Analysis and Classification of Service Interactions for the Scalability of the Internet of Things

DOI:
10.1109/ICIOT.2018.00018

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Proceedings of 2nd IEEE International Congress on Internet of Things

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Analysis and Classification of Service Interactions for the Scalability of the Internet of Things

Damian Arellanes and Kung-Kiu Lau
School of Computer Science
The University of Manchester
Manchester M13 9PL, United Kingdom
{damian.arellanesmolina, kung-kiu.lau}@manchester.ac.uk

Abstract—Scalability is an important concern for Internet of Things (IoT) applications since the amount of service interactions may become overwhelming, due to the huge number of interconnected nodes. In this paper, we present an IoT scenario for real-time Electrocardiogram (ECG) monitoring, in order to analyze how well different kinds of service interactions can fulfill the scalability requirements of IoT applications.

Index Terms—Internet of Things (IoT), service interactions, scalability, Internet of Services (IoS)

I. INTRODUCTION

The Internet of Things (IoT) promises a new era in which not only people interact through Internet, but so do things. Currently, the number of connected devices worldwide is about 17 billion, and it is estimated that this number will grow by a factor of 1.82 in the next three years [1]. For this reason, scalability in terms of the size of IoT applications, rather than vertical or horizontal scalability [2], is an important concern.

For this kind of scalability, four crucial desiderata has been identified: explicit control flow [3], separation between control and computation [4], decentralization [5] and location transparency [6]. In this paper, we analyze how well different kinds of service interactions can fulfill these scalability requirements.

Service interactions play a central role in the Internet of Services (IoS) [7] which will be a key enabler of the IoT goals. IoT services interact in different ways to achieve a common goal in a specific application. Despite an increasing number of proposed network protocols for IoT, there is a lack of understanding about which service interactions best fulfill the scalability requirements of IoT applications.

The rest of the paper is structured as follows. Sect. II presents an IoT scenario for real-time Electrocardiogram (ECG) monitoring. Sect. III describes our classification of service interactions. Sect. IV presents the results of our analysis. Sect. V presents a discussion of our results. Finally, Sect. VI presents the conclusions and the future work.

II. IoT SCENARIO: ELECTROCARDIOGRAM MONITORING NETWORK (EMoNet)

This section introduces a running example for the rest of the paper. The example is an IoT scenario for real-time Electrocardiogram (ECG) monitoring: Electrocardiogram Monitoring Network (EMoNet). EMoNet is a network deployed in a smart city, consisting of patients with cardiac diseases, plenty of ambulances moving around the city, patients’ smartwatches and wearable ECG sensors. Fig. 1 depicts the workflow of EMoNet which corresponds to a timing task triggered every 3 minutes for a particular patient. It basically consists of pulling and analyzing ECG data, and requesting the nearest ambulance if the patient has heart attack signs.

The EMoNet workflow involves four independent IoT nodes shown in Fig. 2: a wearable ECG sensor installed on a patient’s chest, the patient’s smartwatch, a healthcare cloud and an ambulance.

Fig. 1. EMoNet workflow.

Fig. 2. Nodes and services involved in the EMoNet workflow.

The wearable ECG sensor provides the Heart Rate History service as an interface for the records of the electrical activity of a patient’s heart. The smartwatch provides the ECG Analysis service that determines if a patient is showing signs of a heart attack. A healthcare cloud provides the Emergency service to find the nearest ambulance and request it immediately. Ambulances provide the Assistance service to attend to those patients in need on-site. For simplicity, we assume that these services dispatch many requests concurrently. In the next section, we will describe different ways to realize the EMoNet workflow using these services.

III. SERVICE INTERACTION SCHEMAS

IoT services provide low-level functionality implemented in nodes [8]. Resource-constrained nodes (e.g., a pulse sensor) provide fine-grained services for basic functionality (e.g., fetching sensor data). Non resource-constrained nodes (e.g., a smart TV) may offer coarse-grained services in addition.
IoT services interact via a network in order to realize complex functionality. Services can interact by message passing, event exchanges, or any combination thereof. In order to determine what kind of interactions best fulfills the scalability requirements of IoT, we have classified service interactions into four schemas: (i) direct service interactions, (ii) indirect service interactions, (iii) exogenous service interactions and (iv) event-driven service interactions.

