

Self-assemble-featured Internet of Things

Frédéric Lemoine, Tatiana Aubonnet, Noëmie Simoni

▶ To cite this version:

Frédéric Lemoine, Tatiana Aubonnet, Noëmie Simoni. Self-assemble-featured Internet of Things. Future Generation Computer Systems, 2020, 10.1016/j.future.2020.05.012. hal-02607933

HAL Id: hal-02607933

https://hal.science/hal-02607933

Submitted on 3 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Self-Assemble-Featured Internet of Things

Frédéric Lemoine^{a,*}, Tatiana Aubonnet^{a,b}, Noëmie Simoni^b

^a Conservatoire National des Arts et Métiers (CNAM), CEDRIC, 292 rue Saint-Martin,
 Paris, France.
 ^b Télécom-Paris, 19 Place Marguerite Perey, Palaiseau, France.

Abstract

The Internet of Things supports various industrial applications. The cooperation and coordination of smart things are a promising strategy for satisfying requirements that are beyond the capacity of a single smart thing. One of the major challenges for today's software engineering is the management of large and complex computing systems characterized by a high degree of physical distribution. Examples of such systems arise in many application domains. The number of connected devices grows from billions to hundreds of billions, so a maximum of automatisms must be integrated in IoT architectures in order to control and manage them. Software architects migrate to service oriented architecture and applications are now being constructed as service compositions. Since each IoT device includes one or more microservices, the increasing number of devices around the user makes them difficult to assemble in order to achieve a common goal. In this paper, we propose a self-assembling solution based on selfcontrolled service components taking into account non-functional requirements concerning the offered quality of services and the structuration of the resulting assembly. Its aim is to build and maintain an assembly of services (taking into account arrival of new peers or failure of existing ones) that, besides functional requirements, also fulfils global quality-of-service and structural requirements.

Keywords: Quality of service, Service component, Service-oriented achitecture, Self-assembly, Service composition, Internet of Things

Preprint submitted to Journal of LATEX Templates

March 28, 2020

^{*}Corresponding author
Email address: frederic.lemoine@cnam.fr (Frédéric Lemoine)

1. Introduction

25

The Internet of Things (IoT) vision defines a global network of interconnected services and smart objects that support human activities in everyday life using their sensing, computing and communication capabilities. IoT will enable global connectivity between devices, and people. The IoT links physical activities and real life with the virtual world [1].

IoT will have a crucial role to play in the future, when human decisionmaking is necessary, especially in critical and urgent situations in which quality of service must be controlled. Some IoT-related critical systems are listed below:

- Energy and Utilities: Increasing number of projects covering smart grid
 programmes, smart cities, and smart metering looking at ways to improve
 network efficiency and usability. Ensuring critical service requests (monitoring and control, power production, and water pressure monitoring)
 from trusted sources.
- Automotive: Sensors in vehicles provide even more data on things like environmental conditions, tyre pressure, engine performance and environmental conditions. These integrated, safe, and robust embedded systems will in a near future lead to self-driving cars.
 - Railway (improvement of the management of rail networks): Widespread use of automation and sensor technologies, on tracks or trains, provides real-time data on resources in the network. This information can be used to predict when resources need repairing or replacing, reducing so the amount of time that rail resources are unavailable.
 - *Healthcare*: Connected medical devices and applications are already creating an Internet of Medical Things which is contributing to better health monitoring and preventive care. From improving patients' experience through to connecting medical devices to physicians' smartphones, the

IoT is transforming healthcare. Wearable devices and mobile medical applications that allow patients to measure and send their own biometric data, supplement patient treatment through remote monitoring and communication. The efficiency of healthcare facilities is also enhanced by improved tracking of medical devices, personnel and patients (Indoor Location Tracking).

 Aerospace (safety-critical applications): Sensors on aircraft create huge amounts of data for each flight, passenger management systems control huge amounts of complex personal data and air traffic control systems constantly monitor and manage plane flights, sharing data on a global scale.

For any of these applications, failures might lead to serious injury and death (including on a large scale). As the number of objects, including sensors and actuators, increases, IoT becomes more and more complex and should be controlled, especially in these critical domains. Moreover, interactions between critical and non-critical systems in the IoT create problems for design engineers. Consider an application to provide continuous tracking of a container of critical medical supplies (or transplant organs) from a first hospital to a second one far away. The container starts in a first hospital, then it is transferred to an ambulance, which travels by road to an airport. The container is then transferred into a plane, and, after landing, to a train, then to another ambulance, and finally arrives at a second hospital. Many different IoT ecosystems were crossed, each complying with different standards and through a variety of communications links. Building this application in a way that patient lives are not put at risk, will be particularly challenging.

Applications are now being constructed as micro-service compositions [2, 3, 4] integrating more and more functionalities. Everyone can build a composition of services according to its needs. Software architects migrate to service-oriented architecture. We are in an age of services and micro-services are at the heart of architecture. If each IoT device includes one or more microservices, the

increasing number of devices around the user makes them difficult to assemble in order to achieve a common goal.

The number of connected devices grows from billions to hundreds of billions, a maximum of automatisms must be integrated in the IoT architectures so as to control and manage them.

A lot of IoT ecosystems rely on centralized, brokered communication models. All devices are identified, authenticated and connected through cloud servers that provide huge processing and storage capacities. Connection between devices will have to go exclusively through the internet, even if these devices are close to one another. These IoT ecosystems will not be able to manage a growing number of IoT devices. Cloud servers are a bottleneck that may disrupt the entire network.

A distributed service approach would solve the problem mentioned above, by spreading computational and storage requirements among billions of devices that will form the IoT networks of the future. Computing and storage are already widespread in many devices: from home to cars. Devices now carry as much computing power and connectivity as did the first smartphones. Connectivity and intelligence will be embedded in practically every object around us. Some services should stay close to the object instead of far in the cloud. Many human-machine interactions will be replaced by machine-machine interactions. IoT architectures will be more and more distributed and autonomous, acting in the best interest of the user, putting the user first, and designed for user-centric experiences.

Furthermore, when a composition is defective from end to end, it is difficult to know which composite service is responsible. A fixed control centre, sufficiently powerful and capable of controlling the state of the entire system and manoeuvring its behaviour, is generally unreasonable. A maximum of automatisms should be placed on the objects themselves, thus unloading the cloud from this work. Service behavior should be controlled by the object itself rather than by the cloud. So we would gain efficiency and reaction time.

The Internet of Things supports various industrial applications. The coop-

eration and coordination of smart things are a promising strategy for satisfying requirements that are beyond the capacity of any single smart thing. One of the major challenges for today's software engineering is the management of large and complex computing systems characterized by a high degree of physical distribution. Examples of such systems arise in many application domains, such as sensor networks, intelligent transportation systems, or ambient intelligence.

How can be helped the user and architect in building his application based on distributed services?

Since device failures and dynamics in the environment may imply the unavailability of components and devices at any time, adapting and maintaining such a composition is challenging.

How the QoS of such complex systems can be efficiently controlled?

