

# **Kickoff, A Tool For Software Companies To Control Their Labor Cost Safely**

Zhu Yue Cai

ENCS 282 YA/ Assignment 5 Research Proposal

Instructor: Dr. Brandiff Caron

December 2, 2014

## **Abstract**

It is harmful to a software company when there are some developers leave their positions. The damage is manifested as abandoned files on the software system that can be difficult to maintain. Therefore a tool to help software companies to understand the damage and how they can reduce the damage is very useful.

This proposal suggests to build up a software Kickoff to fulfill this demand. The software will first extract change information from the company's version control system to build up the database. Then according to different input from the user, the optimized truck factor algorithm with a stopping condition calculates the damage for one or all possible layoff lists. After the list of abandoned files is generated, it applies the candidate selecting algorithm with to find out the experienced developers who can take over the abandoned files or to mentor new joiners to do so.

The direction for further improvement is also discussed in the proposal.

# Table Of Contens

Cover Page	i
Table of Contents	ii
Executive Summary	iii
1. INTRODUCTION	1
2. THE DESIGN OF KICKOFF	2
3. Conclusion	4
4. SCHEDULE	5
5. BUDGET	6
6. QUALIFICATIONS	7
7. PRESS RELEASE	8
8. Conclusion	9
References	10
Appendix A: The Optimized TF Algorithm	11

# Executive Summary

It is harmful to a software company when there are some developers leave their positions. The damage is manifested as abandoned files on the software system that can be difficult to maintain. Previous researches show that, if the abandoned files are taken over by experienced developers who have the relative knowledge, or by new joiners who are mentored by such an experienced developer, the damage can be reduced. Therefore a tool to help the software companies understand the abandoned files for their layoff list and which remain developer could be able to take over the abandoned files is very useful.

This proposal suggests to build up a software Kickoff to fulfill this demand. The software will first extract change information from the company's version control system to build up the database. Then according to different input from the user, the optimized truck factor algorithm with a stopping condition calculates the damage for one or all possible layoff lists.

After the list of abandoned files is generated, it applies the candidate selecting algorithm to find out the experienced developers who can take over the abandoned files or to mentor new joiners to do so. This algorithm generate 2 matrices: The dev-file matrix contains the the history of how many changes were made by a particular developer to a particular file; The file-file matrix contains the history of how many co-changes happened between a particular file and another particular file. The product of these two matrices how much knowledge of a particular file a particular developer has. Which helps Kickoff to find out demanded experienced developers.

The direction for further improvement is also discussed in the proposal. The optimized truck factor algorithm is still too slow for large size developer team. Further improvement should be how to implement parallel computing technique to speed up the algorithm. The file-file matrix does not consider the files dependency. How to apply this idea at the candidate selecting algorithm is another important improvement.

The rest of the proposal gives out the task schedule and budget approximation, as well as the qualification of the project creator.

# 1. PROJECT DEFINITION

Laying off developers and recruiting new joiners are two common software company activities. But Previous research has shown that high turnover rates leads to a decrease in productivity [4]. Like other jobs that involve knowledge workers, when a software developer leaves, there is a knowledge gap. This gap is manifested as abandoned files on the software system that can be difficult to maintain.

Generally, these abandoned files are taken over by new joiners or remain developers. But the time needed for successors to learn how to maintain the abandoned files could be very long, if the successor has no knowledge of the files. Therefore finding the right successors for the abandoned files is important to software companies when they need to lay off developers.

There is a large literature on mentoring and integrating new developers into software projects. For example, Zhou and Mockus examined the impact of development environment on new developers[8] . Bird et al. [1] looked at the survival rate of new developers. Zhou and Mockus examined the amount of time until a developer becomes productive. Mockus [5] suggested mentors for developers based on past work and Canfora et al. [2] suggested mentors based on the email communication network.

The experienced developers who have the knowledge to the abandoned files will help reduce the damage. They can take over the abandoned files themselves or mentor the new joiner to do so. Thus we believe that a software which can calculate the number of abandoned files and produce a list of experienced developers who have the knowledge to those files for a given layoff list is very helpful. Moreover, If this software could tell the same information for all the possible layoff lists based on input critical variables, the decision process would be simplified dramatically. This is the motivation of Kickoff. It is a useful management tool for software company to manage its knowledge storage.

## 2. THE DESIGN OF KICKOFF

In order to minimize the cost, we will use only opensource tool to build Kickoff. And what we need is "PostgreSQL" as our database.

