ATDx: Building an Architectural Technical Debt Index

Roberto Verdecchia¹, Patricia Lago¹, Ivano Malavolta¹, and Ipek Ozkaya²

¹Vrije Universiteit Amsterdam, The Netherlands
²Software Engineering Institute, Carnegie Mellon University, USA
{r.verdecchia, p.lago, i.malavolta}@vu.nl, ozkaya@sei.cmu.edu

Keywords: Software Architecture, Technical Debt, Software Analytics, Software Metrics, Software Maintenance

Abstract: Architectural technical debt (ATD) in software-intensive systems refers to the architecture design decisions which work as expedient in the short term, but later negatively impact system evolvability and maintainability. Over the years numerous approaches have been proposed to detect particular types of ATD at a refined level of granularity via source code analysis. Nevertheless, how to gain an encompassing overview of the ATD present in a software-intensive system is still an open question. In this study, we present a multi-step approach designed to build an ATD index (ATDx), which provides insights into a set of ATD dimensions building upon existing architectural rules by leveraging statistical analysis. The ATDx approach can be adopted by researchers and practitioners alike in order to gain a better understanding of the nature of the ATD present in software-intensive systems, and provides a systematic framework to implement concrete instances of ATDx according to specific project and organizational needs.

1 INTRODUCTION

Architectural Technical Debt (ATD) in a software-intensive system denotes architectural design choices which, while being suitable or even optimal when adopted, lower the maintainability and evolvability of the system in the long term, hindering future development activities [Li et al., 2015b]. With respect to other types of debt, e.g., test or build debt [Li et al., 2015a], ATD is characterized by being widespread throughout code-bases, mostly invisible [Kruchten et al., 2012], and of high remediation costs.

Due to its impact on software development practices, and its high industrial relevance, ATD is attracting a growing interest within the scientific community and software analysis tool vendors [Verdecchia et al., 2018]. Notably, over the years, numerous approaches have been proposed to detect, mostly via source code analysis, ATD instances present in software-intensive-systems. Such methods rely on the analysis of symptoms through which ATD manifests itself, and are conceived to detect specific types of ATD by adopting heterogeneous strategies, ranging from the inspection of bug-prone components [Xiao et al., 2016], to the analysis of dependency anti-patterns [Arcelli Fontana et al., 2017], maintainability issues [Malavolta et al., 2018], and the evaluation of components modularity [Martini et al., 2018b]. Additionally, numerous static analysis tools, such as NDepend, CAST, and SonarQube are currently available on the market, enabling to keep track of technical debt and architecture-related issues present in code-bases. Both academic and industrial approaches currently focus on fine-grained analysis techniques, considering ad-hoc definitions of technical debt and software architecture, in order to best fit their analysis processes. Nevertheless, to date, how to gain an encompassing overview of the (potentially highly heterogeneous) ATD present in a software-intensive system is still an open question.

In order to fill this gap, in this study we present ATDx, an approach designed to provide data-driven, intuitive, and actionable insights on the ATD present in a software-intensive system. ATDx consists of a theoretical, multi-step, and semi-automated process, concisely entailing (i) the inclusion of architectural rules supported by pre-existing analysis tools, (ii) the identification of outlier values (after normalization) via statistical analysis, and (iii) the aggregation of analysis results into a set of customizable ATD dimensions. ATDx is meant to serve two types of stakeholders: (i) researchers conducting quantitative studies on source-related ATD and (ii) practitioners carrying out software portfolio analysis and management, to suitably detect ATD items and get actionable insights.
We discuss a set of benefits and drawbacks which characterize the current ATDx definition, which were identified from a viability investigation of ATDx. Among other, the most relevant characteristics of ATDx are: analysis tool and programming language independence, (ii) data-driven results at multiple levels of granularity, (iii) extensibility, and (iv) customizability to specific application domains.

