Implementing Microservices through Microtasks

Emad Aghayi, Thomas D. LaToza , Paurav Surendra, Seyedmeysam Abolghasemi

Abstract—Microtask programming is a form of crowdsourcing for programming in which implementation work is decomposed into short, self-contained microtasks. Each microtask offers a specific goal (e.g., write a unit test) as well as all of the required context and environment support necessary to accomplish this goal. Key to microtasking is the choice of workflow, which delineates the microtasks developers may complete and how contributions from each are aggregated to generate the final software product. Existing approaches either rely on a single developer to manually generate all microtasks, limiting their potential scalability, or impose coordination requirements which limit their effectiveness. Inspired by behavior-driven development, we describe a novel workflow for decomposing programming into microtasks in which each microtask involves identifying, testing, implementing, and debugging an individual behavior within a single function. We apply this approach to the implementation of microservices, demonstrating the first approach for implementing a microservice through microtasks. To evaluate our approach, we conducted a user study in which a small crowd worked to implement a simple microservice and test suite. We found that the crowd was able to use a behavior-driven microtask workflow to successfully complete 350 microtasks and implement 13 functions, quickly onboard and submit their first microtask in less than 24 minutes, contribute new behaviors in less than 5 minutes, and together implement a functioning microservice with only four defects. We discuss these findings and their implications for incorporating microtask work into open source projects.

Index Terms—Programming Environments, Behavior-Driven Development, Crowdsourcing, Microservices.

1 INTRODUCTION

MICROTASK programming is a form of crowdsourcing in which programming work is decomposed into short, self-contained units of work. Each microtask offers a specific goal (e.g., write a unit test) as well as all of the required context and environment support necessary to accomplish this goal [1]. Organizing software development work into microtasks can dramatically reduce joining time onto a project, as developers who need to learn less context to contribute are able to make meaningful contributions in as little as five minutes [2]. Microtask programming is often implemented through a specialized programming environment which is responsible for generating, presenting, and aggregating individual contributions [2], [3], [4], [5], [6], [7], [8], [9]. For example, in Apparition [3], a client developer narrates a description for a user interface in natural language, and crowd workers translates this descriptions into user interface elements, visual styles, and behavior. In CodeOn [9], a client developer narrates a request for help from their IDE, and crowd workers use this request and relevant context to author answers and code.

Key to the design of microtask programming environments is the choice of workflow, which delineates the microtasks developers may complete, the context that each offers, and the ways contributions from each are aggregated to generate the final software product. Workflow design is an important challenge in building effective crowdsourcing systems, as small changes in what workers are asked to do and the information that they have to do it may have large consequences in the overall efficiency and effectiveness of the approach [10], [11], [12]. Workflow design is closely tied to the goal that to be achieved, such as reducing time to market through increased parallelism or broadening participation by reducing context barriers to contributing.

While existing systems have demonstrated limited success in microtasking the construction of small programs, their workflows impose fundamental barriers to achieving increased parallelism and low-context contributions. Existing workflows either rely on a single developer to manually generate all microtasks [3] [5], limiting their potential scalability, or, for those automatically generating microtasks [13] [2], impose coordination requirements which limit their effectiveness. For example, in CrowdCode, each microtask requires developers to separately write a function or test, which creates coordination issues when these are not consistent [2]. Moreover, existing systems have focused on building either small GUI prototypes or small libraries, with limited support for interacting with external APIs, limiting their applicability to more realistic software projects.

In this paper, we explore a new form of programming workflow for behavior-driven microtask programming and apply this workflow to microtasking the implementation of microservices. Inspired by behavior-driven development (BDD) [14], work is organized into a single type of microtask where developers continue the implementation of an individual function by identifying, testing, implementing, and debugging an individual behavior within this function. For example, a developer might be given a function which deletes an item in a persistence store and be asked to implement a behavior to return an error if the specified item does not exist. In contrast to existing microtask workflows which distribute implementation, testing, and debugging into separate microtasks done by separate workers, our BDD workflow instead focuses work on a behavior and offers

- *E. Aghayi, Thomas D. LaToza, and P. Surendra are with the Department of Computer Science, George Mason University, 4400 University Drive, Fairfax, VA 22030.*
  E-mail: {eaghayi, tlatoza, psurendr}@gmu.edu
- *P. Surendra is a Software Engineer at Student Opportunity Center Inc. , 1501 Duke St 300, Alexandria, VA 22314*
- *M. Abolghasemi was a summer research intern at the Department of Computer Science, George Mason University, 4400 University Drive, Fairfax, VA 22030. Email: sabolgha@cs.odu.edu.*

an individual developer the opportunity for immediate feedback on their contributions by running their test against their code to identify and debug issues.

We apply this workflow to the task of implementing a microservice. In a microservice, a set of HTTP endpoints characterize requests that can be made and the responses to be generated. In our approach, a client first defines the intended behavior of the microservice by describing these endpoints. These endpoints are then mapped to specific functions which will implement each endpoint, generating initial microtasks for developers to implement each. As developers work through each microtask using the behavior-driven development workflow to identify, test, implement, and debug behaviors, they may choose to interact with third-party APIs. Key to many microservices is persistence, enabling a state to be saved between requests. In our approach, persistence is exposed through a third-party API, which workers can use as to store, update, and delete an object in the persistence store. After the crowd has implemented all behaviors for each of the endpoints, the client may then choose to deploy the microservice, creating a new deployment instance for the new microservice and deploying the code to a hosting site.

To explore this approach, we implemented a cloud IDE which offers a behavior-driven approach for microtasking the creation of microservices, *Crowd Microservices*. The cloud IDE includes an editor for clients to describe requested endpoints, automatic microtask generation and management, and an environment for identifying, testing, implementing, and debugging behaviors. To evaluate our approach, we conducted a user study with 9 crowd workers, who implemented a simple *Todo* microservice and test suite. Participants submitted their first microtask 24 minutes after beginning, successfully submitted 350 microtasks, implemented 13 functions and 36 tests, completed microtasks in a median time under 5 minutes, and correctly implemented 27 of 34 behaviors.

In this paper, we contribute 1) a new behavior-driven microtask workflow which offers crowd workers immediate feedback on their programming contributions as they work, 2) the first system for implementing microservices through microtask contributions 3) evidence that developers can make short BDD microtask contributions which together can be aggregated to create a small microservice. In the rest of the paper, we first illustrate *Crowd Microservices* through an example. We present the behavior-driven workflow for microtask programming and our approach for microtasking the creation of microservices and then present a user study evaluating the system. Finally, we conclude by discussing the opportunities and challenges this approach brings as well as potential future directions.