In computer software engineering, revision control, or version control, is any kind of practice that tracks and provides control over changes to source code. Software companies use version control software to maintain documentation and configuration files as well as source code. The first step of building Kickoff is to develop a mechanism to extract the source code change history from their version control software. Nowadays, a large number of software companies use "Git" as their version control system which is also opensource. Therefore "Kickoff" should be able to extract information from "Git" at the beginning. Since "Git" has a function called blame can identify changes and who made the changes in your project, we only need to input the result in the database. And this can be done by string processing technique and the psql command. We only need to integrate this into the software.

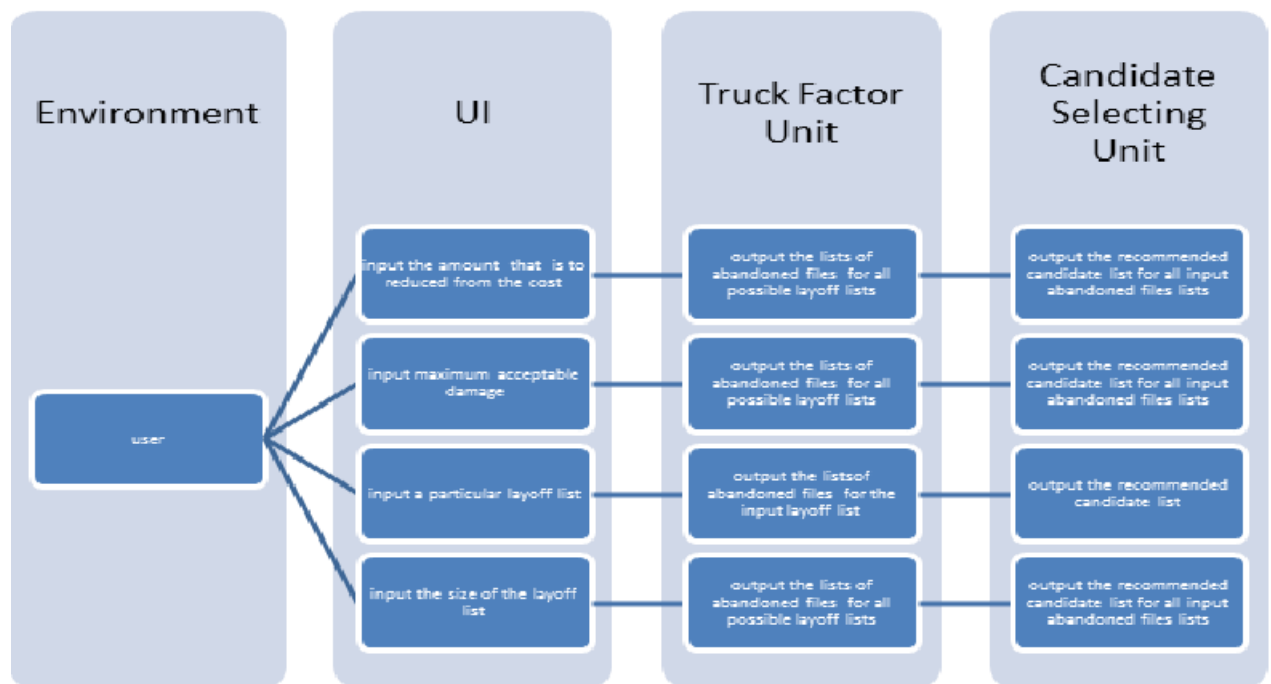
With the database ready, It is easy to calculate the number of abandoned files for a single given layoff list or for all possible lists of a given size. But it is very difficult to calculate the number for all potential layoff lists. We can use the "truck factor" algorithm proposed by Ricca et al [6] for the job. In software development, the truck factor is the number of key developers who would need to be incapacitated to make a project unable to proceed. It is also known as bus factor, or lottery factor and is a measurement of the concentration of information in individual team members. The concept of truck factor is not related to the project, but the algorithm to calculate it is what we need. However, as Ricca et al pointed out[6], the time complexity of this algorithm is too large to implement in the industry. Therefore, optimizing this algorithm is the first problem we need to solved. We introduced a stopping condition to optimize the naive algorithm. What it does is to determine whether the unarticulated lists will have a number of abandoned files greater than the input value. If no, it will stop the program and save time. Then with the user input their maximum acceptable abandoned files amount, The program will work only for the layoff lists that satisfies the condition. For more detail about the optimized truck factor algorithm, please refer to Appendix A.