The main contributions of the paper are: (i) the ATDx approach for building a multi-level index of ATD, (ii) the description of the process for building an instance of ATDx, and (iii) a thorough discussion of the limitations and drawbacks of ATDx.

The remainder of the paper is structured as follows. In the next section, we present the theoretical framework underlying the ATDx approach. In Section 2 we document the steps entailed in the implementation of a concrete instance of ATDx. Section 3 discusses the main benefits and drawbacks of the ATDx approach, which emerged from a viability investigation. In Section 4 the related work to our study is presented. Finally, Section 5 reports on the conclusions of the study and the future work.

2 THE ATDx APPROACH

In this section we present the definitions on which the calculation of ATDx relies (Section 2.1), followed by the description of the approach (Section 2.2).

2.1 Definitions

Definition 1. Architectural rule. Given a source code analysis tool T and its supported analysis rules \( R^T \), the architectural rules \( AR^T \) supported by T are defined as the subset of all rules \( R^T_i \in R^T, i = \{1, \ldots, |R^T|\} \) such that:

- \( R^T_i \) is relevant from an architectural perspective;
- \( R^T_i \) is able to detect a technical debt item, i.e., a design or implementation construct that is expensive in the short term, but set up a technical context that can make future changes more costly or impossible [Avgeriou et al., 2016].

We consider every \( AR_i^T \) as a function \( AR_i^T : E \rightarrow \{0,1\} \), where \( E \) is the set of architectural elements according to a granularity level (see below). In case that an element \( e \in E \) violates rule \( AR_i^T \), then \( AR_i^T(e) \) will return 1, and 0 otherwise.

For example, a rule \( AR_i^T \) checking if the java.lang.Error class is extended in a Java system is (i) architectural since it predicates on the high-level design of Java-based systems, and (ii) related to technical debt since the design of the Java Virtual Machine (JVM) assumes that java.lang.Error and its subclasses are managed only by the JVM.

Definition 2. Granularity level. Given an architectural rule \( AR_i^T \), its granularity level \( Gr_i^T \) is defined as the smallest unit of the system being analysed which may violate \( AR_i^T \), e.g., a class, a method, or a line of code. As an example, if we consider a rule which deals with cloned classes, its corresponding granularity level is “class”.

The ad-hoc mapping of design rules to different granularity levels enables us to evaluate and compare the occurrence of rules violations across different software projects at a refined level of precision, instead of trivially adopting a single metric for the size of software projects for all the rules in \( AR^T \), e.g., source lines of code (SLOC).

Definition 3. ATD Dimension. Given a set of architectural rules \( AR^T \), the set of ATD dimensions \( ATDD^T \) represents a set of clusters of architectural rules \( AR_i^T \subseteq AR^T \) with similar focus. One architectural rule \( AR_i^T \) can belong to one or more ATD dimensions \( ATDD_j^T \), \( ATDD_j^T \subseteq ATDD^T \) and the mapping between \( AR_i^T \) and \( ATDD_j^T \) is established by generalizing the semantic focus of \( AR_i^T \).

For example, if an architectural rule \( AR_i^T \) deals with the conversion of Java classes into Java interfaces, the \( AR_i^T \) could fall under the general “Interface” ATD dimension (for further details on the identified ATD dimensions and their establishment, refer to Section 3.2).

We use the 3-tuple \( \langle AR_i^T , ATDD_j^T , Gr_i^T \rangle \), as the mapping of each architectural rule \( AR_i^T \) to its granularity level \( Gr_i^T \), and one or more ATD dimensions \( ATDD_j^T \). It is important to note that, while an \( AR_i^T \) can be associated to one and only one granularity level \( Gr_i^T \), an \( AR_i^T \) can be mapped to multiple \( ATDD_j^T \)s, and vice versa.

