On Software Modernisation due to Library Obsolescence

Simos Gerasimou
Department of Computer Science
University of York
York, UK
simos.gerasimou@york.ac.uk

Maria Kechagia
Software Engineering Research Group
Delft University of Technology
Delft, The Netherlands
m.kechagia@tudelft.nl

Dimitris Kolovos
Department of Computer Science
University of York
York, UK
dimitris.kolovos@york.ac.uk

Richard Paige
Department of Computer Science
University of York
York, UK
richard.paige@york.ac.uk

Giorgos Gousios
Software Engineering Research Group
Delft University of Technology
Delft, The Netherlands
g.gousios@tudelft.nl

ABSTRACT
Software libraries, typically accessible through Application Programming Interfaces (APIs), enhance modularity and reduce development time. Nevertheless, their use reinforces system dependency on third-party software. When libraries become obsolete or their APIs change, performing the necessary modifications to dependent systems, can be time-consuming, labour intensive and error-prone. In this paper, we propose a methodology that reduces the effort developers must spend to mitigate library obsolescence. We describe the steps comprising the methodology, i.e., source code analysis, visualisation of hot areas, code-based transformation, and verification of the modified system. Also, we present some preliminary results and describe our plan for developing a fully automated software modernisation approach.

CCS CONCEPTS
• Software and its engineering → Software libraries and repositories; Software evolution; Maintaining software;

KEYWORDS
application programming interfaces; software libraries; library evolution; software modernisation; visualisation

1 INTRODUCTION
Modern software development practices advocate the use of third-party libraries as a means of improving maintainability, usability and dependability of software systems [7]. Many of these libraries aggregate functions and services, and expose endpoints, i.e., Application Programming Interfaces (APIs), that enable interacting with the logic of the libraries. An API is considered an implicit agreement between a software system (client) and a third-party library (provider). Based on this agreement, software developers can integrate the library within their client systems and use the provided functions without needing to be aware of the underlying implementation logic.

Despite the benefits accompanying the use of third-party libraries and their APIs, a strong dependency link is created with software systems using these libraries [2]. Since software systems and the libraries they are based on evolve independently, maintaining the system to a fully-functional state requires additional effort when changes to the libraries’ APIs violate the agreement. For instance, library evolution caused by new requirements or architectural changes can lead to API modifications that break compliance with client systems. Likewise, end-of-support of a software library can cause client systems to fall into “technology stagnation” unless they switch to another functionally-equivalent and actively maintained library. Similarly, the introduction of a competing library with improved functionality, better features and reduced overheads might urge software developers to consider adopting this new library in subsequent system versions.

In order to avoid the imminent risks that arise from using obsolete libraries (e.g., unresolved bugs, susceptibility to cyber attacks due to security issues), developers typically sustain the effort of modernising their systems [13]. Nevertheless, modifying the system to start using a new library is a time-consuming, error-prone and, to a great extent, developer-driven task [2, 5]. Performing correctly this task includes identifying system instructions that must change, doing the necessary code changes and, finally, checking that the modernised system preserves its original functionality. As the size of the client system and the legacy third-party library increases, the effort required for performing these steps increases significantly [7].

We argue that developers can reduce significantly the manual effort spent for mitigating library obsolescence by adopting concepts and approaches from the areas of software visualisation and bug finding [5, 12]. For instance, visualisation paradigms can help developers to understand better the architecture of their systems and perform impact analysis. Software visualisation is an active research area, but with rather limited adoption in software development [8, 12]. Also, code-based and pattern-based transformation approaches can assist with automating code refactoring [2].

In this paper, we propose an end-to-end methodology that not only helps developers to assess the required effort, but also automates a great part of the modernisation task. In particular, the methodology involves analysis of the software system’s source code to extract hotspots, i.e., source code elements that invoke the legacy library, visualisation of these hot areas using the city metaphor [12], automatic source code transformation based on templates (patterns), population of the generated templates with suitable code, and, finally, verification of the software system and the transformed code...
to establish that the functional and non-functional requirements of the resulting system are maintained.