Even with this optimization, the algorithm is only piratical to company with less than 100 developers. Further research on how to apply parallel computing technique in the optimized truck factor algorithm is required, such that it is also piratical large software companies.

The second problem is how to integrate previous researches to generate an algorithm for selecting the experienced developers for each abandoned file. We introduce a successor candidate selecting algorithm for this functionality. The general idea of this algorithm is to pick out those developers who have worked with the files that have co-changes with the abandoned files as candidates. The intuition behind is that if file A has changed with file B, and file A becomes abandoned, then the developer who works on file B will likely know something about file A. The idea of this algorithm comes from the matrix multiplication in Cataldo et al.'s[3] The algorithm will first generate two matrices from the database. One is the developer-file matrix whose entries are the number of change a developer has made to a file. The other one is the file-file matrix whose entries are the number of co-changes a file has with

the abandoned file. The product of these two matrices will give us a measure of how much knowledge a developer has to the abandoned file. Then we could select the one with the largest knowledge measure to take over the abandoned file.

With all these function units prepared, the conceptual structure of Kickoff is shown in the following figure.



### 3. DISCUSSION ON FURTHER IMPROVEMENT

As mentioned above, the optimized truck factor algorithm is practical only for companies with less than 100 developers. Since the number of calculation is the factorial of the developer team size, this number is too huge to be practical when the team size is greater than 100. Further improvement of the algorithm is necessary such that the software is practical for large companies. One direction would be applying parallel computing technique to the algorithm. Multicore central processing unit becomes very common now. We could separate the task into a few parts and process them the same time. By this way we could speed up the algorithm.

The other improvement direction would be applying the file dependency idea into the generating of the file-file matrix. In the successor candidate selecting algorithm we simply assume that the file dependency is always the same between 2 files. But this is not always the case. Then how the difference of dependency affects the developer knowledge need to be researched. And also the algorithm need to be modified according to the research result.

## 4. SCHEDULE

Please check Table1 for the detail of the action schedule.

Table 1: Action Schedule

Number of Days required	Task
2	Select the programming language and the operating system
7	recruit developers
7	develop the database building function
7	implement the optimized truck factor algorithm
3	develop the result output function
1	integrating test
7	bugs fixing if any (If no bugs, we don't need to do this task)
7	implement the successor candidate selecting algorithm
1	integrating test
7	bugs fixing if any (If no bugs, we don't need to do this task)
7	integrating all developed function unit and design UI
2	final test
7	bugs fixing if any (If no bugs, we don't need to do this task)
1	first release



## 5. BUDGET

Please check Table2 for the detail of the project budget.

Table 2: Project Budget

Resource	Budget
2 programmers working 3 hours per day	\$7920 on total
2 computers	\$2000 on total
Total	\$9920

## 6. QUALIFICATIONS

**Yue Cai Zhu** Cell-phone: 438-878-9302 Email : zhuyuecai@gmail.com

### EDUCATION

- |                                                                                                                                   |                            |
|-----------------------------------------------------------------------------------------------------------------------------------|----------------------------|
| <b>Bachelor degree, Computer Science option Software System</b><br>Concordia University, Montreal QC - Canada                     | <b>May 2012 – Sep 2016</b> |
| <ul style="list-style-type: none"> <li>Member, Institute for Co-operative Education</li> </ul>                                    |                            |
| <b>Bachelor degree, International Commerce and Economics</b><br>Southwest University of Political Science & Law, Chongqing, China | <b>Sep 2002 – Jul 2006</b> |

### LANGUAGES

**Fluent in English, Mandarin and Cantonese**

### COMPUTER SKILLS

- Operating Systems** : Windows XP , Windows 7, Linux
- Applications**: Word, Excel, PowerPoint, Latex, Postgresql
- Programming languages** : Java, Python, C++, R, sql

### WORK EXPERIENCE

- |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|
| <b>Research Assistant, Computer Science Department</b><br>Concordia University (Montreal Canada)                                                                                                                                                                                                                                                                                                                                                                                                                                           | <b>May 2014 – now</b>      |
| Responsibilities: <ul style="list-style-type: none"> <li><i>clean the data for further analyze</i></li> <li><i>Build math model on demand</i></li> <li><i>Help design algorithm</i></li> </ul> Special achievement: <ul style="list-style-type: none"> <li><i>Contributed a math proof and a combination picking algorithm for the optimized Truck Factor algorithm in the paper "Developer Turnover and Succession: Empirical studies of the susceptibility and damage from developer turnover" (submitted to ICSE 2015)</i></li> </ul>   |                            |
| <b>IT Analyst, Asset Strategy Department</b><br>Hydro One Networks Inc.(Toronto Canada)                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <b>Jan 2014 – May 2014</b> |
| Responsibilities: <ul style="list-style-type: none"> <li><i>Analyzed electricity equipment usage data</i></li> <li><i>Develop Excel applications for internal use</i></li> <li><i>Test the new algorithm of the new asset management system</i></li> </ul> Special achievement: <ul style="list-style-type: none"> <li><i>A small VBA graphing library for Excel</i></li> <li><i>An Excel macro in VBA to simulate the new asset management System</i></li> <li><i>Three test reports on the new asset management algorithm</i></li> </ul> |                            |