100

Our motivation is to propose a self-assemble-featured architecture for IoT in a distributed service-oriented environment. Its aim is to automatically build and maintain an assembly of services that besides functional requirements, fulfill global QoS and structural requirements. The concept of self-assembling covers multiple properties that can be regarded as the basic requirements for enabling the desired behaviors in such systems.

Analysis of existing solutions to meet the needs of IoT/cloud distributed systems, has led us to look at the evolution from component models to service models and their dynamic management. Component models provide a structured programming paradigm, and ensure a very good program re-usability. We focus here on generic and hierarchical component models because they make the design of large-scale systems easier. On one side, QoS based approaches and adaptive contracts for services already exist (video services for example) [5, 6] but they lack the compositional design and the management featured by components. On the other side, dynamic management in time-critical IoT middleware is very compositional and expressive, but the support for an end to end QoS in the specific context of service-oriented components is weak. Time-critical IoT middlewares provide the QoS links but do not take into account the QoS nodes [7, 8, 9, 10]. To complete the modeling of the IoT ecosystem, we refer

to the node/links model allowing us to have an architectural abstraction for dynamicity and a flexible engineering for a personalized user session.

Our proposal, presented in this paper, consists in using strongly structured components to provide a service oriented platform that eases the design and execution of applications with end to end QoS guarantees. The dynamic management will be performed in a modular way. Each service (component) is responsible for its own adaptation and its own QoS while interacting with external services.

The extension of software services with autonomic capabilities has been suggested as a possible way to deal with this challenge. One of the major challenges in IoT is the coordination that is necessary to align the behavior of different objects. Decentralization implies a style of coordination in which the objects cooperate as peers with respect to one another. Self-assembly is an emergent property of decentralized systems. It is the preferred mechanism for growth in the natural world, on scales ranging from molecular to macro-scales. It involves the assembling of components that are governed by a set of local interaction rules and lead to the formation of a global structure. Understanding and applying this emergent property continues to be an important subject in the natural sciences, as well as in engineering and computer science.

The autonomic computing paradigm enables a system to automatically act in response to variations of operating conditions guaranteeing short reaction times and minimal human intervention. The approach we present in this paper is structured on a fully distributed implementation of several autonomic capabilities. As an example, a possible scenario, to which we propose to answer, would be the following: A user buys an IoT device and brings it home. He connects it to his home network. The device is thus aware of its environment and detects the surrounding services and their features. These services are able to assemble themselves to achieve a common goal according to quality of service (QoS) requirements and thus respond to a user need (Control the home temperature, close the window, call emergency teams, etc.). If some IoT devices/services are appearing or disappearing, the system is able to reassemble itself in order to do

its work and to continue to respect the QoS requirements.

To address the problem of complexity, collaboration, and QoS management of the growing number of IoT devices, we propose in this paper to use a distributed service architecture based on our service composition entity. We extend it with autonomic capabilities, and we introduce a way to automatically assemble a set of IoT device microservices in order to achieve a common goal. We will show that our solution is able to dynamically perform the composition and thus react to changes in its execution context. We propose a self-assembling solution based on self-controlled service components taking into account non-functional requirements concerning the offered QoS and the structuration of the resulting assembly. Its aim is to build and maintain an assembly of services (taking into account arrival of new peers or failure of existing ones) that, besides functional requirements, also fulfills global quality-of-service and structural requirements. It is able to drive the system and manage the selection, among the set of functionally feasible assemblies, of an assembly that fulfills these requirements.

After enumerating in Section 2, the properties of Self-Assemble-Featured IoT, presenting our service composition entity on which we base our approach (Section 3) and reviewing related works (Section 4), we present our self-assembly algorithm (Section 5). Section 6 presents a case study. Finally, a conclusion (Section 7) ends the article.

2. Properties of Self-Assemble-Featured IoT

The concept of self-assembling covers multiple properties that can be regarded as the basic requirements for enabling the desired behaviors in such systems. A qualified implementation of a self-assembling scheme should meet the following requirements that will be part of the future Internet of Things: (i) distributed infrastructure, (ii) autonomous, and (iii) efficient collaboration.

1. **Distributed infrastructure.** For the IoT, which is characterized by highly evolved heterogeneity, a fixed controlling center capable and powerful enough to control the status of the whole system and manoeuvre its

behaviors is generally unreasonable. So, a self-assembling scheme based on the distributed infrastructure may be much more suitable for the IoT.

180

185

190

195

200

205

210

- 2. Autonomous. Each node in the IoT has to be implemented with a certain degree of autonomy. Such nodes can therefore collect information about their respective environments and make decisions for their own behaviors. In this case, the data (including instructions and information) transmitted amongst nodes in a fully distributed form, is restricted into a limited area, which decreases heterogeneity of the whole network. Such distributed architecture also enables the self-assembling scheme to be scalable.
 Our approach is organized with respect to several autonomic capabilities
 - [11]. We recall the definitions for our approach in Table 1. Services have to be aware of their environment and therefore of their surrounding services. They need intelligent service discovery process. Services have to broadcast their features (function, offered/nominal QoS, threshold value from which the business component stops responding) to the others (self-defining). A method to implement this service discovery process is proposed in Section 5.1. We recall, in Section 3, that our component fulfills self-control/self-diagnosis, and self-monitoring requirements. In Section 5.3, we present a self-assembling algorithm allowing self-controlled services to be assembled to compose an application (self-assembled). This algorithm test and evaluate scenarios to choose the more suitable for using taking account QoS constraints (self-simulation). We show in Section 5.3.4 that our self-assembling algorithm, with its abilities to reassemble components at any time, supports the self-adjusting, self-adapting,
- 3. Efficient collaboration based on ubiquitous data exchanging and sharing. We suppose in this paper that our components can exchange and share data in an efficient way. This point is out of the scope of this paper. Several approaches have been proposed in literature like "Blockchain" which brings security. A blockchain [12, 13], or chain of blocks, is the implementation of a data storage and transmission technology without a

and self-healing (self-repair) benefits.

control organ. Technically, it is a distributed database whose information, sent by users, is checked and grouped in blocks, linked and secured using cryptography. Our algorithm requires that each node maintains and disseminates global state information consisting of the whole assembly of services and computed link QoS. The amount of data to be shared is small. This can be done with use of blockchain technology but other simplest solutions may be sufficient.

The following section presents our service composition entity on which we base our approach.

220 3. Previous Works

215

235

At the service era, the service is the center of architecture, to profit by all the benefits expected from this concept, we have proposed in [14] a component called "Self-Controlled service Component (SCC)", the description of which we recall. A component encapsulates a micro-service and only one. These services can be of different sizes and natures: real-time image analysis service (face recognition, character recognition, barcode recognition) in the case of the Internet of Things, algorithms (sorting, encryption), image capture, etc. To describe the behavior of our components and allow an homogeneous quality of service management, a generic QoS model [15] is defined. Four criteria are proposed to describe the QoS: availability, reliability, time, and capacity.

- Availability represents the accessibility rate of the service component.
- A system is said to be reliable when its behavior, over a given duration, is in conformity with that expected.
- **Time** represents the time required for request processing.
- Capacity represents the maximum load that the service component can handle.

