# Research Proposal

Name: Ivo Schüepp  
Supervisor: Corinna Breitinger, Thomas Hepp  
Date of this Document: 30.01.2017

# Introduction

These days a huge amount of the success of scientists depend on their release rate of scientific papers. The pressure in many fields is high. To cope with that pressure there are a lot of cases of plagiarism, including self-plagiarism (where content is repacked), and even the release of fake articles (e.g. created with SCIgen, which is a random paper generator which creates papers which sound ‘scientific’, but at least they are non-sense).[[1]](#footnote-1)

Publishing companies try to filter out deceiving publications. But because the amount of papers, whose entry is huge, most of the work is done with computer algorithms for plagiarism detection. Another approach are conference management systems, submission systems or pre-printing services, which are internet based platforms where a paper can be reviewed by other peer-reviewers, scientist or publishing personal, who also should check for correct citation and plagiarism. Lately a case of fraud was in the media. A reviewer rejected a paper including an expensive study, stole the results and released it repacked at another publisher.[[2]](#footnote-2) The originator then blamed the plagiator for the fraud, but what if the author couldn’t prove that he is the originator?

The originator of a media relies often on the integrity of such services and its members, when he’s not looking himself for appropriate steps to ensure that he can provide evidence that he’s the owner.

A possible solution for this problem in conference management systems could be trusted timestamping. With trusted timestamping, we can proof that a certain document (or other digital media) belongs to us at a certain time.

Trusted Timestamping is in use for a long time. But mostly it was dependent on trusted timestamping authorities (TSA). These are central services and due to this one depend on the reliability and availability of the service.

In 2012 Prof. Dr. Gipp developed a new, decentralized approach using the bitcoin blockchain. This approach makes the trusted timestamping independent from TSA’s and even from his own provided service originstamp.org. This approach has several advantages. Since the bitcoin blockchain is worth billions of dollars, it assumable will last for a very long time. As long as a single node with the blockchain exists, the timestamps will be verifiable.

In short, for a certain media is a unique SHA-256 hash computed in the originators browser, which afterwards is sent to the originstamp.org API. The API accumulates all hashes over a period of time (currently 24h) and creates then an aggregated hash. This hash is encoded to a string via Base58. Afterwards a transaction in the blockchain is triggered with the string.

With the owners’ hash and the list of hashes which were used to create the aggregated hash, the media can be verified even without originstamp.org.

# Goal of the Project

Since the integration of originstamp.org in today’s available submission, pre-printing and conference management systems can ease the property evidence for the users of such systems as well as for the providers, we’ll focus on the best possible solution for them.

## Problem Statement

As far as we know, today’s available submission, pre-printing and conference management systems have no decentralized trusted timestamping service included. But rather they use the database, where the media is saved, which have no evidential value as well as they aren’t tamper-proof. Also, such databases are often centrally hosted, which is prone to availability issues.

Therefore, I would like to investigate how the originstamp.org API fits best in existing commercial and non-commercial submission, pre-printing and conference management systems. This may include the most-efficient way to integrate the API in the code. Also, it is to investigate what visual features the users need on the frontend, as well provider at the backend. Further we need to investigate how to keep up an effective workflow, if for example originstamp.org is unavailable for a lapse of time.

A base framework is recently completed by the Information Science group of Prof. Dr. Gipp, called CryptSubmit[[3]](#footnote-3).

## Research Objective(s)

With the goal that we want to bring the framework CryptSubmit to the level that it’s integrated in major submission, pre-printing and conference management systems, such as Arxive [[4]](#footnote-4) and EasyChair[[5]](#footnote-5) we need to investigate the terms of feasibility, usability, reliability and availability.

## Prototyping

The first step to get closer to the latter terms is prototyping, to get an overview for the requirements in commercial/ proprietary systems as well as to get familiar with potential issues. For this reason, we will adapt the submission system HotCRP[[6]](#footnote-6) to the CryptSubmit framework.

The following issues, amongst others, will be examined:

* Design an implementation that fast and cope with possible availability issues of the originstamp.org service.
* Design or adapt an overview page – so that every user can see their submissions and contributions including the timestamps. Also included is a direct verifying function.
* Briefly evaluate the prototype with testers.

## Integration in major systems

After we established contact to major system providers and have the permission to work on their code, we try to integrate the CryptSubmit framework with the experience we’ve gained with the prototype. This should happen in close contact with the providers. Also, the in the above-mentioned section issues will be further examined.

## Expected Results

The expected and/ or desirable result is to have the CryptSubmit framework well integrated in major systems. As well that this process is well documented and easily reproducible. It should meet the terms of feasibility, usability, reliability and availability.

# Work Plan

|  |  |
| --- | --- |
| 06.02 | 1. Get HotCRP working with Vagrant DONE 2. Get Proposal done IN PROGRESS 3. Get familiar with HotCRP code IN PROGRESS 4. Get familiar with originstamp.org API IN PROGRESS 5. Do an E-Mail draft for arxiv.org, EasyChair, … 6. (Implement API in HotCRP) |
| 13.02 | 1. Implement API in HotCRP 2. Create/discuss ideas of best solutions to integrate originstamp    * + 1. UI extensions        2. Backendarchitecture 3. Get the prototype done with HotCRP 4. Correspondence with major services |
| 20.02 | 1. Puffer 2. Try to get access to the code of properitary suppliers 3. Integrate our solution 4. Evaluate our solution |
| 27.02 |
| 06.03 |
| 13.03 | 1. Puffer 2. Start paperwork |
| 20.03 | 1. Finish paperwork 2. (Find Thesis) |
| 27.03 | 1. Puffer 2. End of Project 3. Get grade 4. Start BA |

# References

Gipp, Bela, Norman Meuschke, and André Gernandt. "Decentralized trusted timestamping using the crypto currency bitcoin." arXiv preprint arXiv:1502.04015 (2015).

1. http://news.doccheck.com/de/151679/fake-artikel-der-sechste-unsinn/ [↑](#footnote-ref-1)
2. http://www.faz.net/aktuell/wissen/medizin-ernaehrung/plagiate-in-der-wissenschaft-haengt-die-diebe-hoeher-14602306.html [↑](#footnote-ref-2)
3. “CryptSubmit: Introducing Securely Timestamped Manuscript Submission and Peer Review Feedback using the Blockchain” – 2017 - Bela Gipp, Corinna Breitinger, Norman Meuschke, Joeran Beel – not yet released [↑](#footnote-ref-3)
4. https://arxiv.org/ [↑](#footnote-ref-4)
5. http://easychair.org/ [↑](#footnote-ref-5)
6. https://hotcrp.com/ [↑](#footnote-ref-6)