2 MOTIVATING EXAMPLE

Bob decides that he would like to build a new Todo web application, creating a new open source project dedicated to his idea. Lacking the time and skills to build a backend, he decides to focus on the frontend and open the backend to crowd contributions. He logs in to *Crowd Microservices* and uses the *Client-Request* page (Fig. 1) to create a new project. He names the application Todo, provides a brief description,

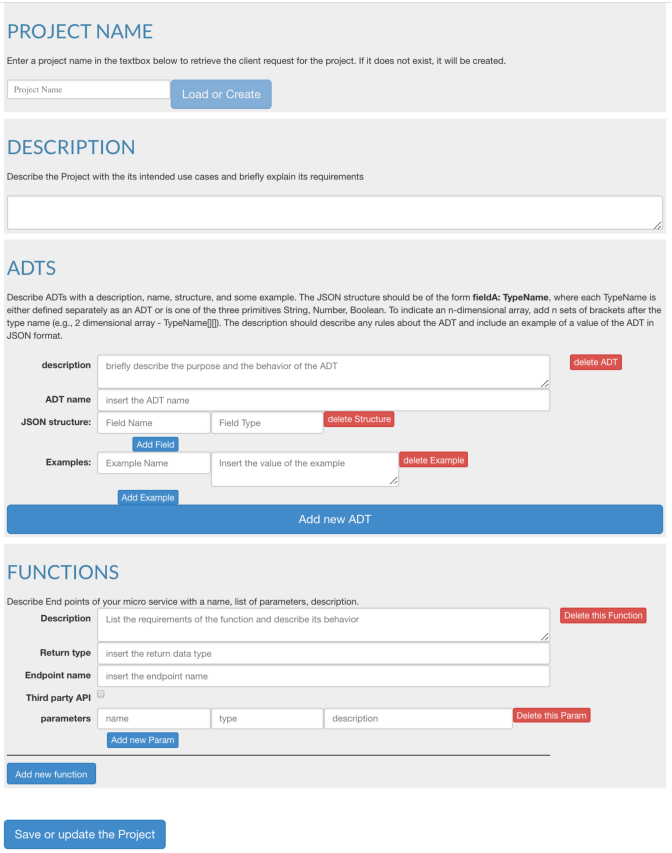

Fig. 1: In the *Client Request*, a client defines a microservice they would like created by describing a set of endpoints and corresponding data structures.

and defines the microservice through a set of endpoints. As some require complex JSON input data, he describes the structure of this data using two abstract data types.

Alice has a few minutes to do some programming and would like to contribute to an open source project. She logs in to *Crowd Microservices*, notices the Todo project has a number of open microtasks, and decides to contribute. As she has never before used *Crowd Microservices*, she first watches a short video[1] and reads a short tutorial, becoming familiar with the environment. Viewing the dashboard for the project, she quickly reads the brief description from the client and a list of descriptions of the functions created by the crowd so far. She clicks *Fetch a Microtask*, and *Crowd Microservices* assigns her an *Implement Function Behavior* microtask. Following the behavior-driven workflow, she first reads the description of the function in the comments above the body, identifying a behavior that does not yet seem to be implemented. Next, she writes a test. *Crowd Microservices* offers two types of tests, input/output tests and traditional unit tests. Alice chooses to write a traditional unit test, writing test code and running the test to verify that it fails. Next, Alice implements the behavior using the function editor. Alice tests her implementation, running the function's tests, but one fails. To understand why, Alice using the debugger to inspect values in the function's execution, hovering over several variables to see their values. Identifying the issue,

1. https://www.youtube.com/watch?v=qQeYOsRaxHc

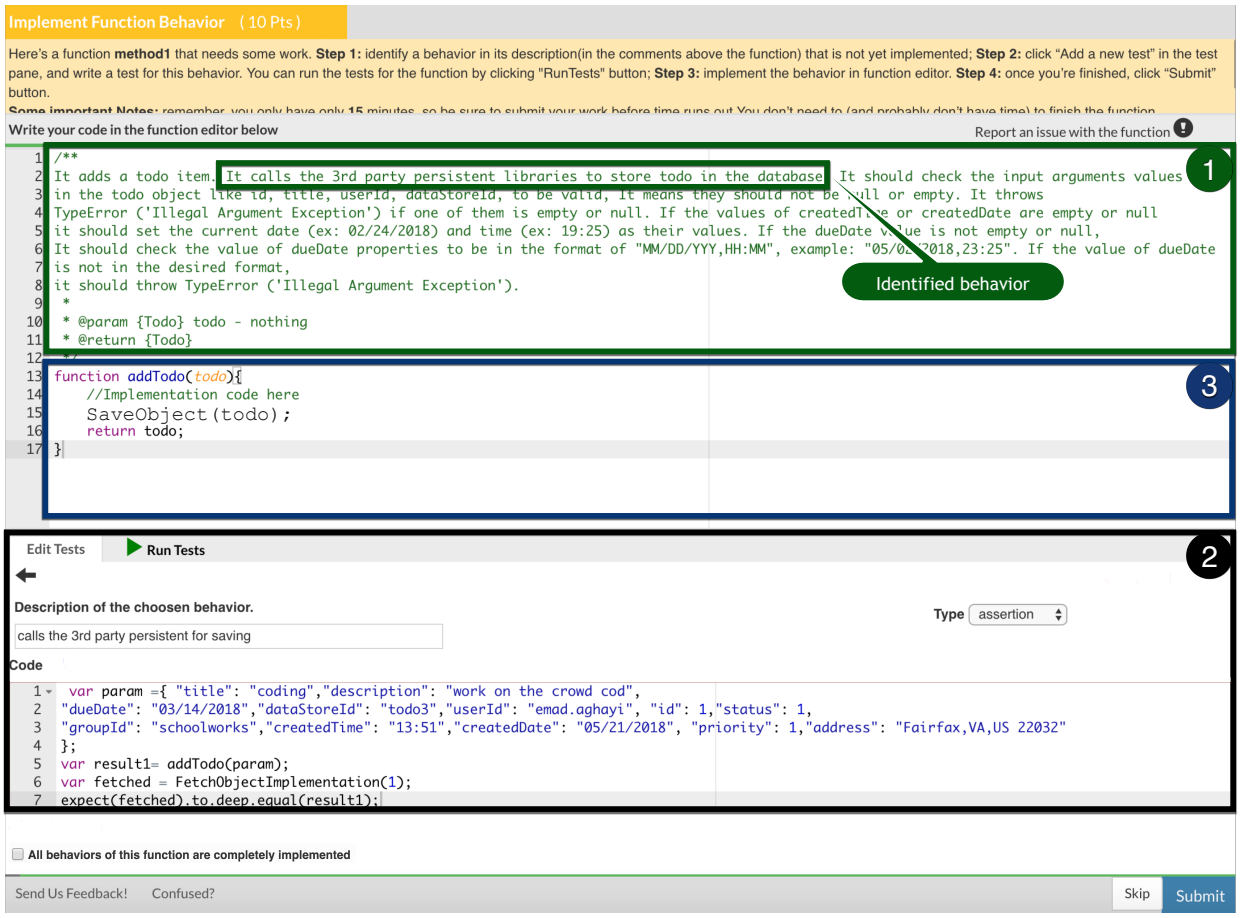

Fig. 2: An example of an *Implement Function Behavior* microtask. Workers first identify a behavior from the description of a function (1). They then write a test in the test editor to verify this behavior (2) and edit the existing code for the function to implement this behavior (3). Finally, they run the tests, debugging and fixing any issues they identify.

Alice fixes the problem, and finds that all of the tests now pass. She clicks *Submit*.