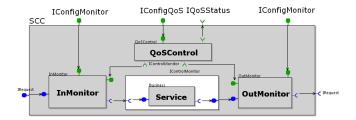


Figure 1: Self-Controlled service Component (SCC)

This proved to be useful and sufficient in all the practical cases we studied. To increase the structural decomposition and the reuse of non-functional QoS components, its internal functions have been separated and an architecture that separates the monitoring and QoS functions of the remaining functions called "control" has been proposed. This model has been specified in the OpenCloudware project [16] to address the behavioral aspects through QoS.

The membrane of our SCC includes (Figure 1):

245

250

255

- Input monitoring (InMonitor) and output monitoring (OutMonitor) components. They play an interceptor role. Incoming service requests are intercepted and transmitted (unchanged) to the functional component via the corresponding internal interfaces. The OutMonitor intercepts outgoing service requests. They provide measurement information on the flow that they intercept.
- A QoS component (QoSControl), associated with the business component.
 - A non-functional interface (client) for QoS control (IQoSStatus), by which it will send the information in case of violation of QoS contracts, i.e. "InContract" notifications when the behaviour is in conformity with the contract or "OutContract" otherwise.
- A non-functional interface (server) of configuration (IConfigQoS, IConfig-Monitor), whose role is to receive component configuration commands.

The QoSControl component checks the current behaviour of the resource and its conformity with the contract. For this, it regularly requests to the monitors (In-

Monitor and OutMonitor) the parameter values. It compares each current value to the corresponding threshold value not to be exceeded. It sends an OutContract notification if the current value is less (or more) than the threshold value. In this case the dynamic management consists in replacing immediately the failing component by a ubiquitous service fulfilling the requirements. Otherwise, it sends an InContract notification. We obtain an SCC component, self-monitored and self-controlled. The sub-components of the membrane (monitors and QoS) are activated in order to perform monitoring of the quality of service and to notify its degradation. This component has been designed so that it can be composed with others in order to form a self-controlled composition. An application is a service composition. The service composition includes different SCC called during the user session.

The SCC component has proved its value in the Cloud [16, 15, 17] and in the IoT [18]. The SCC architecture has many advantages: Our SCC component allows self-control inside by signaling dysfunction (out of contract) and automatic reacting outside. Since we are as close as possible to the functional component, the analysis is faster, more relevant, and reaction times are minimized. The analysis is done on the site itself. Only its result is sent so that the volume of data exchanged and thus the communication resources needed are extremely small. The code is simplified and hence requires less computing resources. Monitoring and controlling components are generic so they are independent of the functional component and may be present at all levels of architecture. They are not intrusive because they are external to the functional component. They are inside the service component membrane and operate in parallel with the functional component. They have no effects on the second. The QoS of each component (hardware or software) is measured, allowing better diagnostic of various dysfunctions whereas most existing tools monitor network traffic or central processing unit usage when they should monitor the functional component performance. At each addition/removal of a functional component, a monitoring and controlling component is therefore added/removed (Scalability, Elasticity).

The service is defined by an offered/nominal QoS and a threshold value.

These values are used by our algorithm (Section 5.3). Offered/nominal QoS is the QoS (processing time for example) provided by the service under normal operating conditions. It is a value given by the service provider in a service level agreement or in its provided services catalogue. It is computed under resource conditions of the underlying level. The threshold value is the number of simultaneous requests not to exceed in order to comply with the nominal preceding defined QoS. These values are computed with a calibration procedure [19].

The following section surveys the related works.

4. Related Works

300

In literature, the term "self-assembling" mainly concerns robotic, mechatronic, chemistry and materials science domains. There are few articles in the field of service-oriented architecture. Assembly is often done manually by a software architect or sometimes computed automatically.

Methodology for middleware and technology selection is detailed as follows. We surveyed 50 middlewares and technologies and made an initial selection of 25 of the most advanced approaches that deal with service discovery and assembly with a preference for the self-assembly property. Assembly can be done in different ways: (i) manually with a visual editor for example, (ii) automatically with an algorithm or predefined scripts (only once in general at the design phase or at demand), and (iii) dynamically (at any time with respect to context changes). Assembly can be done in a centralized (by a central entity) or decentralized way (by the devices themselves). We have chosen to take into account the control of QoS too (QoS aware). The choice has been restricted to a selection of 16 representative solutions: Tile-based approach, MACODO, Schuhmann et al., FlashMob, MOSDEN, UbiROAD, Calvin, Paraimpu, SENSEI, Node-RED, CHOReOS, SenseWrap, SOCRADES, Ubiware, extended GSN, KASOM (Table 2).

Article [20] presents an approach allowing to develop a service-oriented sys-

tem, based on a model called service tiles, by building an assembly of service components that accomplishes a given goal. The assembly is computed automatically: with the specification of a subset of the whole system, a few constraints, and the goals that the application should fulfill.

MACODO [21, 22] uses a partially distributed architecture based on a masterslave scheme. The master has a complete knowledge of the assembly state and controls the dynamics in a centralized way. Masters of different assemblies can cooperate to achieve a given goal.

Article [23] presents algorithms for homogeneous and heterogeneous environments whose goal is to choose the most efficient assembly method for a given environment while minimizing the assembly time. The organization latency is reduced by eaching and reusing partial application assemblies.

FlashMob [24] is based on dynamic service assembly and requires a backtracking phase to explore alternative solutions in case the assembly fails and has no global QoS goal. Its self-assembly procedure is decentralized. Global state information dealing with the whole assembly is disseminated among the services.

MOSDEN [25] supports a sensing as a service model [26]. It is an IoT middleware for resource constrained computational mobile devices. MOSDEN can collect data from multiple different sensors and process them together. It is fully compatible with Global Sensor Network Middleware that runs on the cloud. Sensor discovery and service composition are not automated.

UbiROAD [27] is a specialized platform for smart traffic environments. Semantic technologies are the basis for the discovery of heterogeneous resources and data integration across multiple domains. They are used both for descriptive specification of services delivered by the resources and for prescriptive specification of the expected behavior of resources as well as of the integrated system. UbiROAD guarantees a high level of safety while offering personalization, dynamic behaviour, customization, and autonomy of services and ensures context-awareness, and adaptive/reconfigurable composition.

Calvin [28, 29] is an open-source IoT middleware from Ericsson. IoT ap-

plications building is based on actors which are reusable software components that may represent a device, a computation, or a service. It comprises both a development framework for IoT application developers, and a runtime environment that handles the running application. Compositions are made by writing scripts called CalvinScript. Basically, an application consists of actor instances and connections between the ports of the actors, forming a data flow graph. To allow the reuse of scripts, it is also possible to define components. An application script can also contain deployment rules.

Paraimpu [30] is a social aware IoT middleware that allows consumers to add, use, share, and interconnect RESTful IoT services whether physical or virtual. Things are mapped to abstract concepts of sensors or actuators. Providing connection abstraction between things, allows users to compose simple IoT applications via Javascript.

