



# Determining Microservice Boundaries: A Case Study Using Static and Dynamic Software Analysis

Tiago Matias<sup>1</sup>, Filipe F. Correia<sup>1,2(✉)</sup>, Jonas Fritzsche<sup>4,5</sup>, Justus Bogner<sup>4,5</sup>, Hugo S. Ferreira<sup>1,2</sup>, and André Restivo<sup>1,3</sup>

<sup>1</sup> Faculty of Engineering, University of Porto, Porto, Portugal  
{up201700421, filipe.correia, hugosf, arestivo}@fe.up.pt

<sup>2</sup> INESC TEC, FEUP Campus, Porto, Portugal

<sup>3</sup> LIACC, FEUP Campus, Porto, Portugal

<sup>4</sup> Institute of Software Technology, University of Stuttgart, Stuttgart, Germany  
{jonas.fritzsche, justus.bogner}@iste.uni-stuttgart.de

<sup>5</sup> University of Applied Sciences Reutlingen, Reutlingen, Germany

**Abstract.** A number of approaches have been proposed to identify service boundaries when decomposing a monolith to microservices. However, only a few use systematic methods and have been demonstrated with replicable empirical studies. We describe a systematic approach for refactoring systems to microservice architectures that uses static analysis to determine the system's structure and dynamic analysis to understand its actual behavior. A prototype of a tool was built using this approach (MonoBreaker) and was used to conduct a case study on a real-world software project. The goal was to assess the feasibility and benefits of a systematic approach to decomposition that combines static and dynamic analysis. The three study participants regarded as positive the decomposition proposed by our tool, and considered that it showed improvements over approaches that rely only on static analysis.

**Keywords:** Microservices · Refactoring · Software architecture

## 1 Introduction

The *microservices* architecture steadily gained popularity over the last years. Nowadays, it is often used in greenfield projects, but a lot of the times, systems are first developed as monoliths, which are quicker to develop and to test than microservices. Monoliths can then be broken up into microservices, when and if the need arises [1]. Doing this may promise high scalability, shorter release cycles or better maintainability. However, missing to identify the right boundaries may hinder reaching these benefits [2]. Therefore, an essential part of such a refactoring is the decomposition approach [3], which has the end-goal to identify contextually-related functionality and encapsulate it into different services. These should be characterized by a high cohesion inwards and loose coupling

outwards. To optimally leverage from the microservices architectural pattern, existing functionality has to be split up with appropriate granularity as well.

There have been a number of approaches proposed already to decompose monoliths into microservices [4, 5]. However, Fritzsche et al. found that such refactoring approaches were often not considered by practitioners and that identifying suitable service cuts is still perceived as a major challenge [3]. They asked 16 practitioners from 10 companies who were in the process of migrating their systems. Participants were either not aware of such tools or even convinced that it would be impossible to automate such a complex task. In a review of refactoring approaches, the same authors ascribed a lack of automation and missing tool support to most approaches proposed by academia [5]. This lack of tools inhibits adoption in industrial contexts and makes empirical studies more challenging to conduct.

We address this gap by *a*) identifying a systematic approach that combines principles of the previously proposed methods, *b*) using it to create a prototypical implementation, and *c*) conducting an industry case study with the prototype.

In the remainder of the paper, we discuss **related work** to provide an overview of other approaches, and describe our own **approach**, which relies on static and dynamic analysis. We introduce a **prototype** – MonoBreaker – that embodies this approach, and that identifies service boundaries for monoliths based on the Django web framework. Afterward, we present a **case study**, in which we contrast the results of MonoBreaker with ServiceCutter by surveying three developers of the project.

## 2 Related Work

The subject of decomposing and migrating monolithic applications to microservices is addressed in books such as *Building Microservices* [6], *Monolith to Microservices* [7] and *Microservices Patterns* [8]. Likewise, a variety of research papers describe ways to tackle such transformations.

Building microservices ideally means to create services that are highly cohesive and loosely coupled. Tyszberowicz [9] confirms that Domain-Driven Design (DDD) is the most common technique for modeling microservices. With DDD, the software mirrors business domains and sub-domains as well as the related domain models and bounded contexts. Each bounded context implements a small set of strongly-related behaviors and conforms to the Common Closure Principle [10]. These sets of behaviors shape individual units, resulting in cohesive designs of loosely-coupled services [11]. A system following DDD supports a higher degree of team independence as well as better scalability, testability and changeability [12].

**Meta-studies.** Ponce et al. provide an up-to-date overview in their review of 20 papers of migration and refactoring techniques [4]. Their study focuses on the approaches, the applicability to certain system types, validations of the techniques, and the associated challenges. The authors group works by their underlying decomposition approaches: *model-driven* (involving design elements, e.g.,

DDD), *static analysis* (based on source code) and *dynamic analysis* (based on runtime data).