As Alice works, the crowd simultaneously completes other microtasks. After logging in, Dave is assigned an *Implement Function Behavior* microtask. Unsure how to implement any of the remaining behaviors, he clicks *Skip* and fetches another microtask. This time, he thinks that he can complete this microtask and writes a test and implementation. After a test fails, he realizes he does not know how to correctly call a third-party API function. Using the *Question and Answer* feature, he asks, *How can I store a todo object in the database?* A worker responds, and he figures out how to fix the problem. However, he then sees an alert: *You have spent more than 14 minutes on the current microtask, so try to submit your task in one minute before the system automatically skips it*. Wanting to submit his partially completed work, Dave submits with failed tests. Fetching another microtask and inspecting the implementation, he see that that all behaviors have already been implemented. So he clicks the corresponding checkbox (Figure 2) and submits.

After logging in, Charles is assigned a *Review* microtask which asks him to assess the behavior implemented by Alice. Charles reads the description of the function, a diff of the code written by Alice, and the tests. Looking at the implementation, he finds it seems incomplete, so he rates her contribution a 2 on a 5 point scale and gives her feedback, *"The behavior asked you to evaluate all input arguments of the function, but you just checked the validity of the date."* Charles submits, and Alice receives a notification that her work was reviewed and received 2 stars.

After being assigned an *Implement Function Behavior* microtask, Frank decides to implement a format check for the *todoDate* parameter. Believing this to be fairly complex, he decides it would be best implemented in a separate function. He invokes the *create a new function* feature, creating the function checkTodoDataFormat for others to implement. Specifying its behavior, he writes a description and signature. He then calls this new, currently empty, function from the body of the function he is working on. To verify his work, he runs the tests. But as checkTodoDateFormat is not yet implemented, his tests fail. Frank uses the *Stub* editor to replace the actual output with a stub value representing the desired output. This automatically replaces all calls to this function with the inputs and output Frank specifies. Frank runs the tests again, they pass, and he submits.

As the crowd works, each is assigned a score based on the ratings of their contributions. These scores are visible on a global *leaderboard* visible to the entire crowd, encouraging everyone to work hard to place higher.

While the crowd was working, Bob implemented the frontend, inserting requests based on the behavior of the endpoints he specified. After all the microtasks have been

completed and the implementation finished, he clicks a button to deploy the microservice. *Crowd Microservices* creates a Node.js Express[2] application, pushes the code to GitHub, and deploys this repository to the Heroku[3] hosting site. Now that the backend is complete, he runs the application and sees that the frontend user interactions are correctly handled by the Todo microservice.

3 RELATED WORK

Building on work in crowdsourcing in other domains, a number of approaches have been proposed for applying crowdsourcing to software engineering [15]. One category of approaches is microtask crowdsourcing, in which a large task is decomposed into several smaller, self-contained microtasks and then aggregated to create a finished product [16]. This approach was first popularized by Amazon's Mechanical Turk[4] and then used broadly in a number of systems [12]. In the following sections, we survey how microtasking has been applied to software engineering work, focusing on issues arising in decomposition, parallelism, fast onboarding, and achieving quality.

3.1 Decomposition, Context, and Scope

Decomposing work is a key challenge in microtasking in all crowdsourcing domains, as the choice of decomposition creates a workflow and the associated individual steps, the context and information required, and the types of contributions which can be made [10], [11], [12], [17]. Depending on the choice of microtask boundaries, contributions may be easier or harder, may vary in quality, and may impose differing levels of overhead.

Several systems have explored techniques for decomposing programming work into microtasks. Many systems rely on a developer or client to manually author each microtask. In systems such as CodeOn [9], developers working in a software project manually author requests for a completing a short programming task, which then become microtasks for workers. Apparition [3] offers a similar workflow for building UI prototypes in which requestors describe UI behavior in natural language todo items and workers view an implementation of the entire UI, select a microtask from a todolist, and implement the requested behavior for a UI element. While working on the microtask, workers interact with an individual element, but otherwise have a global view of the entire codebase. Nebeling et.al [7] proposed a technique for building web apps in which work is broken down by component and crowd members work individually in an isolated environment on each component. In CrowdCode, programming work is done through a series of specialized microtasks in which participants write test cases, implement tests, write code, look for existing functions to reuse, and debug [2], [13]. Other work has focused on applying microtasking to testing and verification. Chen and Kim [18] proposed a Puzzle-based Automatic Testing environment (PAT) that decomposes authoring complex mutation tests into short puzzle tasks. VeriWeb [8] decomposes authoring specifications for verification into smaller microtasks.

3.2 Parallelism and conflicts

By decomposing large tasks into smaller tasks, work can be done in parallel and completed faster. For example, crowdsourced testing platforms such as UserTesting [5], TryMyUI [6], and uTest[7] enable software projects to crowdsource functional and usability testing work by utilizing the crowd of tens or hundreds of thousands of contributors on these platforms.

Microtasking approaches for programming envision reducing the necessary time to complete programming tasks through parallelism. A key challenge occurs with conflicts, where two changes made at the same time conflict. For example, in traditional software development each contributor may edit the same artifacts at the same time, resulting in a merge conflict when conflicting changes are committed. This is an example of an optimistic locking discipline, where any artifact may be edited by anyone at any time. Due to the increased parallelism assumed and greater potential for conflicts, microtasking approaches often apply a pessimistic locking discipline, where microtasks are scoped to an individual artifact and further work on these artifacts is locked while they are in progress. For example, in Apparition [3] workers acquire write-locks mechanism to avoid conflicts. Similarly, in CrowdCode [2] each contribution occurs on an individual function or test, which is locked while a microtask is in progress. However, conflicts may still occur when decisions made in separate microtasks must be coordinated [4], [19]. In CrowdCode, conflicts occurred when workers completed separate microtasks to translate function descriptions into an implementation or tests and each made differing interpretations of a function description [2].

3.3 Fast Onboarding

In traditional open source development, developers onboarding onto a new software project must first complete an extensive *joining script*, installing necessary tools, downloading code from a server, identifying and downloading dependencies, and configuring their build environment [20], [21], [22]. Each of these steps serve as barriers which dissuade casual contributors from contributing. Researchers have explored designing environments to alleviate these barriers, often through dedicated, preconfigured, online environments. In the Collabode IDE, multiple developers can use an online editor to synchronously edit code at the same time, enabling new forms of collaborative programming [5], [6]. Apparition offers an online environment for building UI mockups, offering an integrated environment for authoring, viewing, and collaborating on the visual look and feel and behavior of UI elements [3]. CrowdCode offers an online preconfigured environment for implementing libraries, enabling developers to onboard quickly onto programming tasks [2].

2. https://expressjs.com
3. https://www.heroku.com
4. https://www.mturk.com
5. https://www.usertesting.com
6. https://www.trymyui.com
7. https://www.utest.com

Other work has explored preconfigured environments which enable teachers to manage a crowd of programming students. Codeopticon [23] enables instructors to continuously monitor multiple students and help them code. OverCode [24] and Foobaz [25] help mentors to cluster student code submissions, enabling teachers to give feedback on clusters of submissions rather than individual submissions. Codepilot [26] reduces the complexity of programming environments for novice programmers by integrating a preconfigured environment for real-time collaborative programming, testing, bug reporting, and version control into a single, simplified system.