The SENSEI [31] middleware includes context services, context model, actuation tasks, and dynamic service composition of both primitive and advanced services.

Node-RED [32] is an open-source IoT middleware platform from IBM. Its key advantage is a visual tool that simplifies the composition of IoT devices especially if the node for the IoT device is already developed and published by others.

370

CHOReOS [33] composes distributed services by considering a global specification, called Choreography, of the interactions between the participant services. It enables large scale compositions or choreographies of QoS-aware, and heterogeneous services in IoT. It includes i) extensible service discovery to manage protocols and processes for discovery of services and things, and ii) executable service composition to coordinate the composition of services and things. Semantic thing-based service compositions are automatically executed, with no involvement from endusers.

The SenseWrap [34] middleware combines the Zeroconf [35] protocols with hardware abstraction using virtual sensors. A virtual sensor provides transparent discovery of resources, mainly sensors, through the use of Zeroconf protocol.

Applications can thus use it in order to discover sensor-hosted services. It also provides a standardized communication interface to hide the sensor-specific details. This interface depends on sensor modeling and custom wrappers. Virtualization is applied only to sensors, not to actuators or computing resources which makes it unsuitable for ultra-large scale IoT environments and heterogeneous networks.

The SOCRADES [36] middleware architecture consists of a layer for application services and a layer for device services (service discovery, monitor, device and service lifecycle management). By using devices profile for web services (DPWS), it abstracts physical things as services. Its cross-layer service catalogue supports service composition, but it may not be fully dynamic, since composition relies on predefined building blocks.

Ubiware [37] is able to create flexible, autonomous, and complex industrial systems. It supports composition, invocations, monitoring, automatic resource discovery, and execution.

Global Sensor Networks (GSN) [38] is a service-based IoT middleware that proposes virtual sensor abstraction and aims to provide a uniform platform for integration and deployment of heterogeneous IoT devices. Users and developers specify XML-based deployment descriptors to deploy a sensor. GSN does not support composition of multivendor devices via the XML descriptor. The extended GSN [39] provides a limited composition capability.

KASOM [40] is a knowledge-aware and service-oriented middleware. Its aim is to offer advanced and enriched pervasive services to every user connected to the Internet. KASOM implements mechanisms and protocols that allow managing the knowledge generated in pervasive embedded networks in order to expose it to Internet users in a readable way. Its architecture includes communication services (resource monitor), and knowledge management services (context resources and service composition rules). This middleware offers discovery, registration, orchestration, and composition of services. It has proven its qualities in terms of reliability, efficiency, and response time in the field of health management. However, because of predefined service composition rules

provided by in-network agents, it does not provide dynamic service composition in IoT infrastructure.

An application (service composition) is usually monolithic. It offers features created to meet a set of needs even if the user does not use them all. A dynamic assembly would create a customized application according to the user's current need. It would integrate a service only if necessary and would thus gain in performance. The increasing number of IoT devices, however, implies the problem of their management to achieve specific common objectives. A fixed controlling center capable and powerful enough to control the status of the whole system and manoeuvre its behaviors is in general unreasonable. The assembly has to be decentralized. The question is then how to coordinate local services among IoT devices in order to accomplish a goal with global QoS evaluation?

Table 2 summarizes the analysis and categorization of these works.

425

None of the afore mentioned approaches can be simultaneously dynamic, decentralized, covering non-functional and structural constraints and integrates self-control mechanisms as defined in this paper.

Table 1: Autonomic properties covered by our approach.

| Properties | Definitions | See Section | | |
|-----------------|---|-------------|--|--|
| | A system's ability to describe itself to other | | | |
| Self-defining | systems. An autonomic system may need to | 5.1 | | |
| | understand and interpret other systems' | 0.1 | | |
| | descriptions. | | | |
| C-1f1 | A system's ability to analyse itself in order | | | |
| Self-control | to identify existing problems or to anticipate | 3 | | |
| Self-diagnosis | potential issues. | | | |
| Self-monitoring | A system's ability to retrieve information on its | | | |
| | internal state and behaviour, whether globally, | 3 | | |
| | or for any of its constituent elements. | | | |
| Self-assembled | A system's property of being automatically formed | | | |
| | via the distributed assembly of multiple | 5.3 | | |
| | independent elements, which become the system's | | | |
| | constituent elements. | | | |
| Self-simulation | A system's ability to test and evaluate scenarios | | | |
| | without affecting the executing system (e.g. it | | | |
| | should not impact provided services). This allows | 5.3.4 | | |
| | replying to 'what would happen if' questions and | | | |
| | hence facilitate the selection of self-adjusting actions. | | | |
| Self-adjusting | A system's ability to modify itself during runtime, | 5.3.4 | | |
| sen-adjusting | including modifications to its internal structure. | 5.3.4 | | |
| Self-adapting | A system's ability to modify itself (self-adjust) | | | |
| | in reaction to changes in its execution context or | 5.3.4 | | |
| | external environment, in order to continue to meet | | | |
| | its business objectives despite such changes. | | | |
| G 161 3: | A system's ability to recover from the failure of | | | |
| Self-healing | any of its constituent elements (service) or to | 5.3.4 | | |
| Self-repair | predict and prevent the occurrence of such failures. | | | |

| | Table 2: Analysis of w | Table 2: Analysis of works related to assembly of services | services. | |
|--------------------------|--|--|-------------------------------|------------------------------|
| | Assembly | Service discovery | QoS control | Decentralized / centralized |
| Tile-based approach [20] | Automatic | NI | No | IN |
| | | | | Partially distributed. A |
| MACODO $[21, 22]$ | Dynamic | NI | No | master controls the dynamics |
| | | | | in a centralized way. |
| Schuhmann et al. [23] | Dynamic | NI | No | IN |
| FlashMob [24] | Dynamic | NI | No global QoS goal. | Decentralized |
| MOSDEN [25] | Manual. Not automated | Not automated | NI | Cloud centric |
| ThiBOAD [27] | Adantive/reconfigurable | Yes with descriptive | Prescriptive specification of | N |
| | | specifications. | the expected behaviour. | 1 |
| Calvin $[28, 29]$ | Made by writing scripts | NI | No | IN |
| Paraimpu [30] | Manual. Via Javascript | NI | No | Cloud centric |
| SENSEI [31] | Dynamic | NI | No | IN |
| Node-RED $[32]$ | Manual. Via a visual tool | NI | No | Cloud centric |
| CHOReOS [33] | Computed automatically | Discovery of services and things. | Yes | Decentralized |
| SenseWrap [34] | Manual. Not automated | Yes | $N_{\rm O}$ | Decentralized |
| SOCRADES [36] | Not fully dynamic, predefined building blocks | Yes | No | N |
| Ubiware [37] | Not fully dynamic | Yes | No | Decentralized |
| extended GSN [39] | Limited composition capability | NI | No | IN |
| KASOM [40] | Not dynamic, predefined | Ves | Reliability, efficiency, | N |
| | service composition rules | | and response time. | • |

No information (NI)

In the following section, we present our self-assembly algorithm.