Schemas (i), (ii) and (iii) are based on message-passing, where there are two roles: service sender and service receiver. A service sender accesses functionality offered by a service receiver, by passing a message (expressing control) via the network. Schema (iv) is based on events so a service registers itself with events that will be produced by another service(s). In this section, we describe these four schemas in more detail.

Microservice Architecture [9] has gained considerable attention in the last few years, and is becoming increasingly important and popular for the development of IoT applications [10]. Every Microservice Architecture is a Service-Oriented Architecture (SOA), but not the other way round [11]. Hence, the service interaction schemas presented in this section can be used interchangeably in both Microservices and traditional SOA services.

A. Direct Service Interactions

The direct service interaction schema consists of sending a message (e.g., a XML-based document or a JSON-based document) from a sender to a receiver with no mediator between them [12]. The sender interacts with the receiver using Remote Procedure Calls (RPC) [13] or REST API calls over HTTP [14]. RPC is akin to method invocations in traditional Object-Oriented programming languages, the main difference being that the invoked procedures may reside at different network addresses. REST does not require to know procedure names in advance, but only the location of external resources that can be manipulated using HTTP methods. Direct interactions are typically done using the request-response pattern [15].

Fig. 3 illustrates direct interactions for the EMoNet workflow. **ECG Analysis** triggers the control flow periodically by passing control to **Heart Rate History** so as to get the last sensor reading. Then, **Heart Rate History** returns the control to **ECG Analysis**. If there are signs of a heart attack, **ECG Analysis** passes control to the **Emergency** service which forwards control to the **Assistance** service of the nearest ambulance. Control is returned to **ECG Analysis**, after passing through the **Emergency** service. A glance at Fig. 3, reveals that even though a service bus provides indirection between senders and receivers, control and data are originated in service computation, and data follows control.

B. Indirect Service Interactions

The indirect interaction schema consists of using a service bus to broker sender requests, locate an appropriate receiver, transmit requests, and send responses back to senders. Since it passes messages between senders and receivers, a service bus can be thought of as a universal connector that provides a level of indirection between services [19], [20].

Fig. 4 illustrates indirect interactions for the EMoNet workflow, where **ECG Analysis** triggers control flow periodically. EMoNet services register their interfaces with a service bus that forwards control (and data) originated by **ECG Analysis** and **Emergency**, and sends back control (and data) from **Heart Rate History**, **Assistance** and **Emergency**, respectively. A glance at Fig. 4, reveals that even though a service bus provides indirection between senders and receivers, control and data are originated in service computation, and data follows control.

C. Event-Driven Service Interactions

The event-driven interaction schema is based on the publish-subscribe pattern [15] so there are two roles: producer (i.e., service sender) and consumer (i.e., service receiver). Producers trigger events (perhaps carrying data) which are then stored in a queue. Consumers can subscribe to the events they are interested in, retrieve those events from the queue and react
Event-driven interactions can be done with or without a service bus. **P2P event-driven interactions** enable every service to be responsible of its own queue, so events are exchanged with no mediator. ZeroMQ is the most popular library to realize this interaction schema.\(^1\) Fig. 5(a) shows P2P event-driven interactions for the EMoNet workflow.

**ECG Analysis** periodically gets the last sensor readings by consuming events produced by **Heart Rate History**. If it detects a heart attack, **ECG Analysis** announces an emergency situation by producing an event for **Emergency**. After determining the nearest ambulance, **Emergency** produces an event that is consumed only by the **Assistance** service of that ambulance. Finally, **Assistance** produces an event for **ECG Analysis** to indicate the status of the current emergency.

**Broker-based event-driven interactions** use an event bus to manage event queues for a particular IoT application. An event bus is generally implemented using a messaging protocol such as the Advanced Message Queuing Protocol (AMQP) or the Message Queue Telemetry Transport (MQTT). RabbitMQ is the most popular implementation of the AMQP protocol.\(^2\) The EMoNet services shown in Fig. 5(b) interact in the same way as the ones shown in Fig. 5(a), with the fundamental difference that events are now stored in the queue of an event bus.