3.4 Achieving Quality

Another important challenge in crowdsourcing systems is achieving high quality output from contributions of varying quality [27], [28], [29], [30]. There are many potential causes of low-quality contributions, including workers who do not have sufficient knowledge, who put forth little effort, or who are malicious. A study of the TopCoder[8] crowdsourcing platform revealed six factors related to project quality, including the average quality score on the platform, the number of contemporary projects, the length of documents, the number of registered developers, the maximum rating of submitted developers, and the design score [30]. In TopCoder, senior contributors assist in managing the process of creating and administering each task and ensuring quality work is done [31]. Interviews with several software crowdsourcing companies identified 10 methods used for addressing quality, including ranking/rating, reporting spam, reporting unfair treatment, task pre-approval, and skill filtering [29].

In systems involving microtask approaches to crowdsourcing, achieving quality is a particular challenge, as contributors may be less invested in the community and platform. Systems have explored a number of approaches for giving feedback to contributors. Systems in which microtasks are manually generated often rely on the requestor themselves to review and rate the quality of contributions [3], [6], [9], [32]. Systems where the requestor is less directly involved in work and microtasks are automatically generated may have crowd members review and offer feedback after contributions are made [2]. However, this approach is limited, as contributors who do not receive the traditional feedback offered in programming environments, such as syntax errors, missing references, and unit test failures, may submit work which contains these issues, which other contributors must then address later at higher cost [13].

4 CROWD MICROSERVICES SYSTEM

In this paper, we describe a behavior-driven development workflow for microtasking programming. In behavior-driven development, developers first write a unit test for each behavior they will implement, offering a way to verify that their implementation works as intended by running the test. As a workflow for microtasking, behavior-driven development offers a number of potential advantages. As developers work, they receive feedback before submitting, enabling the developer to revise their own work. Rather than requiring separate developers to test, implement, and debug a function through separate microtasks and coordinate this work to ensure consistency, a single contributor focuses on work related to an individual behavior within a function (Fig. 3).

8. https://www.topcoder.com/

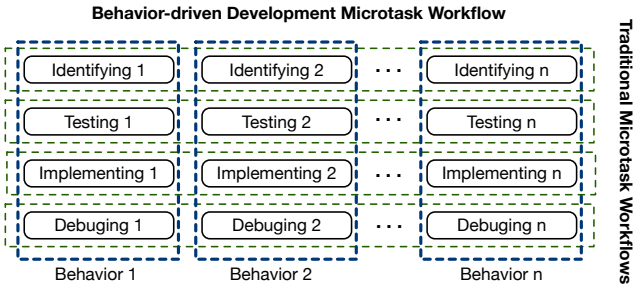

Fig. 3: Traditional microtask programming workflows organize work by the type of contribution made, including separate microtasks to do work such as identifying behaviors, testing, implementing, and debugging. In the behavior-driven microtask workflow, work is instead decomposed by behavior, where an individual microtask incorporates work of several types for an individual behavior.

We apply our approach to the problem of implementing microservices. Modern web application backends are often decomposed into numerous microservices, offering a natural boundary for crowdsourcing a module that is part of a larger system. In our approach, a client, for example a software development team, may choose to crowdsource the creation of an individual microservice. In situations where teams lack sufficient developer resources to complete work sufficiently quickly, a development team might choose to delegate this work to a crowd. Microservices offers a natural boundary between work done inside the development team and that done by the crowd. A client, acting on behalf of the software development team, may define the desired behavior of the microservice by defining a set of endpoints.

In the following sections, we first describe our behavior-driven workflow and then describe how it is applied to the task of implementing a microservice. Fig. 4 surveys our approach.

4.1 Workflow

In our behavior-driven microtask workflow, contributions are made through two microtasks: *Implement Function Behavior* and *Review*. Table 1 summarizes the context and possible contributions of each.

4.1.1 Interacting with Microtasks

After logging in, workers are first taken to a welcome page which includes a demo video and a tutorial describing basic concepts in the *Crowd Microservices* environment. After completing the tutorial, workers are taken to a dashboard page, which includes the client's project description, a list of descriptions for each function, and the currently available microtasks. The system automatically assigns workers a random microtask, which the worker can complete and

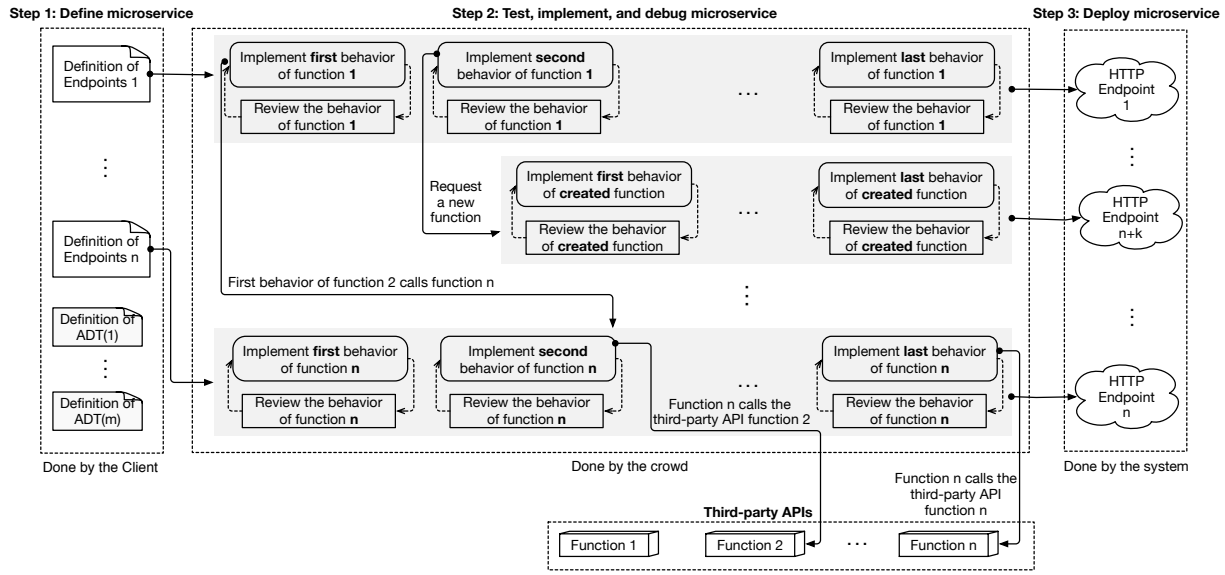

Fig. 4: Work begins with a client request which defines a microservice to implement through a list of endpoints. From this, the system generates initial microtasks to begin implementing each endpoint. In each microtask, an individual behavior is identified, tested, implemented, and debugged, including calling third-party API or crowd-generated functions as needed and requesting additional functions to be written. Each contribution generates a corresponding review microtask, where the contribution is either accepted or issues raised, generating a microtask to resolve any issues. After the crowd has completed implementation of the endpoints, the client may choose to deploy the microservice.

submit or skip. When workers begin a type of microtask which they have not previously worked on, workers are given an additional tutorial explaining the microtask. As participants work and are confused about a design decision to be made or about the environment itself, they may use the global *Question and Answer* feature to post a question, modeled on the question and answer feature in CrowdCode [4]. Posted questions are visible to all workers, who may post answers and view previous questions and answers.