5. Self-assembly Algorithm

In Section 5.1, we describe a method for a service to describe its capabilities especially his QoS nominal/offered to others in order to achieve a requested service composition. In Section 5.2, we show that, using our component, we are able to compute an estimation of the link QoS. In Section 5.3, we define our model, introduce the terminology and notation used in the rest of the paper, present our assembling core algorithm, and enumerate the benefits of our algorithm. Finally, the limitations of the algorithm on a real IoT platform are estimated (Section 5.4).

5.1. Service Self-Definition and Surrounding Awareness

In this section, a method, for a service to describe its capabilities, is proposed. Each service has to broadcast its features (function, offered/nominal QoS, threshold value) to the others. This is proposed as a proof of concept. Our proposition is based on Network Service Discovery (NSD) protocol which we propose to extend. NSD gives to any application access to services that other devices provide on a local network. Devices that support it include webcams, printers, HTTPS servers, and mobile devices. Provided services include web server, file sharing, file transfer, printing, real time streaming, sound server, etc. It's based on the Domain Name System-based Service Discovery (DNS-SD) mechanism [41], which allows the user's app to request services by specifying a type of service and the name of a device instance that provides the desired type of service. DNS-SD is supported both on Android and on other mobile platforms. Presenting DNS-SD is out of the scope of this paper, we only want to underline that it is possible and easy for services to broadcast their features to the other ones. We have chosen DNS-SD to demonstrate that, but other solutions may exist. With NSD, you are able to identify other devices on the local network that support the services on which one wants to connect. This



Figure 2: Self-definition of SCC using NSD on Android

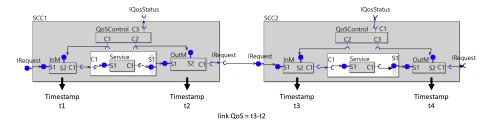


Figure 3: Estimation of the link QoS.

is useful for assembling services as we propose to do so. The surrounding services are in the communication range of the device. From the device's point of view, when a service is out of range, it can no longer be used and disappears from the list of found services. We have implemented, on the Android operating system, an application using NSD to find an image capture service (IP camera) on the near network. Android's NSD APIs simplify the effort required to implement these features. Figure 2 shows the interface. It is divided into two

parts. The top part contains the server logs and two buttons: the first starts the image capture service and the second stops it. The bottom part is for the client who searches for an image capture service. It can start and stop a discovering. When the discovering phase is on, it finds the service and collect his features (nominal/offered QoS, required memory, etc.) in a key/value form and in an eXtensible Markup Language (XML) format form. This is the concretization of the self-defining property defined in Section 2. With NSD, services are aware of their surrounding environment. Services they found are in communication range. In the following section, we show that, thanks to our SCC component, we are able to compute an estimation of the link QoS that will be used by our algorithm.

5.2. Estimation of the Link QoS

475

490

In this section, we describe a method, using our SCC, for calculating the link QoS, that is the response time between services. The estimation of the link QoS is further used in our algorithm (Section 5.3). Our approach allows to compute an estimation of the link QoS by using SCC inside a composition. Indeed, an SCC includes two monitors that intercept requests and responses of the functional part. These are the right places where to put a timestamp as metadata on the intercepted request and response (Figure 3). When the response is provided by the first SCC, the OutMonitor put a timestamp t2 on the response. When the response arrives at the second SCC, the InMonitor puts a timestamp t3 on it. So the link QoS is computed as t3-t2. This is simple, quick, consumes very little computing resources, and allows to estimate the response time between two SCC. This proposition will be used in our algorithm in the next section. Note that the link could also be a full-fledged SCC component. The link QoS represents both a processing time for the seven-layer Open Systems Interconnection (OSI) model of computer networking [42] and a transport time.

In the next section, we present our algorithmic model including our core

5.3. Algorithmic Model

In Section 5.3.1, we define our model, introduce the terminology and notation used in the rest of the paper and present our assembling core algorithm (Section 5.3.2). In Section 5.3.3, we present an example of execution of the assembling core algorithm. The Section 5.3.4 enumerates his benefits.

5.3.1. Model Definition

505

We consider a system containing N distributed services $S=\{S_1,...,S_N\}$, with each service having a type $d \in T=\{T_1,...,T_M\}$. Services are hosted on peer nodes, each node containing one or more services. Nodes can be located anywhere and communicate with one another through a network. Formally, a service S is a tuple (Type, QoS, Threshold), where:

- S.Type \in T denotes the type of the service, its function.
- S.QoS $\in \mathbb{R}$ represents the service QoS nominal value (e.g. response time).
- S.Threshold ∈ N represents the nominal value i.e. the threshold (number of simultaneous requests) not to exceed in order to comply with the nominal preceding defined QoS.

A service assembly A is a graph A = (S,E), where $E \subseteq S \times S$ is the set of bindings. Specifically, a directed edge $(S_i,S_j) \in E$ denotes that S_i is connected to S_j . We allow multiple simultaneous bindings to the same service S by other services. The upperbound for the number of bindings to a service is the threshold value for QoS compliance (Non-Functional constraints). Indeed, in the worst case, services can emit simultaneously without interfering with each other to the same recipient. An application template AT is a tuple (Body, Constraints), where:

AT.Body ⊆ (T × T)^p is a set of p values representing ordered microservice types used to build the desired application. Specifically (t_i ,t_j) ∈ AT.Body denotes that services of type t_i are connected to the services of type t_j.

• AT.Constraints $\in (\mathbb{N}^* \cup \{\bowtie\})^p$ is a set of p values of $\mathbb{N}^* \cup \{\bowtie\}$ representing the number of services that should be connected according to each tuple $(t_i, t_j) \in AT.Body$. $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ is the set of non-zero natural numbers. Specifically: $b_k = (t_i, t_j) \in AT.Body$ and $c_k \in AT.Constraints$ means that all services of type t_i should be connected to c_k services of type t_j . \bowtie means that services of type t_i should be connected to all available services of type t_j . These are **structural constraints**.

520

525

A service assembly A is **AT compliant** if A is compliant with the application template i.e. services are connected according to AT.Body.

A service assembly A is **ATS compliant** (S for structural constraints) if A is AT compliant with respect to AT.Constraints (**structural constraints**).

A service assembly A is **ATSF** compliant (SF for structural and non-functional constraints) if A is ATS compliant and the number of bindings to each service is upper bounded by the threshold value (**Non-Functional constraints**).

For a given template AT, there may be more than one AT compliant assembly.

Given $S_k \in S$ with S_k of type t_k , S_k is a **starting service** if there is no tuple $(t_i,t_j) \in AT.Body$ where $t_j=t_k$.

Given $S_k \in S$ with S_k of type t_k , S_k is an **ending service** if there is no tuple $(t_i,t_j) \in AT.Body$ where $t_i=t_k$.

For an application template AT, in AT.Body, we state that there is only one type of starting service. A second starting type means a second application, so a second template.

For a given AT compliant service assembly A and a starting service $S_i \in S$ of type t_i , we define a subgraph $p_{Si} \subseteq A$, formed by all graphs starting from node S_i .