Fritzsches et al. similarly compare 10 refactoring approaches and likewise provide a classification [5]. They distinguish decompositions based on *Static Code Analysis*, *Meta-Data*, *Workload-Data*, and *Dynamic Microservice Composition*. While the first three classes imply a fixed decomposition result, a dynamic composition of services would be continuously re-calculated, e.g., based on workload constraints. The study moreover reveals that most approaches are only applicable to certain types of applications, require significant amounts of input, or have limited and prototypical tool support.

**Concrete Approaches and Tools.** Nunes et al. pursue an approach based on identifying transactional contexts of business applications and using a clustering algorithm to determine service candidates [13]. Chen et al. similarly base the decomposition on the data flow of the business logic [14]. They compare the resulting service cut with the output of ServiceCutter [15], a freely-available tool implementing the approach by Gysel et al. [16]. ServiceCutter applies a clustering algorithm to identify new services and currently supports the *Girvan-Newman* and *Leung* algorithms for this purpose. To calculate the service cut, it requires that an *Entity-Relationship Model* (ERM) of the system is given in a specific format along with *User Representations* and *Coupling Criteria*. The collection of these partly-exhaustive system specifications is done in a manual process and requires the help of domain experts.

Ren et al. acknowledge the inadequacy of approaches only relying on static analysis [17]. They recognize that not analyzing the runtime behavior would hinder the calculation of a complete and accurate service cut. Therefore, they combine static and dynamic analysis based on the applications' runtime behavior. A subsequent clustering calculates the candidate service cut. Likewise, Taibi et al. propose a combined approach based on dependency analysis and process mining techniques [18]. The decomposition encompasses execution path and frequency analysis. After removing circular dependencies, additionally specified decomposition options are ranked based on coupling and granularity metrics to produce the candidate service cut. The authors employ a tool<sup>1</sup> that is capable of generating graphical visualizations to represent the business processes. Although a tool is referenced to capture the dynamic behavior of the system, the suggestion of service cuts is outside the scope of the work and must be done by experts, even if the authors mention that the process can somehow be automated.

**Implications for our Approach.** The methods described by Richardson [8] (*Decompose by business capability* and *Decompose by subdomain*) provide general guidelines for a partly-automated decomposition process. They support architects in choosing appropriate input values and assessing the resulting candidate service cuts.

The two meta-studies by Ponce and Fritzsches yield a variety of strategies to break down a monolith. Most do not combine static and dynamic analysis

---

<sup>1</sup> More information is found at the tool's website – <https://fluxicon.com/disco/>.

to steer the decomposition. As such, the works by Ren et al. and Taibi et al. comprise the core concepts of the approach described in our work. These works do not provide tools for service decomposition, or for any form of automation, but we will build on the concept of gathering runtime behavior and its analysis.

ServiceCutter is also of importance to our work, as it too implements the deterministic *Girvan-Newman* algorithm. In some aspects, our work is less sophisticated than ServiceCutter, as it does not yet consider quality attributes like security, scalability, and business ownership. However, it trades that for the benefit of being independent from extraneous, subjective, information provided by experts to determine the service cuts.

### 3 Approach

Decomposing a monolith is often done based on insights from software developers on the specific context of the *problem domain* and of the *application's architecture*. The challenge that we aim to address is to reduce subjectivity, making the process more systematic and automated. The approach described below is based on ideas that have been documented before [16–18], but are employed here for determining service boundaries with minimal to no manual input, which so far has not been feasibly demonstrated. Therefore, the approach is described as a hypothesis, and the case study in Sect. 5 as a first step to provide support for its effectiveness.

In more concrete terms, this approach aims to be data-driven and to be independent of sophisticated input from experts. To do this, we do not take into account all the intricacies of the process as it is often done manually today. Instead, we focus on what information can be obtained from the application itself via *static* and *dynamic* analysis to find beneficial service cuts. We rely on the availability of *a)* static software artifacts, namely source code, and *b)* operational data, such as the use of API endpoints, of datastores, and of issued method calls.

**Static Analysis.** Software artifacts are analyzed and the collected information used to build a graph-like model of the system, representing components as *nodes* and the dependencies between them as *edges*. Components and dependencies can be of different types, and identifying them will depend on the used programming languages, frameworks and environments. For example, components can refer to *classes*, *packages* or *modules*, and dependencies to *imports* or *method calls*.

Each edge is assigned a *weight* to represent the strength of the dependency. This is a function of the number and quality of connections between the two components. The weight of edges after static analysis can, for example, be the sum of the number of imports and method calls between its two components.