As workers complete microtasks, each contribution is given a rating through a *Review* microtask. Ratings are then summed to generate a score for each worker. This score is visible to the entire crowd on a global *leaderboard*, motivating contributions and high quality work to place higher.

4.1.2 Implement Function Behavior Microtask

Developers perform each step in the *Implement Function Behavior Microtask* through the *Crowd Microservices* environment (Fig.2).

Identify a Behavior. Workers first work to identify single behavior that is not yet implemented from the comments describing the function located above its body (see (1) in Fig. 2). When the function has been completely implemented and no behaviors remain, the developer may indicate this through a checkbox.

Test the Behavior. In the second step, workers author a test as a simple input/output pair, specifying inputs and an output for the function under test, or as an assertion-based test (see (2) in Fig. 2). The worker may run the test to verify that it fails with the current implementation.

Implement the Behavior. The worker next implements their behavior using a code editor (see (3) in Fig.2). When the behavior to be implemented is complex, the worker may choose to *create a new function*, specifying a description, name, parameters, and return type. They may then call the new function in the body of their main function. After the microtask is submitted, this new function will be created, and a microtask generated to begin work on the function. In some cases, the signature of the function that a worker is asked to implement may not match its intended purpose, such as missing a necessary parameter. In these cases, the worker cannot directly fix the issue, as they do not have access to the source code of each call site for the function. Instead, they may report an issue, halting further work on the function. The client is able to see that this issue has been created, resolve the problem, and begin work again.

Debug the Behavior. A worker may test their implementation by running the function's unit tests. When a test fails, workers may debug by using the *Inspect code* feature to view the value of any expression in the currently selected test. Hovering over an expression in the code invokes a popup listing all values held by the expression during execution. In cases where a function that is called from the function under test has not yet been implemented, any tests exercising this functionality will fail. To enable the worker to still test their contribution, a worker may create a stub for any function call. Creating a stub replaces an actual output (e.g., an undefined return value generated by a function that does not yet exist) with an intended output. Using the stub editor, the worker can view the current output generated by a function call and edit this value to the intended output. This then automatically generates a stub. Whenever the tests are run, all stubs are applied, intercepting function calls and replacing them with the stubbed values.

Submit the microtask. Once finished, the worker may submit their work. To ensure that workers do not lock access to

TABLE 1: *Crowd Microservices* enables workers to make contributions through two types of microtask. Each offers editors for creating content, context views that make the task self-contained by offering necessary background information, and contributions the worker may make.

Microtasks	Editor	Context view	Possible Contributions
Implement Function Behavior	Function and test editor, test runner, stub editor	Description and signature of function, description of requesting function, ADTs	(1) Implement behavior (2) Report an issue in function (3) Mark function as completed
Review	Rating and review text	Description, signature, and implementation of function; function unit tests; ADTs	(1) Rating and review

an artifact for extended periods of time, each microtask has a maximum time limit of 15 minutes. Workers are informed by the system when time is close to expiring. When the time has expired, the system informs the worker and skips the microtask.

4.1.3 Review Microtask

In the *Review* microtask, workers assess contributions submitted by other workers. Workers are given a diff of the code submitted with the previous version as well as the tests of the function. Workers are asked to assign a rating of 1 to 5 stars. If the worker evaluates the work with 1 to 3 stars, the work is marked as needing revision. The worker then describes aspects of the contribution that they feel need improvement, and a microtask is generated to do this work. If the worker evaluates the submitted contribution with 4 or 5 stars, the contribution is accepted as is. In this case, the assessment of the work is optional, which will be provided back to the crowd worker that made the contribution.

4.2 Working with Microservices

Our approach applies the behavior-driven development workflow to implementing microservices. Fig. 4 depicts the steps in our process. The microservice is first described by a client through the *Client-Request* page (Fig 1). Clients define a set of endpoints describing HTTP requests which will be handled by the microservice. Each endpoint is defined as a function, specifying an identifier, parameters, and a description of its behavior. As endpoints may accept complex JSON data structures as input and generate complex JSON data structures as output, clients may also describe a set of abstract data types (ADTs). Each ADT describes a set of fields for a JSON object, assigning each field a type which may be either a primitive or another ADT. In defining endpoints, clients may specify the expected data by giving each parameter and return value a type.

After a client has completed a client request, they may then submit this client request to generate a new *Crowd Microservices* project. As shown in Step 2 of Fig. 4 submitting a client request generates an initial set of microtasks, generating an *Implement Function Behavior* microtask for each endpoint function. Workers may then log into the project to begin completing microtasks. As workers complete microtasks, additional microtasks are automatically generated by the system to review contributions, continue work on each function, and implement any new functions requested by crowd workers.

Microservices often depend on external services exposed through third-party APIs. As identifying, downloading, and configuring these dependencies can serve as a barrier to contributing, *Crowd Microservices* offers a pre-configured environment. As typical microservices often involve persisting data between requests, we chose to offer a simplified API for interacting with a persistence store. Through this API, workers can store, update, and delete JSON objects in a persistence store. Workers may use any of these API functions when working with functions and unit tests in the *Implement Function Behavior* microtask. Any schema-less persistence store may be used as an implementation for this API. In our prototype, a development version used by workers simulates the behavior of a persistence store within the browser and clears the persistence store after every test execution. In the production version used after the microservice is deployed, the persistence API is implemented through a Firebase store.

After the crowd finishes the implementation of a microservice, the client may choose to create and deploy the microservice to a hosting site (Step 3 in Fig. 4). Invoking the Publish command first creates a new node.js GitHub project which includes each function implemented by the crowd. For endpoint functions, the environment automatically generates an HTTP request handler function for the endpoint. For example, a signature *function createGroup(todoArray, groupId)* would generate a POST */createGroup* endpoint with the parameters as fields in the body of the request. Each endpoint then contains the implementation of the function defined by the crowd. Next, this GitHub project is deployed to a hosting site. In our prototype, projects are deployed to the Heroku hosting site. A new project instance is created, and the project deployed. After this process has been completed, the client may then begin using the completed microservice by making HTTP requests against the deployed, publicly available microservice.

4.3 Implementation

We implemented our approach as a prototype *Crowd Microservices* IDE. *Crowd Microservices* is a client-server application with three layers: 1) a web client, implemented in AngularJS, which runs on a worker's browser, 2) a backend, implemented in Node.js, and 3) a persistence store, implemented using Google Firebase [9]. *Crowd Microservices* automatically generates microtasks based on the current state of submitted work. After a client request defines endpoints, the system automatically generates a function and microtask to begin work on each. After an *Implement Function Behavior* microtask is submitted, the system automatically creates a *Review* microtask. After a *Review* microtask is submitted, an *Implement Function Behavior* is generated to continue work, if the contributor has not indicated that work is complete. If a review of an *Implement Function Behavior* contribution

9. https://firebase.google.com

indicates issues that need to be fixed, a new *Implement Function Behavior* microtask is generated, which includes the issue and an instruction to fix it. After a microtask is generated, it is added to a queue. When a worker fetches a microtask, the system automatically assigns the worker the next microtask and removes it from the queue.

