



Universität Stuttgart

A model-based approach for data processing in IoT environments

Von der Fakultät für Informatik, Elektrotechnik und Informationstechnik
der Universität Stuttgart zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von
Ana Cristina Franco da Silva
aus Manaus / Brasilien

Hauptberichter: Prof. Dr. -Ing. habil. Bernhard Mitschang

Mitberichter: Prof. Dr. Marco Aiello

Tag der mündlichen Prüfung: 04.11.2020

Institut für Parallele und Verteilte Systeme

2020

CONTENTS

1 Introduction	17
1.1 Motivation	18
1.2 Research questions and goals	21
1.3 Contributions summary	24
1.4 Structure of this thesis	26
2 Background	29
2.1 Internet of Things	29
2.2 Data stream processing and complex event processing	30
2.3 Operator placement problem	32
2.4 TOSCA	32
3 Thesis overview	37
3.1 Contributions	37
3.2 Methodical approach	40
3.3 Overall architecture	43
4 Modeling of IoT environments and data stream processing	47
4.1 Modeling of IoT environments	49
4.1.1 IoTEM definition	52

4.1.2 IoT object and connection capabilities	54
4.1.3 Architecture component and implementation – IoTEM modeler and manager	58
4.1.4 Related work	60
4.2 Modeling of data stream processing	62
4.2.1 DSPM definition	62
4.2.2 Processing operators	64
4.2.3 Architecture component and implementation – DSPM modeler and manager	67
4.2.4 Related work	69
5 Mapping of DSPMs onto IoTEMs	71
5.1 Automatic mapping approach	72
5.1.1 Matching algorithm – greedy variant	74
5.1.2 Matching algorithm – backtracking variant	77
5.1.3 Case scenario: monitoring of mold levels in smart buildings	81
5.2 Manual mapping approach	84
5.3 Architecture component and implementation – IoTEM and DSPM mapper	88
5.4 Related work	90
6 Deployment of operators onto IoT environments	93
6.1 Automatic deployment approach	94
6.1.1 Deployment states of an operator	94
6.1.2 TOSCA-based operator deployment	95
6.2 Semi-automatic deployment approach	97
6.3 Topic Description Language for the IoT	98
6.4 Architecture component and implementation – Deployment manager	105
6.5 Related work	107
7 Monitoring of deployed DSPMs	111
7.1 Modeling of disturbance recognition	112

7.2 Executing disturbance recognition	117
7.2.1 Customization and provisioning of CEP engines	121
7.2.2 Disturbance classes	126
7.3 Architecture component and implementation – Disturbance recognizer	129
7.4 Related work	130
8 Evaluation	133
8.1 Integration architecture and prototype	134
8.2 MBP overview	142
8.2.1 Modeling IoT environments	143
8.2.2 Deploying operators onto IoT environments	143
8.2.3 Monitoring IoT environments	144
8.2.4 Demonstration: smart office	144
8.3 Further considerations	148
9 Conclusion and future work	153
9.1 Summary	154
9.2 Future work	157
Bibliography	165
List of Figures	187
List of Tables	189
List of Definitions	193

ACRONYMS

- API** Application Programming Interface
- BPEL** Business Process Execution Language
- BPMN** Business Model and Notation
- CEP** Complex Event Processing
- CSAR** Cloud Service Archive
- dDSPM** deployed Data Stream Processing Model
- DBMS** Database Management System
- DSMS** Data Stream Management System
- DSP** Data Stream Processing
- DSPM** Data Stream Processing Model
- ETL** Extract, Transform, and Load
- GPIO** General-purpose Input/Output
- HTTP** Hypertext Transfer Protocol

HVAC	Heating, Ventilation and Air Conditioning
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
IoTEM	IoT Environment Model
IT	Information Technology
JAR	Java Archive
JSON	JavaScript Object Notation
M2M	Machine-to-Machine
MBP	Multi-purpose Binding and Provisioning Platform
MQTT	Message Queuing Telemetry Transport
OASIS	Organization for the Advancement of Structured Information Standards
QoS	Quality of Service
RAM	Random Access Memory
REST	Representational State Transfer
RFID	Radio Frequency Identification
RMP	Resource Management Platform
SOA	Service-oriented Architecture
SSH	Secure Shell
SQL	Structured Query Language
TDLIoT	Topic Description Language for the Internet of Things
TOSCA	Topology and Orchestration Specification for Cloud Applications

UDDI	Universal Description, Discovery and Integration
UI	User Interface
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
XML	Extensible Markup Language
WAR	Web Archive
WSN	Wireless Sensor Networks

ZUSAMMENFASSUNG

Die heutigen Fortschritte in den Bereichen Sensor-Technologie, Netzwerke und Datenverarbeitung haben die Vision des Internet der Dinge mehr und mehr zu einer Realität im alltäglichen Leben gemacht. Dabei ermöglicht das Internet der Dinge die Entwicklung von anspruchsvollen Anwendungen für IoT-Umgebungen, wie Smart Cities, Smart Homes oder Smart Factories. Durch kontinuierliche Sensormessungen sowie hochfrequentem Datenaustausch zwischen sogenannten IoT-Objekten (z.B. IoT-Geräte), nehmen die Daten in IoT-Umgebungen die Form von Datenströmen an. Mit dieser immer größer werdenden Datenmenge, die kontinuierlich als Strom verarbeitet werden muss, treten jedoch einige Herausforderungen auf, die für ein effizientes Verarbeiten von IoT-Daten gelöst werden müssen. Beispielsweise stellt sich die Frage wie IoT-Daten verarbeitet werden können damit einerseits relevante Informationen gewonnen werden können und andererseits die Reaktivität der IoT-Applikationen nicht beeinträchtigt wird. Des Weiteren müssen verschiedene funktionale und nicht-funktionale Anforderungen der IoT-Applikationen durch die Datenverarbeitung erfüllt werden. In dieser Doktorarbeit wird ein neuer holistischer Ansatz vorgestellt, um strombasiere Daten durch IoT-Anwendungen zu verarbeiten. Der Fokus liegt dabei auf der effizienten Platzierung von Datenverarbeitungsoperatoren der datenstrombasierten Anwendungen auf heterogene, verteilte und dynamische