**Dynamic Analysis.** The system is then monitored at runtime to gather *operational data*, which is analyzed to identify how the dependencies are exercised during execution, and gain an understanding of how the system is actually used. Such information is used to compute a new weight for each edge of the graph. The *final weight* values are a function of the *static* and *dynamic weights*, and

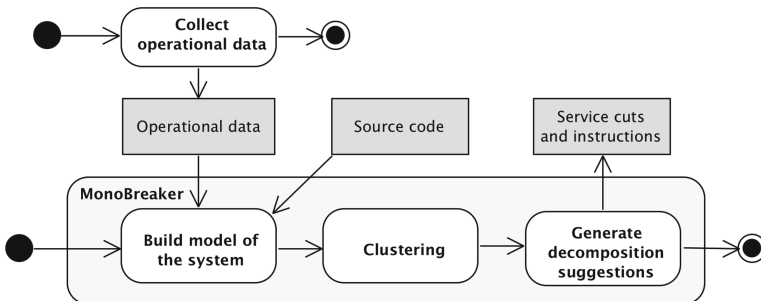
are a measure for how the components in the system are mutually bound. The underlying assumption is that a high amount of interaction between two components correlates with belonging to a common bounded context. Including them in different microservices would imply higher costs in latency and in maintaining resilience and fault tolerance.

**Clustering.** A graph of the service composition will support identifying different clusters of components. The nodes connected by the edges with higher weight values will be grouped to form clusters of relatively *high cohesion*. These clusters will depend on each other through edges with low weight values, representing relatively *low coupling*. The clusters can, therefore, be used to determine a set of possible *service cuts*. The specific clustering algorithm to be used is outside the scope of this approach, but would be interesting to explore (see Sect. 6).

**Decomposition Suggestion.** The identified service cuts serve as a foundation for assigning existing *software artifacts* to each of the new services and advise on the architectural *refactoring process*.

## 4 The MonoBreaker Tool

MonoBreaker aims to demonstrate the feasibility of the approach and was used in the **case study** described in Sect. 5. It is a prototype<sup>2</sup> and currently works with applications using the Django web framework. It takes a project's directory as input and does a **static analysis** of the source code to identify the overall project structure. This information is mapped to a graph-like model together with associated files and their dependencies. The same graph is populated with data collected through **dynamic analysis** to quantify the strength of the dependencies. The graph is then traversed to suggest a decomposition into new services, highlighting the source code files that will be involved and how the resulting services should communicate. This workflow is depicted in Fig. 1 and the several steps are exemplified below.



**Fig. 1.** Operation flow of MonoBreaker with inputs and outputs.

<sup>2</sup> MonoBreaker is freely available at <https://github.com/tiagoCMatias/monoBreaker>.

## 4.1 Collect Operational Data

Operational data is gathered using Silk, which is a profiling tool for Django<sup>3</sup>. The tool is capable of supplying information about the usage of endpoint methods (the ones invoked when a URL is requested), and the model classes and queries involved in the process of returning results from the database. It uses this information to infer some of the internal method calls, as we will see in the next section.

## 4.2 Build Model of the System

The static analysis inspects the domain model, the views, and the dependencies between them. In particular, it tracks the use of Django’s `Model` class, identifying its subclasses (i.e., the *domain model* of the application) and how they are connected through the declared *foreign keys*. It also tracks the use of the `ModelViewSet` class by identifying its subclasses (i.e., the *views* of the application) as well as the connections between these views and the model classes, via the `import` statements.

To illustrate the process, we present a minimalist example and the steps involved in suggesting a service decomposition using MonoBreaker. The file exemplified in Listing 1.1 results in the extraction of the `ViewItem` class as a new graph node. The imports of `Attribute` (line 4) and `Item` (line 6) refer to subclasses of Django’s `Model` class, therefore, these are also extracted as nodes, with graph edges connecting them to the `ViewItem` node. Both the model subclasses `Attribute` and `Item` have a connection to `ViewItem` because it imports them and invokes their methods.

---

```

1 from rest_framework.decorators import action
2 from rest_framework.response import Response
3 from rest_framework.viewsets import ModelViewSet
4 from ..models.Attribute import Attribute
5 from ..serializers.ItemSerializer import ItemSerializer
6 from ..models.Item import Item
7
8 class ViewItem(ModelViewSet):
9     queryset = Item.objects.all()
10    serializer_class = ItemSerializer
11
12    @action(methods=['get'], detail=False)
13    def get_item_details(self, request):
14        if request.GET.get('attributes', None):
15            data = self.serializer_class(self.queryset, many=True).
16                to_representation(self.queryset)
17            for item in data:
18                item['attributes'] = Attribute.objects.get_by_item(item['id'])
19            return Response(data)
20        else:
21            return Response(ItemSerializer(Item.objects.all(), many=True).data)

```

---

**Listing 1.1.** The `ViewItem` class, an example of a *view* in a Django application.

<sup>3</sup> See <https://github.com/jazzband/django-silk> for more information.

Monobreaker uses the graph resulting from the analysis described thus far to generate the visual representation depicted by Fig. 2a. The *weight* values associated to the edges represent the strength of the dependencies and are determined by:

$$\textit{StaticEdgeWeight} = \textit{NumImports} + \textit{NumMethodCalls}$$

After the analysis of all source code files, a global dependency graph of the project is built. In this example, these files would also include Listing 1.2.

