation, extending the webpage with use cases and examples, polishing source code, creating docker images, starting to host the service and reacting to issues submitted by the community (*milestone 5*: passing stress- and security tests, *milestone 6*: providing final version of the webpage, source code and docker images published and documentation created)

The milestones conclude each period and are evaluated at the end of such. When a milestone will not be reached, the repair plan will be prepared and presented to Zcash Foundation for consultancy and approval.

Budget and justification

Total estimated budget for the project is $22,000. The project assumes the work of three team members, whereas two of them are focused mainly on the core development of the platform: coding, testing, documenting. The role of PI is the management of the project, taking responsibility for reaching the milestones, contacting Zcash Foundation, writing the technical report. On a lower level, the PI will be also involved in the development of the platform by: guiding the developers on how to implement network-oriented measures (as well as coding many of them himself), testing the API calls and providing high-level project documentation (including the project webpage). One external consultancy service will be needed, namely the security audit of the platform. When calculating the budget of the project, the assumption was made that the PI is involved in all project months devoting 40% of his time for the project (2 PM), developer 1 is available for first three months of the project for 66% of his time (2 PM) whereas developer 2 is involved in three last months of the project for 66% of his time (also 2 PM). This means that the project consumes 6 PM. The proposed project budget is presented below.

No.	Item	Amount	Responsibilities
1	Principal Investigator	$7,500	Managing the project, coding and testing (especially network science metrics and measures), documenting the project (2 PM in total)
2	Developer 1	$6,500	Coding, testing, documenting the code, mainly responsible for writing the abstraction layer of the platform to support multiple cryptocurrencies, but extensive testing for zcash within the scope of the project (2 PM in total)
3	Developer 2	$6,500	Coding, testing, documenting the code, Python, SMTP, bash (2 PM in total)
4	Security audit	$1,500	Platform security audit performed by experienced pentester (PI has the experience with ordering these consultancy services and later implementing the results of those audits from his past industrial experience as well as from the university, where security audits were required for another software project; name of the pentester can be revealed to Zcash Foundation per request)
	Total:	**$22,000**	

Email address for direct contact

The e-mail address for directo contact is: radoslaw.michalski (at) pwr.edu.pl.