The remainder of the paper is organised as follows. In Section 2 we present an example for motivating our approach. In Section 3 we introduce the proposed approach for mitigating library obsolescence and describe a set of preliminary results. Finally, in Sections 4 and 5, we present related work and outline our conclusions and plans for future work, respectively.

2 MOTIVATING EXAMPLES

We motivate our work using two examples in which library evolution introduced compatibility-breaking changes. The first example comes from TinyXML\(^2\), the new version of the popular C++ XML parsing library. Even though the evolved library is more CPU and memory efficient than its predecessor, it is heavily redesigned and does not maintain backward compatibility. Figure 1 shows a relevant issue\(^3\) from the GitHub repository of the TinyXML project. These obsolescence issues caused modernisation problems to developers of client systems. Some developers performed the migration manually while others preferred to switch to another functionality-equivalent library; for instance, the FileZilla client (https://filezilla-project.org) started using pugiXML from version 12 onwards. Irrespective of the chosen solution, the effort required for updating the system was substantial (e.g., see FileZilla development diary\(^5\)).

Second, we refer to the Lighttpd web-server library which is used by several high-traffic websites (e.g., Wikipedia, YouTube). Research in [4] found that an one-line change to one of the methods in the updated library (Listing 1), broke support for HTTP compression and crashed any client applications. Despite the minimal API changes, this new bug issue could push developers to undertake the effort and migrate their system to an alternative and more reliable library.

Clearly, library and API changes affect the stability of client systems. We aim to assist developers by identifying the affected source code in client systems and simplifying the migration process.

Listing 1: Lighttpd before and after update (as in [4])

```c
// Lighttpd before update
for (h=0, i=0; i < etag->used; ++i)
    h = (h << 5)^(h >> 27)^(etag->ptr[i]);
// Lighttpd after update
for (h=0, i=0; i < etag->used - 1; ++i)
    h = (h << 5) ^ (h >> 27)^(etag->ptr[i]);
```

\(^1\)http://www.grinninglizard.com/tinyxml2/index.html

\(^2\)https://github.com/lethomason/tinyxml2/issues/440

\(^3\)https://tinyurl.com/y7myt4n4
languages (e.g., C/C++, Java, C#) come with publicly available and ready-to-use infrastructures that help with this task; see for instance, Eclipse CDT for C/C++ and Roslyn for C#. We use Eclipse CDT for our prototype implementation (Section 3.2).

Stage 2: Visualisation of software metrics

Manual source code inspection is challenging for complex systems with multiple dependencies, packages, classes, and methods (e.g., FileZilla has more than 300 C/C++ files and more than 130 KLoC). To facilitate reasoning and effort estimation, we automatically generate interactive visual artifacts using the metrics produced during analysis. These artifacts provide a pictorial view of the system and help with system understanding.

The benefits of using suitable visual metaphors both for assisting with software exploration and for reducing the cognitive load is widely accepted [1, 8]. Several visual metaphors have been proposed for representing static aspects of software (cf. Section 4). To the best of our knowledge, however, there is no prior research on visualising the dependencies of a software system on third-party libraries and using this information to guide software modernisation.

We bridge the gap in existing research by adapting the semantics of available visualisation metaphors (e.g., Treemap, Code City [12]) to this problem. In our preliminary realisation (Section 3.2), we use Code City to represent software systems as three-dimensional cities with districts and buildings, and associate the extracted dependency metrics with visual properties of city components.

The development team can study the produce interactive visual artifacts and carry out a risk analysis. The outcome of this study indicates how to best address the obsolescence issues. An estimate about the effort required for completing the modernisation can be also made. For instance, if extensive coupling is identified, it might be more sensible to investigate how to reduce its degree (e.g., by improving the system architecture) before proceeding with the remaining modernisation stages.

Stage 3: Code-based transformation

During this stage, the software system undergoes a three-step transformation in order to become compliant with the new library. The first step includes (1) the automated generation of an abstraction layer (e.g., using the adapter pattern) and its population with elements that delineate the usage of the obsolete library by the software system; and (2) the automated modification of commands in the affected system files to delegate the work to the generated abstraction layer instead of the obsolete library. In this way, our methodology minimises changes in the system’s source code.