---

```

1 from rest_framework.decorators import action
2 from rest_framework.response import Response
3 from rest_framework.viewsets import ModelViewSet
4 from ..serializers.OrderSerializer import OrderSerializer
5 from ..models.Order import Order
6
7 class ViewOrder(ModelViewSet):
8     queryset = Order.objects.all()
9     serializer_class = OrderSerializer
10
11 @action(methods=['get'], detail=False)
12 def get_order_details(self, request):
13     if request.GET.get('items', None):
14         data = self.serializer_class(self.queryset, many=True).
15             to_representation(self.queryset)
16         for order in data:
17             order['items'] = Order.objects.get_order_items(order['id'])
18         return Response(data)
19     else:
20         return Response(OrderSerializer(self.queryset, many=True).data)
21
22 def list(self):
23     return Response(OrderSerializer(Order.objects.all(), many=True).data)

```

---

**Listing 1.2.** The `ViewOrder` class, an example of a *view* in a Django application.

Figure 2b represents the updated version of the graph after the static analysis of the second *view* class. Note also the dependency between `ViewOrder` and `Item` via the call to the `get_order_items()` method. Detecting it could be attempted through deeper static analysis, in particular of chains of method calls that jump into framework code. The static detection of this dependency is a limitation of the current implementation of `MonoBreaker`, but it is one of little consequence, as it can still be detected through dynamic analysis, as we will see next.

The static analysis of the system is followed by the runtime analysis. The operational data that was previously collected (see Sect. 4.1) is processed and the result used to update the graph with *a*) previously undetected dependencies (in this example, the one between `ViewOrder` and `Item`) and *b*) with updated weight values. This ensures that we also consider the existence and the strength of dependencies that cannot be determined solely by inspecting the source code.

The requests received by the application may result in multiple method calls that eventually touch specific model classes. These are determined by `MonoBreaker` via the database queries that are issued during the processing of a specific request. Table 1 shows some of the data resulting from the dynamic analysis, which is used to compute the dynamic weights.

**Table 1.** Data determined through dynamic analysis for this example.

View	Method	# Calls	Related models
ViewOrder	list()	2	Order
ViewOrder	get_order_details()	4	Order, Item
ViewItem	list()	4	Item
ViewItem	get_item_details()	8	Item, Attribute

To keep the weight values calculated by the dynamic analysis in the same order of magnitude as those calculated from static analysis, MonoBreaker normalizes them – the highest weight determined from the dynamic analysis will be at most as high as the highest one calculated from static analysis. Therefore, the equation representing the weight that arises from dynamic analysis becomes:

$$DynaEdgeWeight = NumMethodCalls \times \frac{MaxStaticWeight}{MaxNumMethodCalls}$$

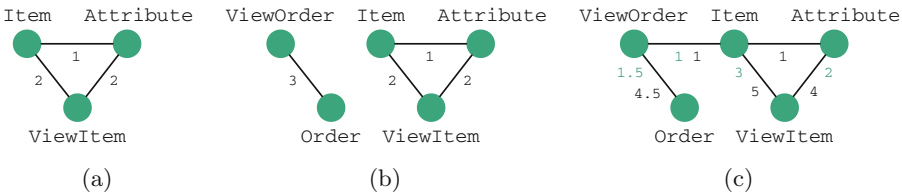
In this implementation, the weights from the static and dynamic analyses were considered in equal parts for determining the final weights, resulting in:

$$EdgeWeight = StaticEdgeWeight + DynaEdgeWeight$$

Figure 2c depicts the resulting graph, showing the computed *DynaEdgeWeight* in green and the final *EdgeWeight* in black.

### 4.3 Clustering

The dependencies collected through the static and dynamic analyses are used by MonoBreaker to create a graph-like model of the system. Nodes consist mainly of Django *model* and *view* classes. A clustering algorithm is then applied to break



**Fig. 2.** Each graph shows a different stage of the example, (a) is after analysing the *ViewItem* class, (b) after analysing the *ViewOrder* class, and (b) after incorporating the results from the dynamic analysis. Values in green are the weights determined by dynamic analysis alone, and those in black are the total weight produced up to that stage.



the network down into smaller communities, thus grouping nodes according to the weights of the edges. We have chosen the Girvan-Newman algorithm<sup>4</sup> [19] given its apparent successful use in tools such as ServiceCutter. The resulting clusters indicate a set of potential service cuts.

#### 4.4 Generate Decomposition Suggestions

After clustering the nodes, MonoBreaker provides an overview of the decomposition. It obtains the service cuts through the Girvan-Newman algorithm and provides the lists of the classes that will be needed for each service. These can be used by the developers to guide the refactoring process. Listing 1.3 shows the output for our simple example.