2.2 Approach Description

ATDx aims to provide a birds-eye view of the ATD present in a software-intensive system by analyzing the set of architectural rules \( AR^T \) supported by an analysis tool T, and subsequently aggregating the analysis results into different ATD dimensions \( ATDD^T \). Intuitively, starting from an already available large-scale dataset of \( AR^T \) executions NORMT and a System Under Analysis SUA, ATDx performs a statistical analysis of all elements in NORM for detecting anomalous occurrences of architectural rule violations in SUA. If the SUA exhibits an anomalous number of violations of a certain rule \( AR_i^T \in AR^T \), then the corresponding ATD dimension \( ATDD_i^T \in ATDD^T \) mapped to \( AR_i^T \) is incremented. In a sense, ATDx represents the synthesized facets of ATD (derived from
the architectural rules in $AR^T$, rather than the total ATD present in the system.

$ATD_x$ leverages the calculation of the number of architectural rule violations of a System $S$ over the size of $S$ in order to compare the occurrence of rule violations across projects of different sizes. Specifically, for each $AR^T_i$, we first calculate the set of violations of $AR^T_i$ in a system by defining $AR^T_i(S)$ as the union of all violations of $AR^T_i$ in $S$, i.e.,

$$AR^T_i(S) = \bigcup_{e \in Gr^T_i(S)} ar^T_i(e)$$

where $Gr^T_i(S)$ is the set of all elements $e$ in $S$ according to the granularity $Gr^T_i$ (e.g., the set of all classes in a Java-based system), and $ar^T_i(e)$ is a function returning $1$ if the element $e$ violates the architectural rule $AR^T_i$, the empty set otherwise.

Subsequently, we calculate $NORM^T_i(S)$, defined as the normalized number of architectural rule violations $|AR^T_i(S)|$ over the total number of elements $|e|$ according to granularity $Gr^T_i$, i.e.,

$$NORM^T_i(S) = \frac{|AR^T_i(S)|}{|Gr^T_i(S)|}$$

where $|Gr^T_i(S)|$ is the cumulative size of $S$ expressed according to granularity level $Gr^T_i$, and $|AR^T_i(S)|$ is the total number of violations of rule $R^T_i$ (see Formula[1]).

Once the $NORM^T_i(S)$ for $S$ is calculated, we statistically detect if it exhibits an anomalous value. In order to do so, we require the set $NORM^T_i$, which contains the values of $NORM^T_i(S)$ for each software project belonging to a pre-defined set of software projects, i.e.,

$$NORM^T = \{NORM^T_i(S_1), \ldots, NORM^T_i(S_n)\}$$

where $n$ is the total number of projects belonging to the dataset of software projects considered. Given the calculation of $NORM^T_i$, we can identify if $NORM^T_i(S)$ constitutes an anomalous measurement by comparing its value with the other ones contained in $NORM^T_i$. Specifically, given the set of values $NORM^T_i$ and the value of $NORM^T_i(S)$, we define the function $outlier$ as: $outlier: X^n \times [0, 1] \rightarrow \{0, 1\}$, where $X = [0, 1]$; $outlier$ returns $1$ if $NORM^T_i(S)$ is greater than the upper inner fence of $NORM^T_i$, and $0$ otherwise[4].

In order to provide an overview of the ATD dimensions of a system $S$, for each ATD dimension $ATDD^T_j \subseteq ATDD^T$, we define the value of $ATDD^T_j(S)$ as the average number of outlier values of the $AR^T_i$ mapped to it, i.e.,

$$ATDD^T_j(S) = \frac{\sum_{i=1}^{n} outlier(\text{NORM}^T_i, \text{NORM}^T_i(S))}{n} \quad (4)$$

where $n$ is the total number of rules in $AR^T$ mapped to $ATDD^T_j$.

Finally, we also define an overall value $ATDD^T(S)$, embodying the overall architectural technical debt of $S$ calculated via our algorithm, as the average value of all the defined ATD dimensions $ATDD^T_i$, i.e.,

$$ATDD^T(S) = \frac{\sum_{i=1}^{n} ATDD^T_i}{n} \quad (5)$$

where $n$ is the total number of ATD dimensions $ATDD^T_i$ considered.