**Event-driven interactions are preferred to direct interactions for implementing Microservice choreographies** [9], [22], [11]. Microservices use the strategy *smart endpoints and dumb pipes* [9] to define event-driven interactions in endpoints.

There is an increasing trend to use event-driven interactions for the exchange of data between IoT applications [24], [25]. In fact, the author in [27] found that a vast majority of current IoT platforms provide support for the event-driven interaction schema. In particular, broker-based event driven interactions are gaining considerable attention since MQTT was particularly designed for resource-constrained nodes [28], [29].

D. **Exogenous Service Interactions**

The exogenous service interaction schema enables a coordinator to define interactions (in the form of a workflow) over mutually anonymous services or other coordinators. Thus, control is always originated in coordinators and services do not interact with each other [30], [31].

Exogenous interactions can be done in one or multiple levels. One-level exogenous interactions are realized by orchestration [32], where the coordinator is a workflow engine running in a specialized server. Fig. 6(a) shows one-level interactions for the EMoNet workflow.

**EMoNet Workflow Engine** is the coordinator for all the involved services. It passes control to **Heart Rate History** and **ECG Analysis** sequentially, in order to pull and analyze the last sensor readings. Then, according to the results of the analysis, the coordinator decides if there are signs of a heart attack. If so, the coordinator passes control to **Emergency** and **Assistance**, in that order. A glance at Fig. 6(a), reveals that control is always originated in the coordinator, and services are only concerned with returning control and data after performing some computation.

Multi-level exogenous interactions are done by hierarchical orchestration [33], [5] or exogenous connectors [34], [35]. In this schema, multiple coordinators create a hierarchy of service interactions. Unlike, one-level exogenous interactions, in this schema control flows over multiple distributed coordinators.

Hierarchical orchestration [5] has multiple workflow engines, each of them responsible for the interaction of services or other workflow engines. In other words, it allows nesting a workflow within another workflow. Fig. 6(b) depicts a two-level hierarchical orchestration for the EMoNet services. **EMoNet Workflow Engine** coordinates the execution of coordinators **Monitoring Workflow Engine** and **Decision-Making Workflow Engine**. First, **EMoNet Workflow Engine** passes control to **Monitoring Workflow Engine** which is responsible for the interactions of the services **Heart Rate History** and **ECG Analysis**. Once control is returned from **Monitoring Workflow Engine** to **EMoNet Workflow Engine**, the latter passes control to **Decision-Making Workflow Engine**. If **Decision-Making Workflow Engine** determines that there are signs of a heart attack, it passes control to **Emergency** and **Assistance** sequentially. Finally, the control flow ends when the **Decision Making Workflow Engine** returns control to **EMoNet Workflow Engine**. Fig. 6(b) shows that a
Fig. 6. Exogenous service interactions for the EMoNet workflow.

Exogenous connectors are lightweight distributed coordinators that define micro-workflows. Fig. 6(c) illustrates how exogenous connectors coordinate service interactions for the EMoNet workflow. The control flow is the same as the one depicted in Fig. 6(b) for hierarchical orchestration. Unlike hierarchical orchestration, where control can be passed from a coordinator to a service, in exogenous connectors control is only passed between coordinators (as data is an orthogonal dimension). Furthermore, the composition of two services results in a composite service (not a workflow) that preserves all the operations from the sub-services. Another difference is that coordinators do not need any specialized server as they can run in any IoT node (including resource-constrained nodes). For the EMoNet workflow, Heart Rate History and ECG Analysis are composed into Monitoring Composite which is deployed in a smart t-shirt; similarly, Emergency and Assistance are composed into Decision-Making Composite which is deployed in the Google Cloud. Exogenous connectors allow composite services to be further composed into even bigger services. For example, the Monitoring Composite and the Decision-Making Composite are composed into the EMoNet composite which is deployed in the patient’s mobile device.

Due to the popularity of one-level exogenous interactions in SOA, in the last years we have seen the emergence of software platforms for IoT service orchestration. Examples include Intel IoT SOL (Service Orchestration Layer) [36] and Compose PaaS [37]. To the best of our knowledge, there are currently no IoT platforms for multi-level exogenous interactions.