---

```

Total Files: 19
Django_Views: 2
Django_Models: 3

GraphNumber: 0
list_of_files: [
    'models.Attribute',
    'models.Item',
    'serializers.ItemSerializer',
    'views.ViewItems'
]

GraphNumber: 1
list_of_files: [
    'models.Item',
    'models.Order',
    'serializers.OrderSerializer',
    'views.ViewOrder'
]

```

---

**Listing 1.3.** Example of an output of MonoBreaker.

#### 4.5 Limitations

The approach described in Sect. 3 is designed to apply to a wide range of contexts. The tool described in this section, on the other hand, was designed with a narrower scope and it is worth highlighting some of its limitations.

**Technologies.** The opportunity, of using a fully developed monolith built with Django to conduct a case study in the industry, led us to develop MonoBreaker specifically for Django-based monoliths that use the object-relational mapper. At this point, the tool will work only for systems developed using these technologies.

**Design Assumptions.** The implementation makes simplistic assumptions about the system to decompose, such as that it was designed around a domain model, and that it avoids cyclic dependencies and other kinds of unnecessary complexity. Such design problems should be approached before running MonoBreaker.

<sup>4</sup> Connectivity-based clustering algorithm, such as Girvan-Newman, are based on the idea that nodes have more affinity to nearby nodes than to those farther away.

**Operational Time Frame.** The quality of the decomposition is sensible to the choice of an appropriate time frame for collecting operational data, as it should be representative of how the system is normally used. Functionality not used during the dynamic analysis time frame will not be considered for calculating dynamic weights.

**Balancing Quality Attributes.** Another assumption is that there is a single *optimal* set of service cuts, but we know that there are often trade-offs when refactoring. Users of MonoBreaker are still not able to specify, for e.g., how the maintainability of the resulting system should be weighed against its scalability.

## 5 Case Study

To assess the feasibility and benefits of a systematic approach to decomposition that combines static and dynamic analysis, we conducted an industry case study using the developed prototype. We were interested in generating insights about the approach, in particular, in understanding its effectiveness for identifying good service boundaries when refactoring a monolith, and the impact that *dynamic* analysis has on the decomposition result. For the latter part of the study, we turned to ServiceCutter for a comparison.

### 5.1 Context

The case study focused on a web application for supporting the collaboration between two centers of a logistics startup company. The application had 15 KLOC and more than 40 domain-model elements, and had recently gone through significant growth in its use, making it an interesting candidate for the study.

We achieved the participation of three of the four developers that form the team responsible for this application. Their professional experience was in the range of 1–5 years for two of the developers and 5–15 years for the third developer.

### 5.2 Process

MonoBreaker was used to analyze the project and produced a suggestion for decomposing it into different services. The process consisted of four steps:

- a) **Run MonoBreaker** – We gathered the project source code and the runtime data collected through Silk and provided them as input to MonoBreaker, which used both *static* and *dynamic* analysis to produce a suggestion of how the system could be decomposed.
- b) **Run ServiceCutter** – The data statically-collected in step **a)** was transformed to the ERM format expected by ServiceCutter and was provided as input to produce an alternative decomposition using *static* analysis only.
- c) **Present MonoBreaker** – A session was scheduled with the development team and included an introduction that explained the goal of the experiment and a showcase of MonoBreaker using an example project.

**Table 2.** Questions and answers in the *approach* group.

Question [ <i>It's important ...</i> ]	Answers
<i>... to know what methods are called between the components of the monolith</i>	5, 4, 3
<i>... to know how frequently each method is called when the monolith is run in production</i>	4, 3, 3
<i>... to identify what the domain objects of the monolith are</i>	5, 5, 5
<i>... to identifying what are the relationships between the monolith components</i>	5, 5, 5
<i>... to know how the relationships between the components are used when the monolith is run in production</i>	5, 5, 5
<i>... to identify what imports are made by each software component of the monolith</i>	5, 4, 3
<i>... to identify what the schema of the database/datastore is</i>	5, 5, 5
<i>... to know the operations made to the database/datastore</i>	5, 4, 4
<i>... to identify how frequently the operations made to the database/datastore are executed when the monolith is run in production</i>	4, 3, 4

- d) **Questionnaire** – Following the MonoBreaker demo, a questionnaire was handed out to the participants. It aimed to assess how the feasibility of the approach and the impact of *dynamic* analysis on the quality of the results were perceived by the team. The participants did not have access to the source code during the questionnaire, and the two service decompositions were presented visually as dependency graphs. Participants were given 30 min to analyze the graphs and answer the questionnaire.

**Table 3.** Questions and answers in the *feasibility* group.

Question [ <i>The proposed decomposition as microservices ...</i> ]	Answers
<i>... is the best one possible</i>	4, 3, 2
<i>... is easier to scale (performance)</i>	4, 3, 2
<i>... is easier to deploy new versions of the system</i>	4, 3, 2
<i>... is easier for maintainability by the existing team(s)</i>	4, 3, 2



Fig. 3. MonoBreaker decomposition result as depicted in the questionnaire.