### 3 ATDx BUILDING STEPS

In this section, we document the steps required in order to build $ATD_x$. Specifically, the implementation process documented in this section is presented in a generic fashion, i.e., it is not bound to any specific analysis tool or technology. The described process can be performed by both (i) researchers investigating ATD phenomena, and (ii) practitioners analyzing their software portfolios. In fact, following the described theoretical steps allows to implement the instance of $ATD_x$ which best fits the specific technical, organizational, and tool-related needs considered.

Fig. [1] presents the main steps of the ATDx approach. Given an analysis tool $T$, five steps are required to build an instance of $ATD_x$, namely: (i) the identification of the architectural rules belonging to $AR^T$; (ii) the formulation of the 3-tuples of the form $\langle AR^T_i, Gr^T_i, ATDD^T_j \rangle$; (iii) the execution of $T$ on a set of already available software projects to form the dataset of $AR^T_i(S)$ measurements, (iv) the execution of the $ATD_x$ analysis on the constructed dataset, and (v) the application of the $ATD_x$ approach on the specific system under analysis (SA).

### 3.1 Step 1: Identification of the $AR^T$ set

The first step of the $ATD_x$ building process consists in the identification of a set of architectural rules $AR^T$ on which the calculation of $ATD_x$ can be based. Specifically, given an analysis tool $T$ and its supported analysis rules $R^T$, a manual inspection is carried out in order to assess if the rules qualify as $AR^T$ according to the criteria presented in [Definition 1]. This process can be carried out either by inspecting the concrete implementation of the rules $R^T$ under scrutiny, or by consulting the documentation of $T$, if available.
3.2 Step 2: Formulation of 3-tuples $\langle AR^T_i, ATDD^T_j, GR^T_i \rangle$

After the identification of $AR^T_i$, the 3-tuples $\langle AR^T_i, ATDD^T_j, GR^T_i \rangle$ are established by mapping each $AR^T_i$ to (i) one or more architectural technical debt dimensions $ATDD^T_j$ and (ii) its granularity level $GR^T_i$.

The process of mapping an $AR^T_i$ to its corresponding architectural dimensions $ATDD^T_j$ is conducted by performing iterative content analysis sessions with open coding [Liddell et al., 2010] targeting the implementation or documentation of the rule in order to extract the semantic meaning of the rule. More in details, once the semantic meaning of each rule is well understood, the $AR^T_i$ under scrutiny is labeled with one or more keywords expressing schematically its semantic meaning. Such analysis is carried out in an iterative fashion, i.e., by continuously comparing the potential $ATDD^T_j$ associated to the $AR^T_i$ under analysis with already identified dimensions, in order to reach a uniform final $ATDD^T_j$ set.

Similarly, the process of mapping an architectural rule $AR^T_i$ to its corresponding level of granularity $GR^T_i$ is also carried out via manual analysis of the architectural rule, and subsequently identifying the unit of analysis that the rule considers (e.g., function, class, or file level).\footnote{In the fortunate instance in which the $GR^T_i$ mapped to $AR^T_i$ is explicitly specified in the source from which the rules $R^T$ are gathered, such information should be preferred over a manual inspection of the rule.}

3.3 Step 3: Building the $AR^T(SUA)$ dataset

After the identification of the $AR^T(S)$ set (Step 1), it is possible to build the dataset of $AR^T(S)$ measurements. This process consists of (i) identifying an initial set of projects to be considered for inclusion in the dataset, (ii) carrying out a quality filtering process in order to ensure the soundness of the included projects, and (iii) identifying the $AR^T(S)$ sets and extracting the $|GR^T(S)|$ values of each project included in the dataset. The selection of the initial dataset of projects to be considered for inclusion is a design choice specific to the concrete instance of ATDx. In other words, such choice is dependent on the analysis goal for which ATDx is adopted, the availability of the software projects to be analyzed, and the tool $T$ adopted to calculate the $AR^T(S)$ sets. It is important to bear in mind that, given the statistical nature of ATDx, having a low number of projects in this step would not lead to meaningful ATDx analysis results (for more insights on this point see Section 4).

