Dr. Marian Bubak
Department of Computer Science
AGH University of Science and Technology
Krakow, Poland
and
Professor of Distributed System Engineering
Informatics Institute, Faculty of Science
University of Amsterdam
The Netherlands
bubak@agh.edu.pl
http://dice.cyfronet.pl/

Kraków, April17, 2018

Report of PhD Thesis Assessment

Thesis: ***Towards a decentralised peer-to-peer cluster***
by ***Josef Gattermayer***

1.**Up-to-dateness of the dissertation.** Although the research presented in this dissertation started before 2012 (according to the date of the first publication of the Author on this topic), it is still vital, and it addresses important issues of the current research agenda related to efficient provisioning of distributed computing resources, namely scalability, efficiency, and fairness.

2. **Formal structure and organization of the dissertation.** The title is adequate to the thesis subject and its contents; it is sound and clear. The dissertation consists of seven chapters which are well organized and put in the right, logical sequence (introduction, related work, proposed solution, experimental evaluation, conclusions). The explanation of topics is clear, the level of technical accuracy is very high. Moreover, the material is very carefully edited.

3. **Completion of the dissertation objectives.** The Author has stated the following research objectives:

1. analysis of scalability of the Clondike cluster as it was available at the beginning of his investigation,
2. design, implementation and evaluation of an efficient protocol for scalable node discovery, bootstrapping and internode communication,
3. design, implementation and evaluation of a decentralized scoring system of single nodes in order to guarantee fairness among the cluster.

In my opinion, the research problem is important and it was defined in a clear way. The material presented in chapters 3 and 4 demonstrates a successful achievement of goal 1, in chapter 5 – goal 2, and in chapter 6 – goal 3.

4. **Assessment of the methods used in the dissertation.** Research methodology which was applied in the investigations leading to this thesis consists of a short analysis of new challenges related to the cluster, a detailed evaluation of the existing Clondike cluster, elaboration of an innovative concept of the communication and node fairness framework, its implementation, and experimental validation. In my opinion, such an approach is appropriate for this type of studies. It was consequently applied having in mind and measuring performance and scalability (chapters 3, 4, 5, and 6).

5. **Evaluation of the results and contributions of the dissertation .** In Chapter 2 the Author presents a a good overview of background of his investigations (clusters and blockchain technology) as well as a description of past and recent developments in the area of cluster computing. The overview is detailed,

and, at the same time, focused on the research objectives of this thesis. It is evident that the Author has acquired a broad and deep knowledge of the today's state-of-the-art in this area.
The main achievement of the Author, besides of the full realization of the planned research objectives and providing a new, scalable cluster for practical usage, in my opinion, is a sound demonstration of applicability of the design science methods: the careful combination of requirements analysis, usage of existing methods (P2P, blockchain) in design of a new solutions, and experimental assessment of the artefacts. In my opinion, the most valuable are Chapters 5 and 6 of the dissertation.
It is worth mentioning that a broad range of results of the research described in this thesis has been published by the Author as four reviewed IEEE conference papers.

6. **Remarks, objections, notes, and questions for the defense.** In my opinion, in section 1.1 the Author should state more clearly a set of requirements related to the new cluster followed by a research hypothesis before presenting the research objectives in section 1.3. Section 2.2, presenting the related work, is missing a complete summary of main drawbacks and shortages of existing solution. An important shortage of the dissertation is lack of a formal model of the cluster. A deeper understanding of the scalability could be achieved with a formal model of the cluster performance which could result in formulation of an isoefficiency function. In my opinion, a better insight into the cluster performance could be achieved by presenting efficiency vs. number of nodes instead of just execution time. The Author should explain in detail advantages of the elaborated solution over typical grid and cloud systems. As the cluster is of general purpose, it would be reasonable to provide also examples of practical use cases run on it with some performance parameters.

7. **The overall evaluation of the dissertation.** The research plan which led to this thesis was well though-out. The investigations presented in this thesis are complete: from an analysis of the-state-of-the-art through design analysis, improvements of protocols and algorithms to implementation and experimental evaluation of solutions. The goals of the dissertation stated in section 1.3 have been fully achieved. This research is a valuable contribution to the field of distributed computing infrastructures and the new cluster is useful as a computing resources.