IoT-Umgebungen. Im Gegensatz zu bestehenden Ansätzen im Bereich der Platzierung von Operatoren werden in dem in dieser Arbeit vorgestellten Ansatz auch zusätzliche Anforderungen beachtet, die sich speziell auf die Eigenschaften des IoT beziehen. Des Weiteren werden auch nicht-funktionale und nutzerspezifische Anforderungen beachtet. Diese Doktorarbeit stützt sich auf verschiedene Informationsmodelle und Techniken, um Operatoren zu platzieren, so dass der gesamte Lebenszyklus von IoT-Umgebungen und datenstrombasierten Anwendungen einfach verwaltet werden kann. IoT-Umgebungen und ihre Fähigkeiten zur Datenverarbeitung werden durch sogenannte IoT Environment Models beschrieben (IoTEM). Analog wird die Geschäftslogik der IoT-Applikationen sowie deren Anforderungen durch ein Informationsmodell, genannt Data Stream Processing Model (DSPM), beschrieben. Basierend auf diesen Informationsmodellen bestimmen Algorithmen eine bestmögliche Platzierung von Operatoren auf die IoT-Objekte der IoT-Umgebung, so dass die vorher definierten Anforderungen der Anwendungen und Fähigkeiten der IoT-Umgebung zusammenpassen. Dabei ist es bei diesem Ansatz das Hauptziel die IoT-Daten so nah wie möglich an den Datenquellen zu verarbeiten, so dass Cloud-Infrastrukturen nur im Falle von Ressourcenmangel der IoT-Umgebung zum Einsatz kommen. Die simultane Ausführung der Datenverarbeitung in der IoT-Umgebung und in der Cloud wird allgemein als Fog-Computing bezeichnet. Durch die Konzepte dieser Doktorarbeit kann die Datenverarbeitung von IoT-Applikationen auf spezifische Szenarien zugeschnitten werden, wobei die charakteristischen Anforderungen der Domänen sowie der Anwender der IoT-Applikation berücksichtigt werden. Sobald eine mögliche Platzierung gefunden wurde, werden die Operatoren auf die zugehörigen IoT-Objekte installiert. Hierfür können etablierte Standards wie TOSCA zum Einsatz kommen. Nach der Installation der Operatoren ist die IoT-Anwendung lauffähig. Während der Laufzeit der IoT-Anwendung wird diese kontinuierlich überwacht, um mögliche Störungen zu bemerken, die während der Datenverarbeitung auftreten. Die Ansätze dieser Doktorarbeit werden unterstützt durch die Multi-Purpose Binding and Provisioning Platform, eine Open-Source IoT-Plattform, die als Proof-of-Concept für die Konzepte dieser Doktorarbeit implementiert wurde.

ABSTRACT

The recent advances in several areas, including sensor technologies, networking, and data processing, have enabled the Internet of Things (IoT) vision to become more and more a reality every day. As a consequence of these advances, the IoT of today allows the development of sophisticated applications for IoT environments, such as smart cities, smart homes, or smart factories. Due to continuous sensor measurements and frequent data exchange among so-called IoT objects, the data generated within an IoT environment incorporate the form of data streams. With this increasing amount of data to be continuously processed, several challenges arise while aiming at an efficient processing of IoT data. For instance, how IoT data processing can be realized, so that meaningful information can be derived without affecting the reactivity of IoT applications. Furthermore, how different functional, non-functional, and user-defined requirements of IoT applications can be satisfied by the IoT data processing. In this PhD thesis, a new holistic approach for processing data stream-based applications within IoT environments is presented. Its focus lies on efficient placement of operators of data stream applications onto heterogeneous, distributed, dynamic IoT environments. In contrast to state-of-the-art operator placement, this approach takes into consideration additional requirements introduced by the peculiar characteristics of the Internet of Things. Furthermore, non-

functional and user-defined requirements are also taken into consideration. This PhD thesis is supported by different informational models and operator placement techniques, so that the entire life cycle of IoT environments and data stream-based applications can be easily managed. IoT environments and their processing capabilities are described by IoT environment models (IoTEM). Likewise, the business logic of IoT applications and their requirements are defined by data stream processing models (DSPM). Based on these informational models, several algorithms determine feasible placements of processing operators onto IoT objects of IoT environments, so that the aforementioned requirements and capabilities are matched. In this approach, one of the main goals is to process IoT data as near to data sources as possible, so that cloud infrastructures are employed only in cases where IoT environments do not offer sufficient processing resources for the IoT application. The execution of data processing on both IoT environments and cloud infrastructures is commonly known as fog computing. Through the approach of this PhD thesis, data processing of IoT applications can be tailored to particular use cases, supporting the specific requirements of the domains, and furthermore, of IoT application users. Once feasible placements are determined, processing operators are then deployed onto corresponding IoT objects using standards, such as TOSCA, and the IoT application is considered up and running. Finally, the IoT environment is continuously monitored in order to recognize and react to disturbances affecting the data processing of deployed IoT applications. The approach of this PhD thesis is supported by the Multi-purpose Binding and Provisioning Platform (MBP), an open-source IoT platform, which has been developed as a proof-of-concept of the contributions of this PhD thesis.