As for the selection of projects to be considered, the step of carrying out a quality-filtering process on the initial set of projects depends on the setting in which ATDx is implemented. In case that ATDx is deployed in an experimental setting, e.g., by considering open-source software (OSS) projects, this step has to be carried out in order to ensure that no toy software-projects, such as demos or software examples written for educational purposes, are included in the final dataset of projects [Kallianvakou et al., 2016].

After the identification of a final set of projects to be considered for analysis, the $AR^T(S)$ sets are calculated for each system $S$ in the dataset. The execution of such process varies according to the adopted analysis tool $T$. In addition to the $AR^T(S)$ calculation, during this phase also the values $|GR^T(S)|$, calculated for each considered $GR^T_i$, are extracted for each project $S$ in the dataset, as such values will also be needed to carry out subsequent ATDx steps.

3.4 Step 4: ATDx Analysis

Once the $AR^T(S)$ and $GR^T(S)$ sets are calculated for the whole dataset of projects, the analysis presented in Section 3 can be executed. Specifically, the ATDx analysis takes as input $AR^T(S)$ and $GR^T(S)$ sets for a set of software projects, and outputs, for each project, its $ATDD^T(S)$ and $ATDx^T(S)$ values. It is worth noticing that, after a first execution of the ATDx ap-
proach, it is possible to carry out further ATDx analyses on additional projects by relying on the previously formulated 3-tuples $\langle AR^T, ATDD^T, Gr^T \rangle$, and the pre-calculated intermediate values of the ATDx analysis $NORM^T$, as further documented in Algorithm 1.

### 3.5 Step 5: Applying ATDx to a SUA

Finally, after the execution of ATDx on all projects in the dataset, the resulting $ATDD^T$ and $ATDx^T$ values of a specific SUA can be computed.

#### Algorithm 1: Computing $ATDD^T$ dimensions and $ATDx^T$ value for a single SUA

**Input:** SUA, $AR^T$, $NORM^T$, $ATDD^T$

**Output:** $ATDD^T$ (SUA), $ATDx^T$ (SUA)

1. dimensions $\leftarrow$ empty dictionary
2. adtx $\leftarrow$ 0
3. for all dimensions $j$ in $ATDD^T$ do
   4. dimensions[$j$] $\leftarrow$ 0
4. for all rules $AR^T_j$ in $AR^T$ do
   5. violations $\leftarrow$ $|AR^T_j| (SUA)$
   6. normalizedViolations $\leftarrow$ $NORM^T_j (SUA)$
   7. if outlier($NORM^T_j$, normalizedViolations) then
      8. dimensions[$j$] $\leftarrow$ dimensions[$j$] + 1
5. for all entries $j$ in dimensions do
   6. dimensions[$j$] $\leftarrow$ dimensions[$j$] / getNumRules()
7. if outlier normalizedViolations $\leftarrow NORM^T_j (SUA)$
8. for all dimensions $j$ do
   9. adtx $\leftarrow$ adtx + dimensions[$j$]
10. return dimensions, adtx