### 5.3 Data Sources

The case study used as data sources: *a)* the source code of the project, *b)* operational data collected through Silk during one week in a production environment and *c)* the answers to the questionnaire that were given by the team of the project.

The source code was obtained from the company’s code repository. The operational information was collected in two tables created by Silk in the application’s database (`silk_request` and `silk_sqlquery`). The questionnaire was built using Google Forms and the answers were gathered in a spreadsheet.

### 5.4 Data Analysis

Most questions were based on a Likert scale [20], ranging from (1) *Strongly Disagree* to (5) *Strongly Agree*. Questions were organized into four groups. Below, we summarize the answers provided by the three interviewees for each group of questions.

**Personal Experience.** These questions support understanding the team’s professional experience, its familiarity with the case study project and with the process of migrating monoliths to microservices. The answers reveal that all team members have some experience migrating monoliths to microservices (3, 4, 3)<sup>5</sup> and that they were very familiar with the case study project (5, 5, 5), as expected. This ensures their ability to evaluate the decomposition approach.

**Approach.** The questions in this group aim to assess the perceived importance of different aspects when decomposing a monolith into microservices. If

<sup>5</sup> Throughout this section, we’ll use this notation to represent the answers of the three team members to a questionnaire item using a five-level Likert scale.

the understanding of these aspects by the study participants revealed to be different from our own, it could explain differences in the answers to questions in the next groups of questions. The questions and answers from the three developers are shown in Table 2. The results show unanimous agreement in that identifying the *domain objects*, the *relationships between components*, how these *relationships are used in production* and the *schema of the data store* are very important factors when determining potential new services (5,5,5).

The answers to the remaining questions were not unanimous, but still show that significant importance is attributed to knowing *what operations are made to the database/datastore* (5,4,4).

These results show the relevance, as perceived by the members of this team, of both structural and behavioral information for service decomposition, and therefore are aligned with the concepts that we used to define our approach.

**Feasibility.** The questions in this group evaluate the perceived feasibility of the approach regarding the quality attributes of the application. Namely, the questions focus on the scalability, ease of deployment, and ease of maintenance. They are supported by the decomposition created by MonoBreaker, which was visually presented as depicted by Fig. 3. Both the questions and the answers are shown in Table 3.

The participants did not agree in their answer to these questions but answered consistently to all the questions (4,3,2). This led us to inspect more closely the answers for the *justification* question (the open-ended question where they could provide further context to their answers) and conclude that the decomposition was perceived as a good basis, but insufficient. Namely, the decomposition consists of 3 services, but team members argued in favor of a more aggressive decomposition. Looking closely at Fig. 3, we can see clusters around three different classes – `CargoMovement`, `MasterdataProducts` and `ShippingTransfer`. From their answers, we understood that the team was expecting the `ShippingTransfer` cluster to be further decomposed into two distinct services. Section 6 outlines a few factors that can be explored in future work to improve the decomposition.

**Comparison With Using Only Static Analysis.** This group has two Likert-scale questions, each accompanied by an open-ended *justification* question.

The first question compared the decomposition using both *dynamic* and *static* analysis with the one using only *static* analysis. To ease the comparison between the outputs, we transported the information to Gephi<sup>6</sup> and extracted both graphs. The graphs were depicted in the beginning of this group of questions as *Decomposition A* and *Decomposition B* (respectively, Fig. 4 and Fig. 3).

The second question directly addressed the usefulness of the output provided by MonoBreaker, listing the classes that would be required by each service.

Table 4 shows the two questions and the associated answers.

---

<sup>6</sup> Gephi is a tool for graph analysis and visualization – <https://gephi.org>.

**Table 4.** Questions and answers in the “comparing with the state-of-the-art” group.

Question	Answers
<i>The decomposition A is better than the decomposition B</i>	2, 2, 1
<i>A tool to support decomposing a monolith into microservices would be useful if it provided this output</i>	5, 5, 5



**Fig. 4.** ServiceCutter decomposition result as depicted in the questionnaire.

The answers dismiss *Decomposition A* as the best, concluding that combining *static* and *dynamic* analysis provided a better decomposition when compared to using *static* only.

Regarding the output provided by MonoBreaker for guidance on the refactoring, the answers were unanimous in that it would be helpful.

### 5.5 Threats to Validity

The purpose of this case study is to gather evidence to support the approach. The design described in Section 5 tries to minimize possible threats to validity, but those that exist need a closer look.

**Projects and Participants.** The sample of our case study was limited to one project and three software developers. The answers to the questionnaire’s *approach* group can be used to confirm if this team valued both structural and behavioral information when decomposing services, as these were base assumptions used to design our approach, but the small scale doesn’t allow to generalize conclusions. We would certainly like to see this case study replicated for other products and larger organizations with different backgrounds, to verify if these preliminary results hold in other contexts.