8. **Recommendation.** The research objectives have been achieved using the right methodology, and, in my opinion, this dissertation fulfils the requirements that are set for PhD thesis, so I recommend the dissertation of Mr Josef Gattermayer for the presentation and defense with the aim of receiving the PhD degree.

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Informatik Institut für Systemarchitektur, Professur für Rechnernetze

Technische Universität Dresden, 01062 Dresden

Prof. Dr. rer. nat. habil. Dr. h. c.
Alexander Schill
Doctor Honoris Causa/ Universidad Nacional de Asunción

Telefon: 0351 463-38261
Telefax: 0351 463-38251
E-Mail: Alexander.Schill@TU-Dresden.de

Dresden, December 15, 2017

Review Report on the PhD Dissertation 'Towards a decentralised peer-to-peer cluster'

of Mr. Josef Gattermayer

Distributed computing in structured clusters of servers has gained significant importance in the age of grid and cloud computing. However, for highly decentralized structures various important research challenges remain, especially regarding performance, scalability, and trust. Based on the existing cluster architecture Clondike, the PhD dissertation of Mr. Gattermayer addresses these issues in a dedicated way and proposes several novel solutions. In this sense, the thesis is up-to-date and is concerned with a timely and relevant topic.

Chapter one introduces a clear problem statement with the major goals of conducting a performance study of Clondike based on real use cases, developing a scalable and decentralized node discovery, communication and bootstrapping system, and establishing a decentralized node reputation check mechanism.

In chapter two, the state of the art regarding Clondike itself, cluster computing in general, and also blockchain networks as a mechanism for trust establishment is discussed in detail. Several cluster systems such as Kerrighed, XtreemOS, Condor, as well as various older projects are analyzed and compared with the Clondike approach. This way, a solid foundation for advanced contributions is provided, and existing work is clearly distinguished. Similarly, several dedicated blockchain mechanisms are discussed with respect to their characteristics and possible use cases.

The third and fourth chapter present the first part of the own contributions, namely the Clondike performance study with respect to a homogeneous and heterogeneous use case on distributed compilation. In order to have a clear benchmark, conventional distributed

compilation with distcc is compared with Clondike-based execution of the same tasks. It is shown that the general-purpose solution Clondike can actually match the performance and scalability properties of the more limited special-purpose compilation solution distcc. In other words, this means that the thesis has proven the suitability of the general cluster solution for a broad range of possible applications.
For the heterogeneous tests, Clondike was ported onto a general cluster of different workstations with highly varying resource characteristics. Several implementation-level problems regarding installation, device access, and bootstrapping were solved successfully. The performance results for the heterogeneous case are also quite interesting and outline possibilities for further optimization regarding task allocation and task migration with respect to the given resources of individual compute nodes.

In chapter five, a new bootstrapping and communication protocol for Clondike is presented, replacing a conventional, scalability-limiting broadcast solution by a more advanced structured peer-to-peer solution. While such mechanisms are already very well-known from file sharing approaches such as Kademlia, the major challenge was to adapt them towards a computing-related instead of data-related infrastructure. After a detailed conceptual analysis of Kademlia, the associated own implementation in Ruby is presented. Over all, the subsequent performance tests confirm impressively that the developed solution behaves logarithmically regarding the required number of messages in relation to the growing number of participating nodes in the experiments. This confirms that the integration of structured peer-to-peer mechanisms into the architecture for cluster computing works well and enables a large degree of improved scalability.

Finally, in chapter six, the thesis describes a new trust mechanism that can ensure that a peer node is a trustable entity, for example in the case of cluster-based task migration onto such a node. The basic idea is to apply mechanisms known from blockchain approaches for crypto-currencies towards the target area of cluster nodes. For each request regarding task migration, both the source and the destination node can check the trustworthiness of their peers based on distributed, non-forgeable log entries that are publicly available throughout the overall distributed system. This way, a multi-level reputation scoring system has been established. The overall behavior of the system has been validated based on a set of experiments that demonstrate the feasibility of the solution.

The formal structure and overall organization of the thesis is consistent and is according to the expected conventions. All the objectives have been met successfully.

Based on the rich scientific contributions, the appropriate methods used, the solid practical validation, and the related publications of the candidate, I would like to recommend the thesis of Mr. Gattermayer to be accepted for the defense as a PhD dissertation by the Faculty of Information Technology of the Czech Technical University in Prague.