5 EVALUATION

To investigate the feasibility of applying behavior-driven microtask programming to implementing microservices, we conducted a user study in which a crowd of workers built a small microservice. Specifically, we investigated 1) the ability of crowd workers to make contributions through a behavior-driven microtask workflow 2) the time necessary to contribute a test and implementation of a behavior, and 3) the feasibility of implementing and testing a microservice entirely through microtasks. We recruited 9 participants to build a small microservice for a Todo application and then analyzed participants' interactions with the environment and the resulting code they created.

5.1 Method

We recruited nine participants by advertising on Facebook, LinkedIn, and Twitter and through flyers (referred to as P1-P9). Participants were from the US, Spain, England, and India. Each had prior experience in JavaScript. Participants included one undergraduate student in computer science or a related field (P5), one instructor (P9), and seven graduate students in computer science or related fields. As typical in open contribution platforms, participants exhibited a diverse range of experience, with prior experience in JavaScript including less than 6 months (P1 and P2), 7-12 months (P3, P4, and P5), and more than 4 years (P6, P7,P8, and P9).

We split our study into two sessions to reduce participant fatigue as well as to simulate participants returning to work after a delay, as is common in microtasking. All nine participants participated in the first session and five (P1, P5, P6, P8, and P9) participated in the second session. The first session was 150 minutes, and the second 120 minutes. One participant (P8) left the second session early after approximately one hour. All participants worked entirely online at their own computers, and their interactions with other participants were only via the *Question and Answer* feature. Participants were paid 20 dollars per hour for their time through gift cards.

The crowd worked to build a microservice for the backend functionality of a Todo app. The microservice included functionality for creating, deleting, updating, fetching, reminding, and archiving of todo items. This functionality was described through 12 endpoints.

We gathered data from several sources. Before beginning the study, participants completed a short demographics survey. As participants worked, *Crowd Microservices* logged each microtask generated, submitted, and skipped as well as each change to a function or test, annotating each with a timestamp and participant id. At the end of the first session, participants completed a survey on their experiences with *Crowd Microservices*, and five of the participants participated in a short interview with the experimenter.

At the beginning of the study, participants logged in to *Crowd Microservices* and worked through tutorial content by watching an initial tutorial video, reading the welcome page, and then reading a second series of 6 tutorials on using the individual microtasks. Participants then began work by fetching a microtask. As in typical work contexts, participants were allowed to use Internet searches as they saw fit.

5.2 Results

5.2.1 Feasibility of behavior-driven microtasks

To investigate the ability of participants to use the behavior-driven microtask workflow, we examined the log data to determine how many microtasks participants were able to successfully complete during the two sessions as well as the functions and tests they created. Overall, participants successfully submitted 350 microtasks and implemented 13 functions, one of which was defined by the crowd (Table 2). Participants created a test suite of 36 unit tests, writing an average of 3 unit tests per function. We analyzed the number of lines of code in each function and test, counting the final numbers of lines in each at the end of the study. Participants wrote 216 lines of code, approximately 16 lines per function. Participants wrote 397 line of code in their test suite.

Several participants reported that identifying a behavior was not difficult. Three participants reported that they preferred to focus on easy behaviors first:

> "I chose the easiest behavior to implement first. This was usually to check if the input was null or empty. If that was already implemented, I just went in order." - (P5)

Others reported that some behaviors were not clear, leading them to focus first on those which were:

> "I chose based on being more clear and simple to me. Sometimes it wasn't clear what exactly that behavior means." - (P8)

Throughout both sessions, workers iteratively implemented and revised function implementations, reflecting contributions from several participants. Participants submitted 175 *Review* microtasks. In 82 of these, they accepted the contribution by giving it a rating of 3 or more stars. One participant reported that

> "Sometimes the feedback offered was helpful. But sometimes I would get 4 stars with no feedback, which was not helpful at all. I think it should be mandatory to write some feedback just so I can know where to improve." - (P5)

5.2.2 Speed of onboarding and contributing

During the two sessions, participants worked for a total of 31.5 hours. Participants spent 21 hours on microtasks that were submitted, including 39% of their time on *Implement Function behaviors* microtasks and 27% on *Review* microtasks. The remaining time was spent familiarizing themselves with the *Crowd Microservices* environment, completing the post-task survey, and working on microtasks that were skipped rather than submitted.

TABLE 2: Microtask completions, skips, and completion times

Microtasks types	Completed		Skipped		Median time (mm:ss)	
	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
Implement Function Behavior	112	63	39	22	4:12	3:27
Review	112	63	5	9	2:41	2:25
Total	224	126	44	31		
Overall Total	350		75			

After participants completed the tutorials and began work, participants spent additional time familiarizing themselves with the environment. On average, participants submitted their first microtask after 24 minutes. In some cases, participants skipped multiple microtasks to find an easy one with which to begin. Participants skipped 17% of all microtasks. Participants skipped microtasks when they were not able to complete a microtask or when they ran out of time.

Participants were able to complete both types of microtasks in a short period of time (Table 2). The median completion time was approximately 4 minutes for *Implement Function Behavior* and 3 minutes for *Review* microtasks. These completion times are similar to the under 5 minute median completion times reported for a microtask workflow which organizes implementation work into separate editing, debugging, and testing microtasks [2].

5.2.3 Feasibility of implementing microservices

To assess the feasibility of using a behavior-driven microtask workflow to implement microservices, we investigated the success of the crowd in building an implementation consistent with the described behavior in the client request. We first constructed a unit test suite, generating a set of 34 unit tests, which is publicly available as part of our replication package[10]. We found that the unit tests for 79% (27) behaviors passed, and the unit tests for 7 of the behaviors failed. To investigate the reasons for these failures, we examined the implementation created by the participants. We found that four of the failures were due to a single defect in one function involving a missing conditional, and the three remaining failures were either due to defects with behaviors not implemented correctly or not implemented. After addressing these issues, we found that all of the unit tests passed.

To further assess the implementation of the microservice built by the crowd, we used the final code written by participants to build a functioning Todo application. We first used the deploy feature in *Crowd Microservices* to deploy and host the microservice. We then implemented a Todo application frontend as a React application, using the deployed microservice as the backend. We found that, apart from the defects we described above, the Todo application worked correctly.

Crowd Microservices offers an API for interacting with a persistence store, which participants made use of in their implementation. Across the final implemented functions and unit tests, participants made 15 calls to the persistence API, or 1.25 per function. In some cases, participants interacted with the persistence store indirectly, by calling other functions implemented by the crowd which made use of the persistence store. When asked in the post-task survey, most participants reported that they used the persistence API without any problems. Some participants reported that additional documentation would be beneficial:

> "I used the API a little bit, and I felt like the documentation could be better with more examples." - (P5)

6 LIMITATIONS AND THREATS TO VALIDITY

Our study had several limitations and potential threats to the validity of the results.

In our study, we chose to recruit a wide range of participants, recruiting participants locally from our university as well as globally through social networking sites. This yielded participants with a wide range of backgrounds, with their experience in JavaScript ranging from 2 months to 6 years. We chose this process as it mirrors the process of an open call, where contributors with a wide range of backgrounds may contribute. However, in practice crowdsourcing communities may exist in many forms, attracting many novice contributors looking for an entry point into more challenging work or attracting experienced developers who attract other experienced developers. Our results might differ if workers were exclusively more or less experienced.