The next step involves the inference of mappings between the obsolete and new libraries [3]. A mapping rule comprises two sets of API commands that perform the same task, one from the obsolete library and the other from the new library. The outcome of this inference step is a list of mappings.

Inferring likely API mappings is challenging and time-consuming, and its automation has been the focus of recent research [11]. The selection of a suitable technique depends on the characteristics of the considered libraries and the expertise of developers. For instance, static analysis and textual similarity techniques [9] can be used when the APIs of the libraries are, to a degree, similar. Alternatively, developers can resort to inspecting the documentation of these APIs and generating manually the list of mappings.

Listing 2: Code fragment that maps the get attribute method in TinyXML to the equivalent methods in PugiXML

```c
const char *XMLElement::Attribute(const char *name, const char *value) const {
    return pugiXmlNode.attribute(name).value();
}
```

The final step involves populating the generated abstraction layer with suitable code fragments that invoke the new library. To achieve this, developers use the extracted list of mappings and write code within the placeholders in the abstraction layer. For instance, Listing 2 shows the TinyXML method (placeholder) for extracting the attribute from an XML element and the corresponding code fragment in PugiXML (https://pugixml.org); this relationship is one-to-many. Once this is done, the functionality that was previously done by the obsolete library is now undertaken by the new library.

The level of difficulty for writing the relevant code fragments depends on the correspondence between the obsolete and the new library. Investigating the time and effort required for completing this task based on this correspondence is part of our future work.

Stage 4: Verification and validation

The next stage involves checking that both the functional and non-functional requirements of the evolved system (transformed and unaffected code) continue to hold. This includes running unit, integration and system tests, or any other type of formal verification (e.g., use model checking to verify the absence of concurrency bugs).

Special focus should be given to the amended system parts, i.e., the abstraction layer and affected components, as these parts are most likely to have introduced erroneous behaviour. The visual artifacts generated earlier can help developers to extract traceability information and decide where to spend most of their efforts.

3.2 Preliminary Realisation

We present a prototype realisation of our methodology which is currently under development as an Eclipse plugin. To evaluate our approach, we use the FileZilla client v.11, a mature open-source C/C++ application, with a non-trivial code base and several functional and non-functional requirements. FileZilla uses a number of third-party libraries, including TinyXML for reading and updating XML files that keep server sites and interface-related properties.

At first, we obtained the ASTs of the FileZilla source code using the parsing facilities provided by Eclipse CDT. The automated analysis of these ASTs enabled the identification of hotspots, i.e., FileZilla code elements that access the legacy XML library. We also gained insight into the usage level of this library by the system.