**Possible Biases from Respondents.** The partnership with the startup company for this case study was only possible due to good working and personal rela-

tionships and commitment between the company and the researchers. Therefore, there is always the possibility that the participants may have been inadvertently influenced. During the MonoBreaker presentation (Sect. 5.2), we took particular caution to take an impartial stance regarding the merits of the tool and of its underlying approach and to not interfere in any way when participants were responding to the questionnaire. Moreover, they didn't know which decomposition had been made using only *static* analysis or using both *static* and *dynamic* analysis. For these reasons, we are confident in discarding this as a threat to validity.

**Representativeness of Sampled Data.** The company supplied the project source code and allowed to alter it to enable the collection of operational data that otherwise would not be possible. As already mentioned, the operational data covered only one week of the application's run time information and collecting data for a longer period may have led to different results. All the relevant functionality of the application seems to have been used during this time, and we believe the amount of data to be sufficient to base a decomposition decision on. For this reason, we are confident in discarding this as a threat to validity.

**Suboptimal Baseline.** To assess the impact of *dynamic* analysis in the decomposition, we compared the result of MonoBreaker (using *static* and *dynamic* analysis) with that of ServiceCutter (using *static* analysis only). The choice of ServiceCutter stemmed from the intention to compare MonoBreaker with leading tools from the current state of the art. ServiceCutter is the only freely-available tool that we could run to automate the decomposition process with minimal manual input<sup>7</sup>.

However, we realized that the specific purpose of assessing the impact of dynamic analysis would have been better served by comparing the output of MonoBreaker when run with *static* and *dynamic* analysis with its output when run with *static* analysis only. We believe that when the Girvan-Newman algorithm is chosen when running ServiceCutter, the resulting output should be identical to MonoBreaker's if only *static* analysis is used, as MonoBreaker uses the same algorithm for clustering dependent components. Notwithstanding, running MonoBreaker with and without dynamic analysis would provide more robust evidence that no other factors had a significant influence on the decomposition result.

## 6 Conclusions and Future Work

In this work we contribute, *a*) a systematic **approach** to decompose monolithic applications to microservices, *b*) a **tool** prototype (MonoBreaker) that implements this approach and *c*) the design and results of an industry **case study**.

<sup>7</sup> This was possible by synthesizing a part of the inputs that it requires – namely, the ERM – and omitting the remaining inputs, which we were unable to create without resourcing to software developers – namely, the *User Representations* and the *Coupling Criteria*.

The approach is based on previous ideas but differs in its focus on fully automating the process of determining service boundaries. It does so by relying on static and dynamic software analysis. The case study uses MonoBreaker to assess the feasibility and merits of the approach. The decomposition obtained by the tool was regarded positively by the participants and seen as an improvement over using only static analysis. MonoBreaker is freely available, and the methodological design is documented to enable the replication of the case study by other researchers.

To improve these contributions, several aspects will be addressed in future work:

**Model Building.** The approach doesn't define a specific way to build the model of the application using the results of *static* and *dynamic* analysis. Future work will evaluate if other algorithms for calculating the weight of dependencies may perform better than our current implementation, which is currently based on a set of simple heuristics.

**Clustering Algorithms.** The approach is also not prescriptive of a particular clustering algorithm. It will be interesting to evaluate if others render better results than Girvan-Newman, the one currently used by MonoBreaker.

**Evaluation Metrics.** To enable a more objective evaluation of the proposed decomposition, the approach could be extended with service-based metrics – e.g., coupling and cohesion [21]. The approach of Taibi et al. [18] already includes metrics to rank decomposition candidates. A set of suitable service-based metrics for our approach would have to be determined, and can help to drive the search for better *model-building* and *clustering* algorithms.

**Comparison with Human Experts.** Future studies will evaluate if a data-driven approach such as ours is, not only able to automate the decomposition process fully, but will also provide a better decomposition than human experts.

**Further Studies.** More industry case studies will need to be conducted to improve our understanding of the effectiveness and limitations of the approach, ideally with a diverse and significant number of applications and participants.

**Representativeness of Sampled Data.** Future studies will compare the number of requests – per request type – that are received during the collection of operational data with those of more extended periods where operational data wasn't captured, but for which we are able to collect request statistics nonetheless. This will reinforce our confidence that the operational data collected is representative enough of a *normal* use of the application.

**Fully Automatic Decomposition.** MonoBreaker can identify file contents affected by the suggested decomposition, e.g., which class has to be extracted for each resulting service. The next step could be to suggest a sequence of lower-level refactorings required for the decomposition or even to automatically apply such refactorings to decompose the system.



**Acknowledgment.** João Paiva Pinto and Isabel Azevedo discussed different forms of this work with us. We thank them for all the precious feedback.