For a subgraph $g_{Si,j}$ of p_{Si} , we define $F(g_{Si,j})$ as Max (processing time from the starting service S_i to an ending service). Processing time is computed as the sum of transfer times between services (computed link QoS, see algorithm 1) and

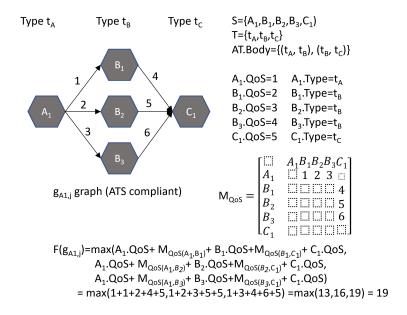


Figure 4: $F(g_{A1,j})$ calculation example.

processing times (nominal S.QoS) of each service composing the path (Figure 4). We choose the worst case, i.e. the maximum of the processing times. We try, in our algorithms, to choose the best organizations that minimize the set of $F(g_{Si,j})$ i.e. the maximum processing time of all subgraphs $g_{Si,j}$.

Since our services are stateless, they do not maintain the interaction state between service invocations, i.e., a client request is served in complete isolation, without relying on information from previous requests. Hence, we assume that the state of the interaction between the client and the service is kept on the user's side, and requests include all information necessary for their processing. Service statelessness enhances (i) decoupling of interacting services (ii) flexibility of the model, since it allows for easily rearranging the assembly at run time and, (iii) scalability. The table 3 summarizes the notation used in this paper. The table also includes additional symbols which will be introduced in the next sections.

5.3.2. Assembling Core Algorithm

The assembling algorithm is structured by three sub-algorithms, here noted 65 1 to 3.

| | Table 3: Symbols Used in this Paper |
|---------------------|--|
| symbols | definitions |
| N | Number of services (peers) |
| S | Set of services $S = \{S_1,, S_N\}$ |
| M | Number of types |
| Т | Set of service types $T = \{t_1,,t_M\}$ |
| A | Service Assembly $A=(S,E)$ |
| S.Type | Service type of S |
| S.Threshold | Number of services maximum bindings |
| \mathbf{p}_{Si} | Subgraph with S_i starting service |
| $\mathrm{g}_{Si,j}$ | Subgraph j of p_{Si} ATS compliant |
| AT.Body | Application body template |
| AT.Constraints | Structural constraints |
| S.QoS | QoS nominal value |
| $F(g_{Si,j})$ | Maximum processing time of the subgraph $g_{Si,j}$ |
| G_{AT} | AT compliant graph |
| \mathcal{M}_{QoS} | Matrix of all computed link QoS (time transfer) |

Algorithm 1 build an **AT compliant** graph and compute all link QoS. We call G_{AT} this graph and $M_{QoS} \in M_{N,N}(\mathbb{R})$ the matrix of all computed link QoS.

```
Algorithm 1 AT compliant
```

```
1: for all S_i starting service of type t_i do
      for all (t_i,t_j) \in AT.Body: do
 3:
        S_i sends a request to all services of type t_i.
      end for
 4:
 5: end for
 6:
 7: if a service S_i service of type t_i, receives a request then
      it computes the QoS of the link between him and the sender.
      for all (t_i, t_j) \in AT.Body: do
 9:
        S_i sends a request to all services of type t_j.
10:
11:
      end for
12: end if
```

We will now take into account AT. Constraints (Algorithm 2). AT. Constraints introduces restrictions in G_{AT} so that the resulting graph will be a subset of G_{AT} . Taking into account AT. Constraints is equivalent to make a choice among available paths. Restriction can be of two types for a path: $k \in N^*$ or \bowtie . k means that we have to choose k services among services of the same type. \bowtie means that we choose all available services of the specified type. Algorithm 1 builds G_{AT} as if there are no restrictions (\bowtie case). We sort all subgraphs $g_{Si,j}$ of p_{Si} in $F(g_{Si,j})$ ascending order. Note that this is not necessary if we don't want to minimize the set of $F(g_{Si,j})$ but only to find a suitable assembly. Note that if there were no non-functional constraints, the first combination would be optimal since all $g_{Si,j}$ are sorted ascending.

Algorithm 3 checks if functional constraints are fulfilled and will find a suitable combination if it exists but will not be necessarily the optimal in terms of $\min(F(g_{Si,j}))$. Indeed, to avoid time-consuming tasks, we are not aiming at the optimal value. The combination will, however, be near the optimal solution

Algorithm 2 ATS compliant

- 1: for all p_{Si} of G_{AT} with S_i starting service do
- 2: We compute all subgraphs $g_{Si,j}$ of p_{Si} taking into account AT.Constraints.
- 3: Sort of all subgraphs $g_{Si,j}$ in $F(g_{Si,j})$ ascending order.
- 4: end for

since $g_{Si,j}$ are sorted in ascending order as specified in algorithm 2.

Algorithm 3 ATSF compliant

- 1: We choose one subgraph $g_{Si,j}$ for each p_{Si} taking into account AT.Constraints.
- 2: for all combination do
- 3: **if** number of bindings to each service is less or equal than the threshold value (S.Threshold) **then**
- 4: This combination has been built according to the application template AT in respect of structural (number of bindings) and non-functional constraints (QoS compliance).
- 5: Break
- 6: **else**
- 7: The combination is rejected.
- 8: end if
- 9: end for

5.3.3. Example of Execution of the Assembling Core Algorithm

We present in this section an example of execution of the assembling core algorithm. It includes seven services named A_1 , A_2 , A_3 of type t_A , B_1 , B_2 , B_3 of type t_B and C_1 of type t_C .

$$S = \{A_1, A_2, A_3, B_1, B_2, B_3, C_1\}$$

590 $T = \{t_A, t_B, t_C\}$

585

 $A_1.Type=t_A$

```
A_2.Type=t_A
A_3.Type=t_A
B_1.Type=t_B
B_2.Type=t_B
B_3.Type=t_B
C_1.Type=t_C
```

These are the service QoS nominal values (e.g. response time) obtained with a calibration procedure. Specified thresholds (number of simultaneous requests) are the values not to exceed in order to comply with the defined nominal QoS.

```
A_1.\mathrm{QoS}{=}1,\ A_1.\mathrm{Threshold}{=}1 A_2.\mathrm{QoS}{=}1,\ A_2.\mathrm{Threshold}{=}1 B_3.\mathrm{QoS}{=}1,\ A_3.\mathrm{Threshold}{=}1 B_1.\mathrm{QoS}{=}2,\ B_1.\mathrm{Threshold}{=}2 B_2.\mathrm{QoS}{=}3,\ B_2.\mathrm{Threshold}{=}3 B_3.\mathrm{QoS}{=}4,\ B_3.\mathrm{Threshold}{=}1 C_1.\mathrm{QoS}{=}5,\ C_1.\mathrm{Threshold}{=}3
```

610

All services should be assembled according to the following application template AT:

AT.Body=
$$\{(t_A, t_B), (t_B, t_C)\}$$
AT.Constraints= $\{2,1\}$

The AT template means that the \mathbf{t}_A services should be connected to two \mathbf{t}_B services and all \mathbf{t}_B services should be connected to one \mathbf{t}_C service.