IV. EVALUATION AND RESULTS

This section presents the results of a qualitative evaluation of our service interaction schemas. A tick mark indicates that a specific interaction schema fulfills the requirement being analyzed, while a cross mark means the opposite. NA means that the analysis is not applicable for a particular interaction schema. In order to determine which schema best fulfills the scalability requirements of IoT applications, we specifically investigate the following research questions:

- **RQ1:** Which schemas allow the visualization of control flow?
- **RQ2:** Which schemas allow a separate reasoning between control and computation?
- **RQ3:** Which schemas allow decentralized interactions?
- **RQ4:** Which schemas enable services that are unaware of the location of other services?

A. RQ1: Explicit Control Flow vs Implicit Control Flow

Control flow can be explicit or implicit. Explicit control flow is visible as an entity defines the order in which individual services are executed. Conversely, implicit control flow is opaque since it is not defined anywhere, but it is implicit in the interactions of the participant services. Table I shows that event-driven interactions do not support visible control flow as it is implicit in the collaborative exchange of events (see Fig. 5) [16], [26]. In both direct interactions and indirect interactions, services are the entities who control the application flow, e.g., ECG Analysis defines a guard to execute Emergency when a heart attack is detected (see Figs. 3 and 4). In exogenous interactions, coordinators define control, e.g., EMoNet Workflow Engine defines control structures to realize one-level exogenous interactions for the EMoNet workflow (see Fig. 6(a)).

The amount of service interactions in IoT applications may become overwhelming due to the huge number of nodes involved. Since it is not visible, implicit control flow limits the scalability of IoT applications as the number of services grows and the complexity of service interactions increases.
Implicit control flow has been an issue for software companies over many years and it is undoubtedly a barrier for IoT. For instance, Netflix has recently expressed that implicit control flow limits the scalability of distributed applications, as they found that process flows are spread across multiple applications and it is difficult to monitor workflow processes. As an attempt to visualize control flow, Netflix recently moved (provided by a service), and control defines the logic to Emergency when a heart attack is detected (see Figs. 3 and 4).

This evidently makes it hard to monitor workflow execution, visual representation of the blueprint with the paths the control for the conditional control structure that passes control to ECG Analysis interactions or direct interactions, control is always originated in the sender’s computation. For that reason, ECG Analysis and ECG Planning must be changed to accommodate the new requirement. In particular, the conditional control structure is removed from ECG Analysis and added into ECG Planning which is now responsible for passing control to Emergency (when a heart attack is detected). For that reason, ECG Analysis is not reused in the new application.

Our analysis of the separation between control and computation is not applicable for event-driven interactions, since control flow is implicit in this schema. Nevertheless, in event-driven interactions, events are originated inside service computation (see Fig. 5). For example, Emergency and ECG Planning would require changes in their computation so as to accommodate the planning phase. In particular, ECG Planning needs to consume the events produced by ECG Analysis, while Emergency needs to consume the events produced by ECG Planning. For that reason, Emergency is not reused in the new application.

Table II shows that only exogenous interactions separate control from computation, as control is always originated in the coordinator(s) (see Fig. 6). In contrast to the rest of the schemas, exogenous interactions do not require changing any service to support the planning phase, but only changing the application logic defined in the coordinator(s). Thus, as business requirements change, developers can manage changes in the application logic without taking care of the functionality provided by IoT nodes [40].

<table>
<thead>
<tr>
<th>Service interactions</th>
<th>Separation between control and computation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct interactions</td>
<td>✗</td>
</tr>
<tr>
<td>Indirect interactions</td>
<td>✗</td>
</tr>
<tr>
<td>Event-driven interactions</td>
<td>ECG Analysis, Broker-based</td>
</tr>
<tr>
<td>Event-driven interactions</td>
<td>ECG Planning</td>
</tr>
<tr>
<td>Exogenous interactions</td>
<td>One-level</td>
</tr>
<tr>
<td></td>
<td>Multi-level</td>
</tr>
</tbody>
</table>

**TABLE II**

**SEPARATION BETWEEN CONTROL AND COMPUTATION IN SERVICE INTERACTION SCHEMAS.**

When events or control are originated in service computation, an application workflow is embedded in the code of plenty of services. This is in fact one of the reasons for which Netflix stop using event-driven interactions. Exogenous interactions is the only schema that enables the development of workflow-agnostic services, as a consequence of the separation between control and computation. For that reason, Netflix preferred the use of exogenous interactions to event-driven interactions.