As shown in Algorithm 1, the computation takes as input 4 parameters: (i) the SUA, the set of rules $AR^T$, the values computed in the previous step $NORM^T$, and the set of dimensions $ATDD^T$. The output of the algorithm is the set of ATD values of the SUA $ATDD^T (SUA)$ (one for each dimension) and $ATDx^T (SUA)$. The $ATDD^T$ values provide support in gaining more insights in the severity of the ATD according to the identified ATD dimensions, while the $ATDx^T$ value provides a unified overview of the relative ATD present in the SUA. After setting up the initial variables for the output to be produced (lines 1-2), the algorithm proceeds by building a dictionary containing an entry for each dimension in $ATDD^T$, with the name of the dimension as key and 0 as value (lines 3-4). Then, the algorithm iterates over each rule in $AR^T$ (line 5) and collects the number of its violations, both raw (line 6) and normalized by the level of granularity of the current rule (line 7). If the normalized number of violations is an outlier with respect to the values collected for the whole dataset of projects (line 8), then the entry of the dimensions dictionary corresponding to the dimension of the current rule is incremented (line 9). For each dimension $j$ we normalize its current value according to the total number of rules belonging to $j$ in order to mitigate the potential effect that the number of rules belonging to the dimension may have, and increment the current $ATDx^T$ with the computed score (lines 10-12). Finally, the $ATDx^T$ value is normalized by the total number of dimensions supported by all $AR^T$ rules (line 13) and both dimensions and $ATDx$ values are returned (line 14).

### 4 DISCUSSION

In order to get further insights into the viability of ATDx, we implemented a prototype of ATDx, built by considering a set of predefined SonarQube rules [Ernst et al. 2017], and a large-scale dataset of over 6.7K software projects. The raw data, analysis scripts, and results of such investigation, accompanied by a companion technical report, is made available online for further consultation. Following, the benefits and drawbacks characterizing ATDx, and/or emerging from the prototype implementation, are reported.

#### 4.1 Approach Benefits

##### 4.1.1 Tool independence

ATDx is a tool-independent approach by design. This allows its future adopters to tailor it according to the specific development context considered, by basing the ATDx calculation on the analysis tools which are deemed more fitting. Specifically, the adoption of ATDx, rather than imposing a technological constraint on its own tools and practices, is intended to leverage already deployed analysis tools, hence lowering the effort required for its adoption.

##### 4.1.2 Language independence

ATDx allows combining the results of different language-specific analyses via the ATDx building process discussed in Section 4.2. This enables its adopters to aggregate analysis results into a single comprehensive overview of the ATD present in a software-intensive system, encompassing architectural components implemented in heterogeneous programming languages.

##### 4.1.3 Tool composability

In our viability investigation we considered a single analysis tool, namely SonarQube. Nevertheless, the ATDx approach, thanks to its entailed abstraction and normalization steps, allows to aggregate analysis results gathered via heterogeneous tools, e.g., static and dynamic analyzers, in order to gain a simple, unified, overview of the ATD present in a software system.

[https://github.com/ATDindex/ATDx](https://github.com/ATDindex/ATDx)


4.1.4 Multi-level granularity results

ATDx provides the means to inspect ATD issues at various levels of granularity. At the highest level of abstraction, the ATDx value, representing the total ATD present in a software-intensive system, enables to swiftly gain knowledge of the ATD present in a software system, and to compare its normalized value with the ones of other projects. At a more refined level of granularity lies the set of architectural technical debt dimensions $ATDD^T$. Such semantic decomposition provides the possibility to intuitively gain an overview of the ATD items of different nature which are present in the considered system. The $ATDD^T$ values enable adopters to effectively compare the ATD status across different projects and to visualize it, supporting the communication of related concepts and a deeper understanding of the root causes of ATD. Finally, at the finest level of granularity, the derived data utilized to build the overview can be analyzed. This provides the lowest level of the ATDx calculation, namely the normalized $AR^T$ outlier values, in order to gain a deep, actionable, and data-driven guidance on where the architectural debt is rooted. This last level of granularity enables developers to pinpoint the precise location of the source of ATD.

4.1.5 Extensibility

After the implementation of an ATDx instance, it is possible to add new rules to the existing $AR^T$ set by re-executing Steps 1 to 4 of ATDx exclusively on the newly added rules. This accounts for scenarios such as the support of new rules added by tool vendors, or the inclusion a posteriori of an an additional architectural rule. Additionally, it is possible to include additional tools in the ATDx analysis by integrating the new results with the ones obtained from already adopted tools. This process is carried out by expanding the previously identified $ATDD^T$ dimensions, and possibly including new emerging ATD dimensions.