Algorithm 1 execution: We applied the Algorithm 1 to determine the M_{QoS} matrix including all computed link QoS. Services send a request according to the AT template (See Figure 5). Step 1 shows the initial state of services. Each service is aware of the existence of others. According to AT template, t_A services should be connected to t_B services. So, t_A services send a request to t_B

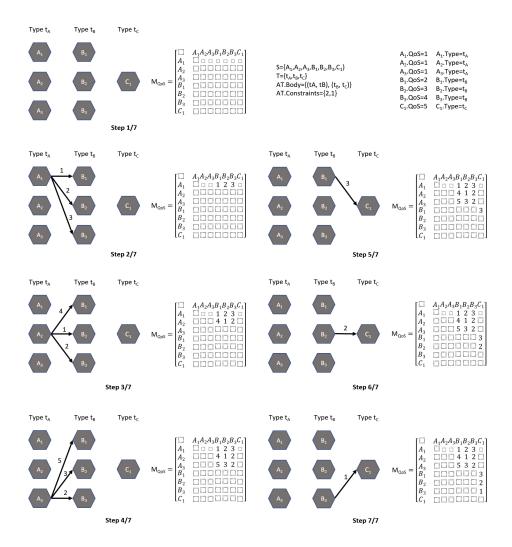


Figure 5: Algorithm 1 - AT compliant graph / \mathcal{M}_{QoS} design.

services. Step 2 shows the service A1 sending a request to B_1 , B_2 , and B_3 . B_1 , B_2 , and B_3 compute the link QoS (response time) and inscribe it in M_{QoS} . The same procedure is repeated for each of the first four steps. According to AT template, t_B services should be connected to t_C service. So, t_B services send a request to t_C services (steps 5 to 7). Algorithm 1 build an AT compliant graph called G_{AT} (Figure 6) and compute all link QoS.

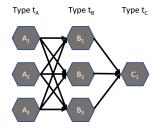


Figure 6: AT compliant graph: G_{AT} .

Algorithm 2 execution: We take into account the structural constraints
AT.Constraints= $\{2,1\}$ by applying the algorithm 2. There are three subgraphs
to consider, each with a different starting service: p_{A1} , p_{A2} , and p_{A3} . For each
of them, we determine all subgraphs taking into account AT.Constraints i.e. t_A services should be connected to two t_B services and all t_B services should be
connected to one t_C service. For each subgraph s, we compute F(s) i.e. the
maximum processing time of the subgraph. Figure 7 deals with p_{A1} , Figure 8
deals with p_{A2} , and Figure 9 deals with p_{A3} . We also indicate which services
are used for each of the subgraph. Figure 10 shows results sorted by ascending
order.

Algorithm 3 execution: For each combination, choosing one subgraph for each starting service (A_1, A_2, A_3) , we check if non-functional constraints are fulfilled i.e. if the number of bindings to each service is less than or equal to the threshold value (S.Threshold). Otherwise, the combination is rejected (Figure 11). We compute the number of used B_1 , B_2 , B_3 , and C_1 service for each combination. After five steps, the algorithm finds a suitable combination. The final assembly is given in Figure 12.

5.3.4. Benefits of the Assembling Core Algorithm

650

The assembling core algorithm can potentially be run at anytime:

- When a service appears, or disappears on the network. A new service can be more efficient (better QoS) or can have a better link.
- When a service has failed and should be replaced. The assembling can

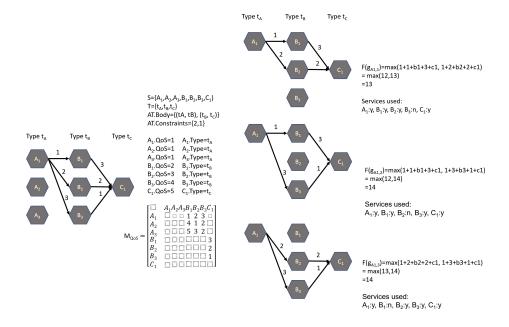


Figure 7: Algorithm 2 - ATS compliant graph: Computing of all subgraphs $g_{A1,j}$ of p_{A1} taking into account AT.Constraints.

thus be considered as a kind of adaptation mechanism. The algorithm avoids the service and proposes a new assembling. The composition is more robust and reliable.

When communication (link) is becoming bad. For example, when a service
moves away from another one or in case of electromagnetic disturbed environment. A too bad communication can be the cause of the disappearance
of a service.

655

660

The assembling core algorithm can thus potentially be used as self-adjusting, self-adapting, and self-healing mechanisms.

For the user, it simplifies fastidious manual assembling of a lot of devices. The service composition is performed automatically without human intervention.

The assembling is dynamic. The application template AT is provided to

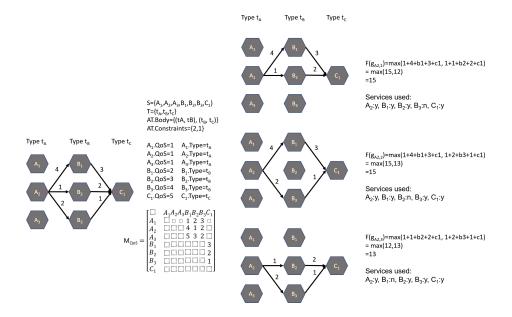


Figure 8: Algorithm 2 - ATS compliant graph: Computing of all subgraphs $g_{A2,j}$ of p_{A2} taking into account AT.Constraints.

perform a function. We can change later the template and thus change the function by reusing some or all of the current services.

Some assembling results lead to well-known organizations like Edge [43, 44], Dew [45] or Fog computing [46, 47] [48].

The algorithm can be useful for the run time, as mentioned, but also for the designing of the whole application. The architect has to make choices like the number of services of each type, in particular, to prepare and deploy in the environment so that a future need is answered. He can use the algorithm to simulate different organizations before deploying and implementing them in the real world (Self-simulation property).

5.4. Resource Consumption and Limitations of the Algorithm

675

In this section, we estimate the limitations of the algorithm on a real IoT platform. The aim of this section is to estimate the maximum number of services, SCC in our case, that could be assembled with our algorithm in real

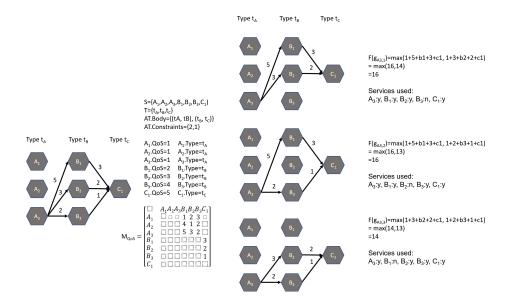


Figure 9: Algorithm 2 - ATS compliant graph: Computing of all subgraphs $g_{A3,j}$ of p_{A3} taking into account AT.Constraints.

| | | F | B ₁ | B ₂ | B ₃ | C ₁ |
|----------------|-------------------|----|----------------|----------------|----------------|----------------|
| A ₁ | g _{A1,1} | 13 | У | У | | У |
| | g _{A1,2} | 14 | У | | У | У |
| | g _{A1,3} | 14 | | у | У | У |
| A ₂ | $g_{A2,3}$ | 13 | | у | У | У |
| | g _{A2,1} | 15 | У | У | | У |
| | $g_{A2,2}$ | 15 | У | | У | У |
| A ₃ | g _{A3,3} | 14 | | у | у | У |
| | g _{A3,1} | 16 | У | У | | У |
| | g _{A3,2} | 16 | У | | У | У |

Figure 10: Results of the algorithm 2 (y means that the corresponding service is used).

conditions, in a reasonable amount of time and, if possible, low-resource consumption (memory). In Section 5.4.1, we present the equipment used for the implementation of our algorithm. In Section 5.4.2, we determine the worst case for assembling constraints: when the algorithm might experience difficulties. Finally, in Section 5.4.3, we present the chosen layouts and we analyse the experimental results.