C. RQ3: Decentralized Service Interactions

Service interactions can be centralized or decentralized. On the one hand, centralized service interactions means that control, events (or even data) pass through a single central entity. On
the other hand, decentralized service interactions means that control, events (or even data) are passed in a P2P fashion, since workflow (expressed by control or events) is distributed over two or more entities.

Table III shows that indirect interactions, one-level exogenous interactions (i.e., orchestration) and broker-based event-driven interactions are centralized schemas. Indirect interactions require a service bus for passing control and data between services (see Fig. 4). Broker-based event-driven interactions use an event bus to handle events (see Fig. 5(b)). In one-level exogenous interactions, a central engine defines a workflow for passing control (and frequently data) between services (see Fig. 6(a)).

Table III shows that decentralization is present in direct interactions, multi-level exogenous interactions (i.e., hierarchical orchestration and exogenous connectors) and P2P event-driven interactions. Direct interactions do not require any mediator for passing control between services (see Fig. 3). In multi-level exogenous interactions, coordinators are the only entities that pass control to services or other coordinators (see Figs. 6(b) and 6(c)). Similarly, P2P event-driven interactions do not rely on a bus for event management, as every service is responsible of its own queue (see Fig. 5(b)).

D. RQ4: Location Transparency

IoT is highly dynamic due to the intermittent connection and spontaneous failures of IoT nodes, resulting in nodes (and therefore services) frequently changing locations over time. For that reason, churn is one of the main challenges of IoT applications as they usually operate in a dynamic and uncertain environment [6], [45]. For example, the Wearable ECG Sensor is a resource constrained-node that can run out of battery with the subsequent disconnection from the network. Similarly, an Ambulance may experience frequent disconnections due to its high mobility.

Service location transparency is crucial to mitigate churn in IoT applications, as it enables services to be unaware of the physical location of other services. Table IV shows that indirect interactions, event-driven interactions and exogenous interactions provide location transparency. In indirect interactions, the service bus is the only entity aware of services’ locations. In event-driven interactions, publishers and subscribers do not know the location of one another, but they only know what events to produce and consume, respectively. In exogenous interactions, coordinators encapsulate services’ locations as they are responsible for service interactions.

Table IV shows that location transparency is crucial to mitigate churn in IoT applications. As it enables services to be unaware of the physical location of other services. Table IV shows that indirect interactions, event-driven interactions and exogenous interactions provide location transparency. In indirect interactions, the service bus is the only entity aware of services’ locations. In event-driven interactions, publishers and subscribers do not know the location of one another, but they only know what events to produce and consume, respectively. In exogenous interactions, coordinators encapsulate services’ locations as they are responsible for service interactions.

Table IV Location Transparency in Service Interaction Schemas.

<table>
<thead>
<tr>
<th>Location transparency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct interactions</td>
</tr>
<tr>
<td>Indirect interactions</td>
</tr>
<tr>
<td>Event-driven interactions</td>
</tr>
<tr>
<td>Event-driven interactions</td>
</tr>
<tr>
<td>Exogenous interactions</td>
</tr>
<tr>
<td>Exogenous interactions</td>
</tr>
</tbody>
</table>

TABLE IV Location Transparency in Service Interaction Schemas.
V. Discussion

Table V summarizes the results of our qualitative evaluation. It particularly shows how well service interaction schemas fulfill the scalability requirements of IoT applications: explicit control flow, separation between control and computation, decentralized interactions, and service location transparency.

Direct interactions, indirect interactions and P2P event-driven interactions cover 50% of the requirements, respectively. Broker-based event-driven interactions is the worst schema since it only meets 25% of those requirements. Lacking only decentralization, one-level exogenous interactions cover 75% of the desiderata. Multi-level exogenous interactions is the only schema that fulfills all the scalability requirements of IoT applications.