A. Schill

(Prof. Dr. Alexander Schill)

Prague, March 3, 2018

Doctoral Thesis Review

Thesis: Towards a Decentralized Peer-to-Peer Cluster
Author: Josef Gattermayer
Reviewer: Petr Tůma

In computing environments that integrate personal workstations into local networks, we can often observe bursty workload patterns with low average utilization – most workstations are powerful enough to run most considered workloads, but are also idle most of the time because it is rare for all users to execute their workloads simultaneously. This has long invited research into harnessing the combined power of the workstations by forming workstation clusters.

Interestingly, many technical challenges of workstation clusters – such as live process migration or remote resource access – have been solved in many contexts, but we are still far from the ultimate goal of utilizing local workstations in a seamless experience. Not only there are hard open problems, such as the security issues related to hosting potentially malicious code or utilizing potentially malicious platforms, but it is also not clear whether enough common workloads would benefit. Many solutions also require complex and fragile installation procedures that hinder practical adoption.

In this context, the thesis aims to advance the state of the art towards the ideal of a decentralized peer-to-peer cluster. The work is done in the scope of Clondike, an open source workstation cluster project hosted by the Czech Technical University. There are three specific goals – (i) an evaluation of Clondike performance on a practically relevant workload and a common workstation cluster, (ii) an extension of Clondike with peer-to-peer workstation discovery mechanism, and (iii) an extension of Clondike with reputation system to prevent cluster abuse. I am convinced these goals are both sufficiently relevant and sufficiently challenging for the purpose of the thesis defense.

In the following, I first present some observations that concern the thesis as a whole, and then focus on each of the three contributions in turn. I also include questions that may help better assess the thesis – as I explain later, I believe these questions should be addressed during thesis defense.

Overall Observations. The thesis is accompanied by a software artifact of significant size. This is often the case with software oriented research, and the research community has long accepted that software artifacts can represent valid research results. Because the thesis text does not talk much about the artifact, there is a risk that it remains an unappreciated part of the contribution. ***Q:*** *What are the basic dimensions of the Clondike project in terms of software size and what is the contribution of the thesis author to this artifact ?*

My other overall observation concerns the thesis structure. When introducing the individual contributions, the thesis tends to delve into detail with little regard for the big picture. As a result, certain design decisions seem to lack clear justification, which should probably be derived from the overall goals of the Clondike project. These are never really spelled out. ***Q:*** *When the thesis talks about "a global peer-to-peer cluster", what are the envisioned dimensions in terms of nodes and users and network distances ? What are the target workloads that the cluster should accelerate ?*

(i) Performance Evaluation. The first of the three specific contributions presents a performance evaluation of Clondike, designed as a comparison between a generic cluster framework and a purpose built tool on the task of distributed kernel compilation. A major factor in the evaluation is the choice of workload. The thesis argues (Section 1.4.1) that the workload exhibited by kernel compilation should be reasonably representative, but it is not clear whether the thesis refers to other compilation workloads or other workloads in general. In any case, kernel compilation has some quite specific features that may not apply to other projects, and the performance evaluation would benefit from more discussion on the workload selection topic.

I also find it surprising that the evaluation did not provide any data about network traffic. The very basic assumption that process migration hinges on is that, at least for some workloads, the performance of migrating and executing remotely is better than the performance of executing locally. A performance evaluation that omits network traffic information is therefore missing an important component, even more so when the cluster has global ambitions and fast connectivity cannot always be assumed. ***Q:*** *What was the impact of network traffic and network utilization on the evaluation, and how would the balance change if a slower or a faster transport were available ?*

Another set of questions is related to the presented conclusions. For the homogeneous environment, the thesis only mentions one limiting factor for Clondike performance (Section 3.8), namely the fact that even small tasks were migrated. Hypothetically, the effort needed to migrate small tasks is not compensated by the benefits of running more tasks in parallel on more machines, however, the thesis does not mention any investigation done to back this conclusion. The conclusions for heterogeneous environments (Section 4.5) point to a naive scheduler design and since the relevant article mentions this as a subject of ongoing research back in 2012, it would be interesting to know what improvements to the scheduler were done between 2012 and 2018.