4.1.6 Data-driven

Instead of relying on predefined thresholds to identify outlier values of architectural rules, ATDx takes advantage of normalization and statistical analysis of data gathered from the analysis of a dataset of software projects. Hence, ATDx is not prone to design problems related to the definition of metric thresholds, as it identifies severe architectural rule violations based on the evaluation of empirical values collected from a multitude of (possibly third-party) projects.

4.1.7 Lightweight analysis for additional SUA

Once an initial dataset of $NORM^T_i$ values is collected, the evaluation of additional systems via ATDx is straightforward. In fact, as a dataset of $NORM^T_i$ values is already available, the analysis can be carried out exclusively on the new SUA, without the need to re-execute the analysis for all projects belonging to the dataset. Additionally, as ATDx is designed to leverage tools already deployed in a specific industrial setting, pre-computed data can be used as input for ATDx, enabling to carry out the analysis by investing a negligible amount of effort.

4.1.8 Domain-specific customization

As ATDx relies on the analysis of a dataset of software projects, rather than considering a single one, it is possible to tailor the approach by including in the dataset exclusively software projects belonging to a specific domain, e.g., safety-critical systems, mobile applications. In fact, ATDx analysis results may vary according to the application domain considered. Hence, to get more refined and accurate results, a curated dataset of software projects belonging to the same domain can be considered, enabling to fine-tune the detection of severe rule violations according to the specific domain considered.

4.2 Approach Drawbacks

4.2.1 Dataset dependence

Naturally, the ATDx approach is dependent, by definition, on a dataset of software projects. Hence, while numerous ATD analysis approaches require exclusively the SUA [Verdecchia et al., 2018], our approach needs instead various software systems in order to detect outlier values. This implies that the ATDx results of a SUA are only “relative” to the other projects considered, and are not representing an universal result. Additionally, to provide accurate ATDx analysis results, it is necessary to ensure the quality of the utilized dataset. In order to mitigate this latter drawback, a quality filtering process on a preliminary dataset of software projects should be executed (see also Section 3.3). Regarding the scale of the dataset, we plan to carry out quantitative and qualitative experiments by considering different domains in order to identify the minimum number of projects required to achieve saturation of results (i.e., when the inclusion of additional projects does not lead to statistically significant changes of results).

4.2.2 ATDx implementation validation

In order to verify the correctness of an implemented instance of ATDx, it is necessary to evaluate and eventually tune it in a real development context. This process entails (i) the selection of a set of software systems, (ii) their analysis via ATDx, and (iii) the inspection of the analysis results by means of focus groups with the developers who implemented the selected software projects. As the focus of our study
is defining a generic approach through which specific instantiations of ATDx can be implemented, this process was deemed as out of scope for the ATDx prototype implementation carried out for this study.

4.2.3 Emphasis of outlier values

The design of ATDx relies on the identification of outlier normalized values of AR\textsuperscript{T} violations (i.e., values within the upper inner fence and the upper outer fence). Hence, the emphasis of ATDx is on severe ATD issues present in software-intensive systems. While in our ATDx viability investigation we observed numerous projects characterized by outlier values (4179/6706), such process could be deemed as “lossy” if a more fine-grained analysis is expected. To mitigate this potential drawback, it is possible to modify the approach by considering more inclusive strategies, e.g., by considering as “outliers” the values between the third quartile and the upper outer fence, or even mapping a discrete value according to the quartile associated to an AR(S) value.

4.2.4 Balance of AR number mapped to ATDD

As the different ATDD values constituting ATDx are computed by considering distinct sets of ARs, it is necessary that the number of rules across the different sets is balanced. In fact, if the distinct sets exhibit notable differences in cardinality, the weight of under-represented sets could lead to an unfair representation of an ATDD. This drawback has to be considered and mitigated while designing an ATDx instance, by carefully selecting ARs, considering their recurrence, relevance, and the cardinality of the mapped ATDDs.