### HONOURS, AWARDS, & DISTINCTIONS

- |                                                                     |                 |
|---------------------------------------------------------------------|-----------------|
| <b>Campaign for the New Millennium Student Contribution Bursary</b> | <b>Jan 2013</b> |
| <b>Faculty of Arts and Science Scholar for 2012-2013</b>            | <b>Oct 2013</b> |
| <b>Faculty of Arts and Science Dean's List</b>                      | <b>Nov 2013</b> |
| <b>Undergraduate Student Research Award</b>                         | <b>Jun 2014</b> |

## 7. PRESS RELEASE

KICKFF HELPS REDUCE LABOR COST BY PROVIDING A SAFE DEVELOPER LAYOFF SOLUTION FOR SOFTWARE COMPANIES.

**Concordia University, November 19, 2014 TODO**

A group of Concordia University students have developed a software application called Kickoff to help software companies control their labor cost. Based on the desired salary cut off amount, it can give out a programmer layoff list that is the least harmful to the software project and an abandoned source code take over solution.

Laying off developer is dangerous to software companies. Since the part of project files maintained only by the layoff developers will become unknown to the company. If there was a bug reported in those files, nothing could be done until the remain developers learned how to maintain them. The learning process is time consuming, but the customer may select other product if the bug could not be fixed in a short time.

However, In order to survive in a critical financial situation, software companies sometimes have to lay off developers to reduce the labor cost. Laying off senior developers cuts off more cost but lose more project files than laying off junior developers. Therefore deciding the layoff solution is difficult for software companies.

Kickoff first generates a set of possible solutions according to the target cutoff cost, then applies the "Optimized Truck Factor" algorithm to find out the least harmful solution and uses the "candidate finding system" to generate the project remedy solution.

With Kickoff, it is easier for a software company to cut off their labor cost.

## 8. CONCLUSION

It is harmful for a software project when there are some developers leaves. The damage is manifested as abandoned files on the software system that can be difficult to maintain.

We believe that this damage can be reduced by finding experienced developers who have worked with the files that have co-changes with the abandoned files before to take over the abandoned files or to mentor new joiners to do so. Therefore we propose to apply the optimized truck factor algorithm to find out the abandoned files, and the candidate selecting algorithm to find out the experienced developers to build up the software Kickoff.

With Kickoff, a software company can understand how harmful a particular lay-off list is and how to reduce the damage. With Kickoff, a software company can also obtain information about all possible layoff solution by inputting the critical variable. Kickoff will be a very useful tool in software company management.

In this proposal we've already scratched out the conceptual structure of it and discussed on its further improvement.

# References

- [1] BIRD, C., GOURLEY, A., DEVANBU, P., SWAMINATHAN, A., AND HSU, G. Open borders? immigration in open source projects. In *MSR: Proceedings of the Fourth International Workshop on Mining Software Repositories* (2007), IEEE Computer Society, p. 8.
- [2] CANFORA, G., DI PENTA, M., OLIVETO, R., AND PANICHELLA, S. Who is going to mentor newcomers in open source projects? 44.
- [3] CATALDO, MARCELO AND WAGSTROM, P. A., HERBSLEB, J. D., AND CARLEY, K. M. Identification of coordination requirements: implications for the design of collaboration and awareness tools. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (New York, NY, USA, 2006), CSCW'06, ACM, pp. 353–362.
- [4] GUTHRIE, J. P. High-involvement work practices, turnover, and productivity: Evidence from new zealand. *Academy of management Journal* 44, 1 (2001), 180–190.
- [5] MOCKUS, A. Organizational volatility and its effects on software defects. In *Proceedings of the eighteenth ACM SIGSOFT international symposium on Foundations of software engineering* (New York, NY, USA, 2010), FSE '10, ACM, pp. 117–126.
- [6] RICCA, F., MARCHETTO, A., AND TORCHIANO, M. On the difficulty of computing the truck factor. In *Proceedings of the 12th International Conference on Product-focused Software Process Improvement* (Berlin, Heidelberg, 2011), PROFES'11, Springer-Verlag, pp. 337–351.
- [7] TORCHIANO, M., RICCA, F., AND MARCHETTO, A. Is my project's truck factor low?: theoretical and empirical considerations about the truck factor threshold. In *Proceedings of the 2nd International Workshop on Emerging Trends in Software Metrics* (New York, NY, USA, 2011), WETSoM '11, ACM, pp. 12–18.
- [8] ZHOU, M., AND MOCKUS, A. Does the initial environment impact the future of developers? In *Proceedings of the 33rd International Conference on Software Engineering* (New York, NY, USA, 2011), ICSE '11, ACM, pp. 271–280.