(ii) Peer-to-Peer Node Discovery. The second contribution of the thesis deals with node discovery in peer-to-peer environments. Here, the Kademlia protocol is used to connect the nodes in a cluster. Perhaps the most striking feature in this part of the thesis is the obsession with node disconnection support – even though the thesis talks about a "deeper analysis of possible solutions" (Section 5.1), the choice of Kademlia is made quite quickly and disconnection with zero communication is the only reason provided. This is quite strange. A node that disconnects must have connected before that, hence it would make sense to look at both the connection and the disconnection cost before choosing a protocol. Also, the ability to recover from sudden node disconnection is not unique to Kademlia, and the essence of the solution – network reorganization in response to communication failure, possibly with heartbeat messages thrown in – is the same in many other systems. Finally, if abrupt client disconnection is an issue, then I would expect the thesis to discuss what happens to the migrated tasks. ***Q:*** *If client disconnection is expected to be a frequent occurrence, how does this impact Clondike usability, assuming migrated tasks are lost ?*

I also find it somewhat confusing that the thesis lists network address discovery and NAT traversal as two requirements that the protocol must meet (Section 5.2.2). The thesis uses vague technical terms, such as "forcing a connection to a given node in the cluster by providing a static

Department of Distributed and Dependable Systems
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
phone +420-951554245, fax +420-951554323
http://d3s.mff.cuni.cz

link" (Section 5.2.2) or "direct access to IP addresses and ports of single nodes" (Section 5.3.2), but it really does seem like the two requirements are simply technically incompatible. When the thesis concludes that none of the existing Kademlia implementations meets the requirements (Section 5.4), this seems to be merely a natural consequence of invalid requirements. ***Q:*** *What exactly is meant by the direct connection requirement and does the new implementation of Kademlia delivered in the thesis meet both this requirement and NAT traversal ?*

Most importantly, the thesis focuses on the role of Kademlia for node discovery, but it only ever stays with the concept of connecting to arbitrarily chosen nodes. Clearly, a peer-to-peer cluster would benefit from features such as discovering nearby (in terms of network connectivity) nodes, discovering idle nodes, discovering nodes with particular properties (processor, memory, software), or even a combination of these features. ***Q:*** *How does the proposed node discovery mechanism support (eventual) smart node selection that the Clondike cluster evidently needs, based on the performance evaluation conclusions ?*

(iii) Reputation System. As the third contribution, the thesis describes a reputation system based on a process migration ledger, which can be used to compute reputation values. These are then used to govern acceptance decisions for incoming migration requests. The proposal offers some interesting ideas, such as the use of canary tasks for fraud detection. This could be extended with a deeper elaboration, such as an estimate of what share of canary tasks would have to be used to achieve reasonable protection from fraud, or an explanation of how the canary tasks would escape premature recognition.

The reputation system overview does not discuss behavior with long term execution. Here are some scenarios that make the reputation system look less credible. Since the reputation values are only ever increasing, even fraudulent nodes will eventually reach reputations that will make them acceptable, especially if a network of fraudulent nodes collaborates. New joiners will be forced to use fraudulent partners because they will not have the reputation needed to enlist reliable partners. High reputation nodes get more opportunities to further increase reputation, creating a positive feedback loop. ***Q:*** *How will the reputation system behave in a long running cluster with high node churn and highly variable membership duration ?*

I also find it strange that the thesis uses round robin node selection for evaluation (Section 6.4.3) but uses other (fairly vague) strategies to provide algorithm validation sketches (Section 6.5.1). ***Q:*** *What (single) strategy would be needed to meet the use case requirements (Section 6.3.1) and how would that strategy perform in the evaluation (Section 6.5.2) ?*

As I have observed before, most of my questions may be due to the lack of overall strategy description in the thesis. If so, such an omission should not completely and automatically eclipse the individual contributions, which must be weighted carefully. While there are other details in the thesis that might warrant attention (such as related work coverage), I believe that appropriate answers to the questions raised in this review can establish a convincing framework towards successful defense. In this light, **I recommend the thesis for presentation and defense.**

Petr Tůma

Department of Distributed and Dependable Systems
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
phone +420-951554245, fax +420-951554323
http://d3s.mff.cuni.cz