4.2.5 Potential empirically unreachable ATDD maximum values

Reaching a maximum value in a certain dimension ATDD\textsuperscript{T}_i is by definition theoretically possible, but empirically extremely improbable. By design, a system S reaches the maximum in one dimension if and only if it possesses outlier values in all ARs composing ATDD\textsuperscript{T}_i. If S possesses a maximum value of ATDD\textsuperscript{T} in all ARs, this would indicate that S is characterized by exceptionally severe and recurrent issues in that dimension. In our viability investigation with SonarQube such project was not present for all dimensions. A possible heuristic to solve this potential drawback would be to rescale the values of a dimension i to the \([0, \text{max}_S]\) range, where \text{max}_S is the maximum value in the dimension i across all rules in AR\textsuperscript{T} mapped to i. Nevertheless, we refrained from such solution for the sake of clarity.

5 RELATED WORK

Numerous software analysis approaches have been proposed to detect ATD in software-intensive systems. Among the most prominent and current ones, the approach of Arcelli Fontana et al. [Arcelli Fontana et al., 2016] and Martini et al. [Martini et al., 2018a]. Roveda et al. [Roveda et al., 2018] focuses on the identification of ATD by analyzing dependency architectural smells, which could lead to the emergence of an additional ATD dimension, namely “Dependency”. Similarly, Kazman et al. [Kazman et al., 2015] and Xia et al. [Xia et al., 2016] Cai and Kazman, 2017], analyzed ATD by inspecting antipatterns of semantically related architectural components, e.g., via the analysis of bug-prone components. In [Roveda et al., 2018], another ATD index is presented. Differently from our ATDx, this concentrates entirely on architectural smells, notably related to dependency violations. Finally, Le et al. [Le et al., 2018] reported on an empirical investigation of architectural decay via the analysis of 8 architectural smells of different nature. Interestingly, in a similar fashion to ATDx, the smell violations are evaluated by adopting interquartile analysis [Tukey, 1977]. Nevertheless, such analysis is carried out at an intra architectural level and, in contrast to ATDx, values are not normalized per system-size. More ATD identification approaches are reported in a secondary study of Verdecchia et al. [Verdecchia et al., 2018]. Regarding the identification of metrics thresholds, similarly to ATDx, in [Alves et al., 2010] an interquartile strategy is adopted to identify the severeness of metric values. Differently, this study does not focus on ATD and, while adopting a system-size normalization strategy, it takes only one level of granularity (NCLOC). In a recent work, Ulan et al. [Ulan et al., 2019] proposed a software metric aggregation approach based on distribution. Our approach is different by (i) focusing specifically on outlier architectural rule violations, (ii) considering sizes according to distinct granularities, and (iii) clustering results into semantic dimensions.

6 CONCLUSIONS

In this study we presented ATDx, an approach designed to gain an encompassing overview of the source-code detectable ATD present in a software-intensive system. The ATDx approach concisely entails (i) the manual inspection of pre-existing source-code rules, (ii) the identification of outlier normalized values via statistical analysis, and (iii) the aggregation of analysis results into a set of ATD dimensions. As future work, we plan to address the identified drawbacks by refining our theoretical framework of ATDx. Additionally, we would like to enhance the current approach by including the temporal factor into our anal-
ysis. Subsequently, we plan to implement a concrete instance of AT Dx, and validate it by conducting a large-scale study involving industrial partners in order to assess (i) the usefulness and actionability of the proposed index, (ii) its sensitivity w.r.t. different analysis tools, and its (iii) performance when dealing with larger datasets. As stated in [Nord et al., 2012], ATD is a complex, heterogeneous, and multifaceted problem. Our approach contributes to advance the field towards establishing a holistic, encompassing, view of the ATD present in a software-intensive system.

ACKNOWLEDGMENTS
This material is partially based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center. DM20-0240.

REFERENCES