We adapted the code city metaphor [12] to visualise the information extracted during the analysis stage. Figure 3 shows a representation of FileZilla using our adapted city metaphor. We employ the basic metaphor semantics, i.e., represent an application as a city, a package/sub-package as a district/sub-district, and a class as a building. However, we introduce new semantics that help visualising the dependency level between the software system and the obsolete library, and capturing the hotspots. First, we use red-coloured buildings to show a class that invokes the obsolete library, i.e., an affected class. Second, we set the height of a red-coloured building
Recent research has been focused on identifying API Evolution. with FileZilla and confirmed that the functionality and properties to improve software robustness. associated with mitigating library obsolescence, we develop a visualisation insights from previous work and considering the challenges. Raemaekers et al. proposed a framework of metrics that can empirical studies with real software projects, and they reported changes in used api. 4 RELATED WORK abstraction layer has been generated and populated correctly, and of the system still hold. This assessment provided insight that the inference and code population is part of our future work. Given the modernised system, we used the test suite coming with FileZilla and confirmed that the functionality and properties of the system still hold. This assessment provided insight that the abstraction layer has been generated and populated correctly, and also that the system started exercising the new library. 4 RELATED WORK API Evolution. Recent research has been focused on identifying issues that can break client applications’ source code, because of changes in used APIs. Jazek et al. studied this problem in real word programs and argued that better tools and methods need to support library updates in client applications [5]. Xavier et al. also investigated API changes that may break previously established contracts, resulting into software crashes [13]. They conducted empirical studies with real software projects, and they reported lessons for better library support and maintainability of client applications. Raemaekers et al. proposed a framework of metrics that can be used to measure the stability of a software library [10]. Based on insights from previous work and considering the challenges associated with mitigating library obsolescence, we develop a visualisation approach that can assist developers of client applications to improve software robustness. Visualisation Techniques. Software visualisation techniques are widely used in program comprehension; see for instance the recent survey in [8] about existing software visualisation approaches. This survey highlights the number of visualisation tools for supporting developers’ needs on dependency management. A few studies have attempted to provide visualisations of evolving software systems and their library dependencies. Kula et al. developed a heat-map metaphor, which can help maintainers to navigate to library dependencies and gives an overview of the users across the different versions of a library [6]. Wettel et al. have used the city metaphor to represent software projects and they conducted a controlled experiment with developers to examine how the subjects comprehend the structure of a program [12]. In addition to the existing related work, we are using the city metaphor to highlight areas of software projects using third-party libraries. 5 CONCLUSIONS The increasing dependency of software systems on third-party libraries challenges system maintainability. Developers need guidance on identifying and modernising client source code that can be affected due to library obsolescence. We proposed a modernisation methodology that can help developers of client systems to maintain their source code when these obsolescence issues occur. Also, we introduced a visualisation approach, using the city metaphor, for visualising source code areas that invoke the legacy library. We demonstrated a prototype realisation of our methodology and applied it on the FileZilla client. In the future, we plan to provide a mechanism that can automatically adapt client source code to new versions of third-party libraries. We will also explore the automated test generation for the changed parts of client source code. REFERENCES [1] C. R. B. de Souza and D. L. M. Bentolila. 2009. Automatic evaluation of API usability using complexity metrics and visualizations. In ICSE'09 - Companion Volume. 299–302. [2] Danny Dig and Ralph Johnson. 2005. The Role of Refactorings in API Evolution. In ICSE'05. 389–398. [3] A. Gokhale, V. Ganapathy, and Y. Padmanaban. 2013. Inferring likely mappings between APIs. In ICSE'13. 82–91. [4] Petr Hosek and Cristian Cadar. 2013. Safe Software Updates via Multi-version Execution. In ICSE'13. 612–621. [5] Kamil Jezek, Jens Dietrich, and Premek Brada. 2015. How Java APIs break - An empirical study. Information and Software Technology 65 (2015), 129–146. [6] Raula Gaikovina Kula, Coen De Roover, Daniel German, Takashi Ishio, and Katsuro Inoue. 2014. Visualizing the Evolution of Systems and Their Library Dependencies. In VISSOFT'14. 127–136. [7] Ralf Lammel, Ekaterina Pek, and Jurgen Starek. 2011. Large-scale, AST-based API-usage Analysis of Open-source Java Projects. In SAC'11. 1317–1324. [8] Leonel Merino, Mohammad Ghaafari, and Oscar Nierstrasz. 2017. Towards actionable visualization for software developers. JSEP’17 (2017). [9] Hoan Anh Nguyen, Tung Thanh Nguyen, Gary Wilson, Jr., Anh Tuan Nguyen, Miryung Kim, and Tien N. Nguyen. 2010. A Graph-based Approach to API Usage Adaptation. In OOPSLA'10. 101–321. [10] S. Raemaekers, A. van Deursen, and J. Visser. 2012. Visualising software library stability through historical version analysis. In ICSE’12. 378–387. [11] Martin P. Robillard, Eric Bodden, David Kawrykow, Mira Mezini, and Tristan Ratchford. 2013. Automated API Property Inference Techniques. TSE 39, 5 (2013), 613–637. [12] Richard Wettel, Michele Lanza, and Romain Robbes. 2011. Software Systems As Cities: A Controlled Experiment. In ICSE'11. 551–560. [13] L. Xavier, A. Brito, A. Hora, and M. T. Valente. 2017. Historical and impact analysis of API breaking changes: A large-scale study. In SANER'17. 138–147.