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project UIDB/50014/2020.

## References

1. Fowler, M.: Monolith first. Martin Fowler's Bliki (2015). <https://martinfowler.com/bliki/MonolithFirst.html>. Accessed 27 Nov 2019
2. Balalaie, A., Heydarnoori, A., Jamshidi, P.: Migrating to cloud-native architectures using microservices: an experience report. In: Celesti, A., Leitner, P. (eds.) ESOCC Workshops 2015. CCIS, vol. 567, pp. 201–215. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-33313-7\\_15](https://doi.org/10.1007/978-3-319-33313-7_15)
3. Fritzsche, J., Bogner, J., Wagner, S., Zimmermann, A.: Microservices migration in industry: intentions, strategies, and challenges. In: 2019 IEEE International Conference on Software Maintenance and Evolution (ICSME), pp. 481–490 (2019)
4. Ponce, F., Márquez, G., Astudillo, H.: Migrating from monolithic architecture to microservices: a rapid review. In: Proceedings of 38th International Conference of the Chilean Computer Science Society (SCCC 2019), Chile (2019)
5. Fritzsche, J., Bogner, J., Zimmermann, A., Wagner, S.: From monolith to microservices: a classification of refactoring approaches. In: Bruel, J.-M., Mazzara, M., Meyer, B. (eds.) DEVOPS 2018. LNCS, vol. 11350, pp. 128–141. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-06019-0\\_10](https://doi.org/10.1007/978-3-030-06019-0_10)
6. Newman, S.: Building Microservices: Designing Fine-Grained Systems, 1st edn. O'Reilly Media, Sebastopol (2015)
7. Newman, S.: Monolith to Microservices: Evolutionary Patterns to Transform Your Monolith, 1st edn. O'Reilly Media, Sebastopol (2019)
8. Richardson, C.: Microservices Patterns: With examples in Java. Manning, Shelter Island (2018)
9. Tyszberowicz, S., Heinrich, R., Liu, B., Liu, Z.: Identifying microservices using functional decomposition. In: Feng, X., Müller-Olm, M., Yang, Z. (eds.) SETTA 2018. LNCS, vol. 10998, pp. 50–65. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-99933-3\\_4](https://doi.org/10.1007/978-3-319-99933-3_4)
10. Martin, R.C.: Clean Architecture: A Craftsman's Guide to Software Structure and Design. Prentice Hall, Upper Saddle River (2017)
11. Evans, E.: Domain-Driven Design. Addison-Wesley Professional, Boston (2003)
12. Millett, S., Tune, N.: Patterns, Principles, and Practices of Domain-driven Design. Wiley, Hoboken (2015)
13. Nunes, L., Santos, N., Rito Silva, A.: From a monolith to a microservices architecture: an approach based on transactional contexts. In: Bures, T., Duchien, L., Inverardi, P. (eds.) ECSA 2019. LNCS, vol. 11681, pp. 37–52. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-29983-5\\_3](https://doi.org/10.1007/978-3-030-29983-5_3)
14. Chen, R., Li, S., Li, Z.: From monolith to microservices: a dataflow-driven approach. In: Proceedings of the 24th Asia-Pacific Software Engineering Conference – APSEC 2017, pp. 466–475. IEEE (2018)
15. Kälbener, L., Gysel, M.: Service cutter: a structured way to service decomposition. <https://servicecutter.github.io/>

16. Gysel, M., Kölbener, L., Giersche, W., Zimmermann, O.: Service cutter: a systematic approach to service decomposition. In: Aiello, M., Johnsen, E.B., Dustdar, S., Georgievski, I. (eds.) ESOC 2016. LNCS, vol. 9846, pp. 185–200. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-44482-6\\_12](https://doi.org/10.1007/978-3-319-44482-6_12)
17. Ren, Z., et al.: Migrating web applications from monolithic structure to microservices architecture. In: Proceedings of the Tenth Asia-Pacific Symposium on Internetware, series ICPS, Internetware 2018. ACM, New York (2018)
18. Taibi, D., Systä, K.: From monolithic systems to microservices: a decomposition framework based on process mining. In: Proceedings of the 9th International Conference on Cloud Computing and Services Science—CLOSER 2019 (2019)
19. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Phys. Rev. E* **69**(2) (2004). <https://journals.aps.org/pre/abstract/10.1103/PhysRevE.69.026113>
20. Likert, R.: A technique for measurement of attitudes. *Arch. Psychol.* **22**, 5–55 (1932)
21. Bogner, J., Wagner, S., Zimmermann, A.: Automatically measuring the maintainability of service- and microservice-based systems: a literature review. In: Proceedings of the 27th International Workshop on Software Measurement and 12th International Conference on Software Process and Product Measurement, series ICPS, IWSM Mensura 2017, pp. 107–115. ACM, New York (2017)