In some scenarios, it could be useful to combine interaction schemas. For example, in order to provide asynchronous interactions in EMoNet, services can combine event-driven interactions with direct interactions. ECG Analysis can interact via an event bus with both Heart Rate History and Emergency, whilst Emergency can use direct interactions to request the Assistance service of the nearest ambulance.

A service bus can be [20]: (i) distributed, (ii) with technical intelligence or (iii) with business intelligence. Options (i) and (ii) are used only for data and control routing, whilst (iii) can be used to define coordination logic in addition [21]. Even though it is typically used only for straightforward workflows, (iii) is a special case of one-level exogenous interactions.

Although the Microservices community recommends the avoidance of (iii) as they do not want business logic embedded in a service bus [21], there is an increasing tendency to use exogenous interactions for Microservices in traditional SOA applications [47], [3]. By contrast, in the context of IoT, event-driven interactions are currently more popular. However, given the advantages of exogenous interactions, as evidenced by their increasing adoption in traditional SOA applications, we envision that exogenous interactions will increase in popularity in Microservice-based IoT applications in the next years.

A Distributed Service Bus (DSB) [48] is often seen as a decentralized approach due to the existence of a federation of brokers. However, it consists solely of a distribution of middleware components over different nodes. According to our view of decentralization presented in Sec. IV-C, the existence of an intermediary (or intermediaries) for service interactions leads to a centralized approach. As we noted in Sec. IV-C, a purely decentralized approach removes the need of a middleman (or middlemen) which, among other issues, introduces an extra network hop for service interactions.

In order to achieve decentralization, the Microservices community fosters direct interactions between Microservices. Nevertheless, direct interactions impact performance because a connection must be open for the entire duration of an interaction, and a Microservice participant needs a reference (i.e., a client library) for every Microservice it communicates directly with. Maintaining references to other Microservices is costly. Furthermore, a HTTP connection may become a bottleneck, especially for long running Microservices. This is undoubtedly a problem for resource-constrained IoT nodes which do not have communication and storage capabilities to support long-running transactions or to store multiple references. To solve this issue, the Internet Engineering Task Force (IETF) has developed the Constrained Application Protocol (CoAP) [49]. CoAP has been proved to be a simpler and more cost-efficient alternative to HTTP/REST in several IoT scenarios involving resource-constrained nodes [50]. Nevertheless, CoAP does not support the separation between control and computation.

The separation between control and data is also crucial for the scalability of IoT applications. It means that data is never passed alongside control. This separation of concerns allows a separate reasoning between data flow and control flow, which could result in the development of an efficient data exchange approach. For instance, a P2P data exchange can be used to reduce the number of network hops, thereby avoiding network congestion as shown by [43]. The separation between control and data also enables reuse of data flow without the need of modifying control flow. Hence, data flow and control flow can evolve separately. Exogenous connectors in multi-level exogenous interactions provide semantics for the separation between control and data. Although data typically follows control in orchestration approaches, the separation between control and data has already been done in hierarchical orchestration [43] and traditional orchestration [51]. For the EMoNet workflow based on exogenous interactions, we assumed that there is no separation between control and data.

VI. Conclusions and Future Work

In this paper, we classified and analyzed service interactions into four schemas, namely direct interactions, indirect interactions, event-driven interactions and exogenous interactions.

We conducted a qualitative evaluation to determine which interaction schema best fulfills the scalability requirements of
IoT applications: explicit control flow, decentralized interactions, separation between control and computation, and service location transparency. We showed that multi-level exogenous interactions is the most promising schema since it meets all the desiderata for the scalability of IoT applications.

Network performance is another aspect that needs to be considered when tackling scalability. We would like to conduct experiments to quantitatively evaluate the throughput of the service interaction schemas presented in this paper.

To the best of our knowledge, there are no IoT platforms based on multi-level exogenous interactions. As this is the most promising schema for IoT, we hope to see its realization in the coming years. In fact, we are currently working on the development of such a platform.

REFERENCES