Another potential threat to validity is the choice of task. In selecting a task, we sought to identify a task that is representative of typical microservices that developers create. We chose the *Todo* application as a canonical example of a web application, often used to compare different approaches to building web applications[11]. Larger microservices may involve more complex endpoints where individual behaviors are more challenging to identify.

Our results might also vary with different contexts in which work took places. To simulate the constant process of hand-offs that occur in microtask work, where workers complete tasks that others began, we chose to have participants work synchronously, maximizing the number of hand-offs that occur. To simulate participants coming back to work that they had begun earlier, we divided our study into two sessions. Of course, in practice, microtask work involves less predictable schedules, where contributors may come and go at arbitrary times. This may introduce additional challenges, where new participants that are unfamiliar with either the environment or anything about the project are constantly introduced. On the one hand, this might reduce performance, as such participants are less experienced. On the other, compared to participants in our study who had access to no workers who were already familiar with the environment and project, such participants might have an easier time onboarding, as more experienced workers would

10. https://github.com/devuxd/crowd-todo-microservices

11. http://todomvc.com

be available to answer their questions. Better understanding the impact of transient behavior on microtasking in programming is an important focus for future work.

In microtask work, workers are often assumed to not have any prior information about the environment, the project, or other workers. However, in practice, over time participants do gain experience with the project and especially the environment. We would expect that over time a contributor to our environment would experience fewer challenges using the environment and become more productive, reducing the average time for completing microtasks or increasing the amount of work contributors complete in each microtask.

7 DISCUSSION

Microtask programming envisions a software development process in which large crowds of transient workers build software through short and self-contained microtasks, reducing barriers to onboarding and increasing participation in open source projects. In this paper, we explored a novel workflow for organizing microtask work and offered the first approach capable of microtasking the creation of web microservices. In our behavior-driven microtask workflow, each microtask involves a developer identifying, testing, implementing, and debugging an individual behavior within a single function. We found that, using this approach, developers were able to successfully submit 350 microtasks and implement 13 functions, quickly onboard and submit their first microtask in less than 24 minutes, contribute new behaviors in less than 5 minutes time, and together implement a functioning microservice back-end containing only 4 defects. Participants were able to receive feedback on their contributions as they worked by running their code against their tests and debugging their implementation to address issues.

While our method has demonstrated success in implementing an individual microservice, there remain a number of additional challenges to address before a large software project could be built entirely through microtasks. In our method today, the client performs the high-level design tasks of determining the endpoints, data structures, and other design work. Moreover, while microtasking reduces the context a developer must learn to successfully contribute to a project, this context is not zero. Developers must still learn about the function they are working on and the current state of its implementation. This overhead is visible in the productivity of work. In 21 hours working on submitted microtasks, 9 participants wrote only 216 lines of code and 397 lines of test code.

Other work has explored techniques for decomposing design work into microtasks, such as through structuring work around tables of design dimensions and design alternatives [33]. Such techniques might be adapted to a microtask programming workflow to enable the crowd to design the initial endpoints and data structures, as well as other high-level decisions, which might then be handed off to others who then implement this design. Similar to workflows such as TopCoder's, a senior crowd worker might also help ensure consistency across the design. Beyond upfront design, support is also necessary for maintenance situations where requirements change. This might result in changes in endpoints, data structures, and design decisions, requiring further downstream changes in the implementation. Such challenges might be addressed through new types of microtasks which identify changes, map these changes to specific impacted artifacts, and ask workers to update the corresponding implementation.

Short of microtasking entire projects, an alternative approach to incorporating microtask contributions into large software projects may be through microservices. Microservices offer a natural decomposition boundary in large web applications, grouping related functionality into a module which is accessed through a set of well-defined endpoints. While a large project might be architected, designed, tested, and built using traditional methods, individual microservices within this project might be built using *Crowd Microservices*. While traditional software development methods may be more efficient for developers who have months or years to build the context necessary to be effective, *Crowd Microservices* may enable new ways to involve contributors more quickly without the extended onboarding process traditionally associated with adding a developer to a software project. A project might discover a need for new functionality but not have the resources or ability to create this functionality as quickly as desired. *Crowd Microservices* might be used to offload some of this work to a crowd, enabling new workers to be brought into a project without needing to get up to speed on the entire project.

A key advantage of microtask programming approaches which incorporate automatic microtask generation is in enabling the potential for scale. Rather than requiring a single developer acting as a client to manually generate each microtask, microtasks are generated automatically as the crowd works. Rather than potentially exposing contributors to the entire codebase and all of its ongoing changes, contributors must only understand an individual function. In this way, in principle, large crowds might be able to work together to build large applications quickly. For example, if a microservice ultimately resulted in 1,000 behaviors being identified, each behavior could then be worked on by a separate developer in a separate microtask. To the extent that these 1,000 microtasks can be done in parallel, this would then enable software development to occur with 1,000 concurrent microtasks, dramatically decreasing the time to complete work. Of course, many sequential dependencies might still exist, where, for example, the necessary existence of a function is not revealed until a previous function has already been implemented. Understanding just how many sequential dependencies exist in software development work and how much parallelism is truly possible is thus an important focus for future work.

8 CONCLUSION

In this paper, we described an approach for implementing microservices through microtasks. Our results demonstrate initial evidence for feasibility, demonstrating that developers could successfully complete a microtask which asked them to identify, test, implement, and debug a behavior in less than 5 minutes as well as onboard onto the project in less than 30 minutes. Our results also offer the first example

of a microservice implemented entirely through microtask contributions. Important future work remains to investigate how this approach might be incorporated into a larger software project as well as exploring how a higher degree of parallelism might reduce the time to market in building software.

Acknowledgments

We thank the participants in the study for their participation. This work was supported in part by the National Science Foundation under grant CCF-1414197.

References

[1] T. D. LaToza and A. van der Hoek, "Crowdsourcing in software engineering: Models, motivations, and challenges," *IEEE software*, pp. 74–80, 2016.

[2] T. D. LaToza, A. Di Lecce, F. Ricci, B. Towne, and A. Van der Hoek, "Microtask programming," *Transactions on Software Engineering*, pp. 1–20, 2018.

[3] W. S. Lasecki, J. Kim, N. Rafter, O. Sen, J. P. Bigham, and M. S. Bernstein, "Apparition: Crowdsourced user interfaces that come to life as you sketch them," in *Human Factors in Computing Systems*, 2015, pp. 1925–1934.

[4] T. D. LaToza, A. D. Lecce, F. Ricci, W. B. Towne, and A. van der Hoek, "Ask the crowd: Scaffolding coordination and knowledge sharing in microtask programming," in *Symposium on Visual Languages and Human-Centric Computing*, 2015, pp. 23–27.

[5] M. Goldman, G. Little, and R. C. Miller, "Real-time collaborative coding in a web ide," in *Symposium on User Interface Software and Technology*, 2011, pp. 155–164.

[6] M. Goldman, "Software development with real-time collaborative editing," Ph.D. dissertation, Massachusetts Institute of Technology, 2012.