## The Optimized TF Algorithm

The ‘truck factor’ is a measure in the agile community of the number of developers that must leave (EX: get hit by a truck) before the project fails [7]. Previous work has calculated the maximum number of files that are lost when a group of developers leave. Since this calculation has a  $O(n!)$ , they are able to examine only small projects at one single point in time. We contribute a stopping condition that allows us to calculate the maximum loss quickly, so that we can examine the truck factor on large projects. We prove that we attain the optimal solution and show how the truck factor for Chrome and Linux changes over time. We compare the maximum loss to the actual loss that Chrome and Linux have experienced suggesting how likely they are to suffer from similar turnover events.

The naive TF algorithm proposed by Ricca et al [6] has a high time complexity. When  $n$  is the size of developer team and  $m$  is the total number of files in the project then the time complexity is given by:

$$T(n, m) = n * \sum_{i=1}^n \frac{n!}{i!(n-i)!} * m \quad (1)$$

For example, if a project has 30 developers then the number of developer combinations that is over 17 million. In previous work, the largest project had only 38 developers [7]. We are considering much larger projects, for example, Chrome and Linux have 1000’s of contributors. It is impractical to compute all developer combination. We introduce a stopping condition which we prove identifies, for a given group size, the set of developers who’s loss will result in the maximum file loss.

### Truck factor definitions

The following is the definition of symbols which will be used later:

$D$  = the set of all developers in the project.

$d_i$  = a particular developer in  $D$ ,  $i$  is the ID of the developer.

$F$  = the set of all files in the project.

$f_j$  = a particular file in  $F$ ,  $j$  is the ID of it.

$M(f_j)$  = the set of developers who have modified the file  $f_j$ .

$I()$  is the logic function defined as:

$$I(condition) = \begin{cases} 1 & \text{if condition is true} \\ 0 & \text{otherwise} \end{cases}$$

We first calculate the proportion of developer  $d_i$  who have modified the file  $f_j$  by equation:

$$L(d_i, f_j) = \frac{1}{|M(f_j)|} * I(d_i \in M(f_j)) \quad (2)$$

Then the file loss (FL) function returns how many files would be abandoned if a given developer combination  $C$  left the project:

$$FL(C) = \sum_{d_i \in C} \sum_{f_i \in F} (I(M(f_i) \subseteq C) * L(d_i, f_i)) \quad (3)$$

Then the maximum file loss for a given group size is returned by the truck factor function:  $TF(C) = \max(FL(C))$ . We can see that a naive implementation of this algorithm will have a complexity of  $O(n!)$ , which is impractical to compute for large projects.

### Stopping Condition

The shared proportion of the number of files a developer  $d_i$  has modified on the project:

$$L(d_i) = \sum_{f_j \in F} L(d_i, f_j) \quad (4)$$

The upper bound of the file loss for a given developer combination  $C$  is:

$$UFL(C) = \sum_{d_i \in C} L(d_i) \quad (5)$$

Since

$$\forall condition : I(condition) \leq 1$$

We have:

$$FL(C) \leq \sum_{d_i \in C} \sum_{f_i \in F} [1 * L(d_i, f_i)] = UFL(C) \quad (6)$$

For each group size, we order the developer combinations by their  $UFL$  value and calculate the  $FL$  for each developer combination until the stopping condition,  $UFL(C) < \max(FL)$  is met. In other words, we have shown that the number of files that a given group of developers modify,  $UFL$ , will be less than or equal to the number of files that these developers own exclusively,  $FL$ . Provided that we order the developers by the number of files that they modify,  $UFL$ , we can stop when the maximum loss that we calculated from the  $FL$  function is greater than or equal to the total number of files that the subsequent group of developers modify.