# Research Proposal

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

# Introduction

In conference management systems, submission systems or pre-printing services digital content is uploaded for peer-reviewers to view, rate and making corrections or contributions.

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 as well as other uncredited use of information. Especially in the early stadium scientific work is fragile.

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.[[1]](#footnote-2) The originator then blamed the plagiarist for the fraud, but what if the author couldn’t prove that he is the originator?

The originator of a digital content 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 decentralized trusted timestamping. With trusted timestamping, we can proof that a certain document (or other digital content) existed at a certain time and thus we’re the originator. Also, this can be used to timestamp any contribution to scientific work, which makes sense when contributions are used uncredited. This makes it saver to use such systems for all participants.

Trusted Timestamping is in use for a long time. But it was dependent on trusted timestamping authorities (TSA). These are central services and due to this one depend on the trustworthiness and availability of the service. They also have fees, which makes it expensive when a high number of timestamps are needed.

In 2012, Prof. Dr. Gipp developed a new, decentralized approach using the bitcoin blockchain.[[2]](#footnote-3) This approach makes the trusted timestamping independent from a TSA and even from his own provided service originstamp.org. Since the bitcoin blockchain is worth billions of dollars, it presumably 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 digital content is a unique SHA-256 hash computed in the originators computer, 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, which is then a valid Bitcoin address. Afterwards a transaction in the blockchain is triggered to the address. After a time the transaction gets confirmed, which means that the transaction is in the blockchain with an assigned timestamp.

With the owners’ hash and the list of hashes which were used to create the aggregated hash, the task is reproducible, so the content can be verified in Bitcoin’s blockchain even without originstamp.org.

The approach through the aggregation of hashes makes it possible to supply a very high number of trusted timestamps at very low cost. Also, the overhead for management systems, submission systems or pre-printing services is minimal, since the main computations are done by originstamp.org and the originators computer.

# Goal of the Project

Since the integration of originstamp.org into 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, I will focus on the best possible solution for them. With the goal that it’s integrated in at least one major submission, pre-printing or conference management system, such as Arxiv [[3]](#footnote-4) and EasyChair[[4]](#footnote-5) .

## Problem Statement

As far as I know from website-inspection of arxiv.org, easychair.org and hotCRP, they have no decentralized trusted timestamping service included. For example, hotCRP uses the database to store a timestamp as well as a SHA-1 checksum of the content, but this has no evidential value.

Also, such databases are often centrally hosted, which is prone to availability issues. This approach isn’t bad, but the integration of DTT would simplify solving origin issues of content as well that it is serves evidence. A big plus for all participants would be that it consistently timestamps every submission (content, reviews, messaging).

Therefore, I will investigate how the originstamp.org API fits best in existing commercial and non-commercial submission, pre-printing and conference management systems.

## Research Objective(s)

A base framework for the design of such a system was recently completed by the Information Science group of Prof. Dr. Gipp, called CryptSubmit[[5]](#footnote-7).

This will 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.

## 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-10) to the CryptSubmit framework.

The following issues, amongst others, will be examined:

* Design a implementation that copes with possible availability issues of the originstamp.org service. There are different possibilities:
  + The simplest one is just to tell the user to upload his content later or to disclaim the DTT service.
  + Another one would be that the SHA-2 is calculated and given to a holding stack on the provider side until the service is up again. Of course, the user is informed also.
* Design or adapt an overview page – so that every user and provider, with corresponding DTT, can see:
  + Uploaded content (papers, additional data, …) and its different versions
  + Contributions/ reviews
  + Messaging
  + All DTTs are downloadable
  + Verifying function is given as well
* Provide on the option for DTT well communicated at any relevant uploading/ form spot.
* 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.

## Expected Results

The expected and/ or desirable result is to have the CryptSubmit framework well integrated in hopefully at least one major system like arxiv.org or easychair.org. As well that this process is well documented and easily reproducible.

With this approach users of available submission, pre-printing and conference management systems get more aware of origin and license issues as well that the use of the tools to protect content evidently proof is much easier and built in. So, the services get safer to use. This prevents the credulous nature of many users to blind trust in services.

# 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 |
| 13.02 | 1. Get familiar with HotCRP code IN PROGRESS 2. Get familiar with originstamp.org API IN PROGRESS 3. Implement API in HotCRP 4. Do an E-Mail draft for arxiv.org, EasyChair, … 5. Create/discuss ideas of best solutions to integrate originstamp    * + 1. UI extensions        2. Backend architecture |
| 20.02 | 1. Get the prototype done with HotCRP 2. Test/ brief evaluate implementation 3. Correspondence with proprietary providers: arxiv.org, easychair.org, .. |
| 27.02 | 1. Try to get access to the code of proprietary providers. 2. Integrate our solution 3. Test/ Evaluate our solution |
| 06.03 |
| 13.03 |
| 20.03 | 1. Start and iterate project-report |
| 27.03 | 1. End of Project 2. Start Thesis |

The tasks may overlap each other, as well that the planning is rough.

# References

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

1. http://www.faz.net/aktuell/wissen/medizin-ernaehrung/plagiate-in-der-wissenschaft-haengt-die-diebe-hoeher-14602306.html [↑](#footnote-ref-2)
2. Decentralized Trusted Timestamping using the Crypto Currency Bitcoin – 2015 - Bela Gipp, Norman Meuschke, André Gernandt - National Institute of Informatics Tokyo, Japan [↑](#footnote-ref-3)
3. https://arxiv.org/ [↑](#footnote-ref-4)
4. http://easychair.org/ [↑](#footnote-ref-5)
5. “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-7)
6. https://hotcrp.com/ [↑](#footnote-ref-10)