[7] M. Nebeling, S. Leone, and M. C. Norrie, "Crowdsourced web engineering and design," in *Web Engineering*, 2012, pp. 31–45.

[8] T. W. Schiller and M. D. Ernst, "Reducing the barriers to writing verified specifications," *Special Interest Group on Programming Languages Notices*, pp. 95–112, 2012.

[9] Y. Chen, S. W. Lee, Y. Xie, Y. Yang, W. S. Lasecki, and S. Oney, "Codeon: On-demand software development assistance," in *Conference on Human Factors in Computing Systems*, 2017, pp. 6220–6231.

[10] D. Retelny, M. S. Bernstein, and M. A. Valentine, "No workflow can ever be enough: How crowdsourcing workflows constrain complex work," *Conference on Computer-Supported Cooperative Work and Social Computing*, pp. 1–23, 2017.

[11] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, and K. Panovich, "Soylent: a word processor with a crowd inside," in *Symposium on User Interface Software and Technology*, 2010, pp. 313–322.

[12] A. Kittur, J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton, "the future of crowd work," in *Conference on Computer Supported Cooperative Work*, 2013, pp. 1301–1318.

[13] T. D. LaToza, W. B. Towne, C. M. Adriano, and A. Van Der Hoek, "Microtask programming: Building software with a crowd," in *Symposium on User Interface Software and Technology*, 2014, pp. 43–54.

[14] K. Beck, *Test-driven development: by example*. Addison-Wesley Professional, 2003.

[15] K. Mao, L. Capra, M. Harman, and Y. Jia, "A survey of the use of crowdsourcing in software engineering," *Journal of Systems and Software*, vol. 126, pp. 57 – 84, 2017.

[16] A. Doan, R. Ramakrishnan, and A. Y. Halevy, "Crowdsourcing systems on the world-wide web," *Communications of the ACM*, pp. 86–96, 2011.

[17] A. Kittur, B. Smus, S. Khamkar, and R. E. Kraut, "Crowdforge: Crowdsourcing complex work," in *Symposium on User Interface Software and Technology*, 2011, pp. 43–52.

[18] N. Chen and S. Kim, "Puzzle-based automatic testing: Bringing humans into the loop by solving puzzles," in *Automated Software Engineering*, 2012, pp. 140–149.

[19] A. L. Zanatta, L. Machado, and I. Steinmacher, "Competence, collaboration, and time management: Barriers and recommendations for crowdworkers," in *Workshop on Crowd Sourcing in Software Engineering*, 2018, pp. 9–16.

[20] I. Steinmacher, M. A. G. Silva, M. A. Gerosa, and D. F. Redmiles, "A systematic literature review on the barriers faced by newcomers to open source software projects," *Information and Software Technology*, pp. 67 – 85, 2015.

[21] G. Von Krogh, S. Spaeth, and K. R. Lakhani, "Community, joining, and specialization in open source software innovation: a case study," *Research Policy*, pp. 1217–1241, 2003.

[22] C. Jergensen, A. Sarma, and P. Wagstrom, "The onion patch: migration in open source ecosystems," in *Special Interest Group on Software Engineering Symposium and the European Conference on Foundations of Software Engineering*, 2011, pp. 70–80.

[23] P. J. Guo, "Codeopticon: Real-time, one-to-many human tutoring for computer programming," in *Symposium on User Interface Software and Technology*, 2015, pp. 599–608.

[24] E. L. Glassman, J. Scott, R. Singh, P. J. Guo, and R. C. Miller, "Overcode: Visualizing variation in student solutions to programming problems at scale," *Transactions on Computer-Human Interaction*, pp. 1–7, 2015.

[25] E. L. Glassman, L. Fischer, J. Scott, and R. C. Miller, "Foobaz: Variable name feedback for student code at scale," in *Symposium on User Interface Software and Technology*, 2015, pp. 609–617.

[26] J. Warner and P. J. Guo, "Codepilot: Scaffolding end-to-end collaborative software development for novice programmers," in *Conference on Human Factors in Computing Systems*, 2017, pp. 1136–1141.

[27] T. D. LaToza, W. B. Towne, A. van der Hoek, and J. D. Herbsleb, "Crowd development," in *Workshop on Cooperative and Human Aspects of Software Engineering*, 2013, pp. 85–88.

[28] K.-J. Stol and B. Fitzgerald, "Two's company, three's a crowd: a case study of crowdsourcing software development," in *Conference on Software Engineering*, 2014, pp. 187–198.

[29] M. Saengkhattiya, M. Sevandersson, and U. Vallejo, "Quality in crowdsourcing-how software quality is ensured in software crowdsourcing," Master's thesis, Department of Informatics, Lund University, 2012.

[30] K. Li, J. Xiao, Y. Wang, and Q. Wang, "Analysis of the key factors for software quality in crowdsourcing development: An empirical study on topcoder. com," in *Computer Software and Applications Conference*, 2013, pp. 812–817.

[31] K.-J. Stol and B. Fitzgerald, "Two's company, three's a crowd: A case study of crowdsourcing software development," in *Conference on Software Engineering*, 2014, pp. 187–198.

[32] S. W. Lee, R. Krosnick, S. Y. Park, B. Keelean, S. Vaidya, S. D. O'Keefe, and W. S. Lasecki, "Exploring real-time collaboration in crowd-powered systems through a ui design tool," *Computer-Supported Cooperative Work and Social Computing*, pp. 104:1–104:23, 2018.

[33] E. R. Q. Weidema, C. López, S. Nayebaziz, F. Spanghero, and A. van der Hoek, "Toward microtask crowdsourcing software design work," in *Workshop on CrowdSourcing in Software Engineering*, 2016, pp. 41–44.

Emad Aghayi received a BS degree in information technology at the Shiraz University of Technology in 2010 and an MS in information technology at the University of Tehran in 2014. He is a Ph.D. student in the Department of Computer Science at George Mason University. His research interests are at the intersection of software engineering and human-computer interaction.

Thomas D. LaToza received a Ph.D. in software engineering from the Institute for Software Research at Carnegie Mellon University in 2012 and degrees in psychology and computer science from the University of Illinois, Urbana-Champaign in 2004. He is currently an Assistant Professor of Computer Science at George Mason University. His research investigates how humans interacts with code and how programming tools can better support software developers in their work. He has served as co-chair of the Workshop on the Evaluation and Usability of Programming Languages and Tools, guest editor of the IEEE Software Theme Issue on Crowdsourcing for Software Engineering, and as co-chair of the International Workshop on Crowdsourcing in Software Engineering.

Paurav Surendra received a Masters degree in Software Engineering at George Mason University, Virginia and his Bachelors degree in Information Science and Engineering at Visvesvaraya Technological University in India. He is currently working as a Software Engineer at the Student Opportunity Center. His interests are in Software Architecture, Software Design, and Software Development for Web Platforms.

Seyedmeysam Abolghasemi received a Masters degree in Computer Science at Old Dominion University in 2017 and a Bachelors degree in Information Technology and Systems from Monash University in 2012. He is currently working as a Senior Software Developer at Old Dominion University. His area of interests is Software architecture, High-performance computing, and cryptography.