Figure 11: Algorithm 3 - ATSF compliant graph (The chosen graphs are colored in dark gray).

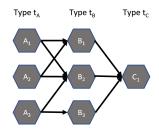


Figure 12: Final suitable assembly.

5.4.1. The Equipment Used for the Implementation

The raspberry pi has been chosen to make our experimentations. Our experimental devices are composed of:

- Raspberry Pi card, proposed by the British Raspberry Pi Foundation, is a one-board nano-computer, about the size of an ARM processor-based credit card [49].
- Android Things operating system is an embedded operating system based on Android designed by Google [50]. This system is meant to be used on devices linked to the Internet of Things. It is therefore designed to use as little memory as possible and to be energy-efficient.
 - 3.5-inch touch-screen monitor, compatible with the Raspberry Pi card, allows users to create all-in-one projects such as tablets, entertainment systems, embedded projects and devices for the Internet of things using interaction with the touch screen [51].

5.4.2. Combination Analysis

685

690

695

Constrains are defined for each level. They represent the number of services that should be connected. The architect has to specify the number of services that should be chosen among available services of the same type. At each level, k services have to be chosen among n available services. The number of combinations is

$$\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!} \tag{1}$$

 $\binom{n}{k}$ is read as "n choose k". There are a few combinations if k is very small or very big but become maximal if k is close to n/2. Choosing the case of n/2 services should be avoided and is extremely rare in real life. Common cases consist to choose only one service (1 of n) or all available services (n of n).

Notice that our study seems to be similar to the shortest path problem which can be solved by several well-known algorithms (Dijkstra's [52], Bellman-Ford [53], A* search [52], Floyd-Warshall [54], Johnson [55], or Viterbi [56]) but is

very different and more complex because (i) starting and ending services are not unique, (ii) they are searching for a path while we are searching for a sub-graph and (iii) the number of bindings to a node is upper bounded.

Usually, an application is monolithic, i.e. the same for each user. It offers features to meet a set of needs even if the user does not use them all. With our approach, the performance of the application can only be better because the assembly is dynamic. Customized applications are created according to the user's current needs. A service is integrated into our assembly only if necessary.

5.4.3. Layouts and Result Analysis

715

We have chosen two services layouts. Threshold values, link QoS and SCC processing time are random. The first layout has only one level. It is the simplest one and allows us to evaluate the maximum number of SCC able to be assembled for a given layer. This arrangement is made to evaluate the limitation of chosen all, 1, 2, or n/2 services. The scheme of the layout is given in Figure 13. If the architect asks to choose all available services or only one of them, which are the most commonly used requirements, the algorithm is able to assemble respectively 1 million and 500 000 services. If the demand becomes more combinatorial like choosing 2 or the worst n/2 services among all available SCC of a given type, the number of assembled SCC decreases respectively to 1200 and 20. The amount of memory necessary for the assembling is reasonable and doesn't exceed 100 MB in all cases. The number of $g_{Si,j}$ computed by the algorithm is also given for information for all four cases. The processing time does not exceed 20s in all cases. It might become, however, excessive depending on the case study but can easily be reduced if we accept to reduce the size of the assembling.

The second layout is excessively combinatory and is designed to push the algorithm as far as we can to its very limits. This layout consists in building a pyramidal organization by decreasing the number of SCC of one at each level until the raspberry pi is at the end (Figure 14). If the architect asks to choose all available services or only one among them, the algorithm is able to assemble

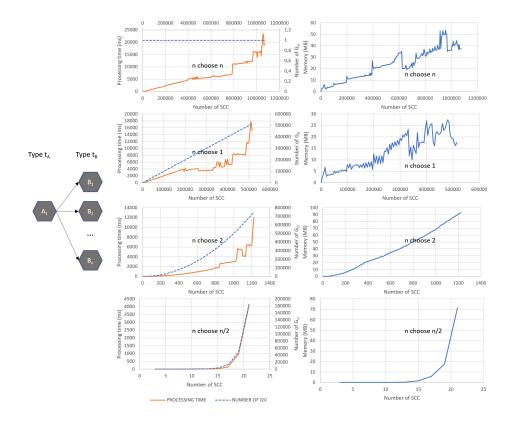


Figure 13: One-layer layout and results of its implementation.

respectively 91 and 28 SCCs. If the demand consists in choosing 2 or the n/2 services among all available SCC of a given type, the number of assembled SCC decreases respectively to 20 and 28. The amount of memory necessary for the assembling remains reasonable and doesn't exceed 37 MB in all cases. The processing time is reasonable and doesn't exceed 3.5s for all the last three cases (n choose 1, n choose 2 and n choose n/2) but becomes excessive if we choose all services at each level (n choose n) especially with 91 SCCs. This remains, however, correct for 55 SCCs (9 layers, 10 types), the processing time being 2.3s.

Measurements are taken in a way that the measuring process has none or few effects on the evaluated service. To minimize the impact of collecting mea-

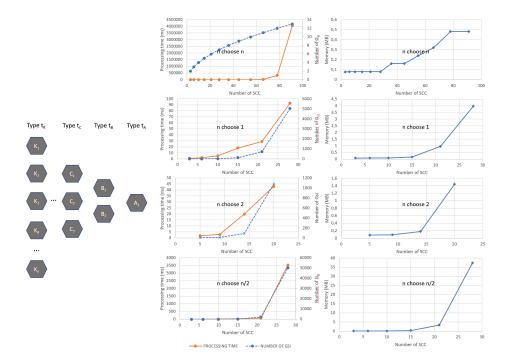


Figure 14: Pyramidal decreasing layout and results of its implementation.

surement data, the QOsControl component is a thread that regularly requests their measurements to the two monitors and write them in a log file which is collected at the end of the experiment.

750

Analysis of the approximate resource overhead (computation and memory) has been written in [18]. We have analyzed the extra code needed by the sub-components located in the SCC membrane: InMonitor, OutMonitor and QoSControl. We have used profiling tools to measure memory and CPU consumption of our approach at runtime. The memory and CPU consumption of our control mechanism is very low for common usage. However, it must be compared to the size of the functional part. The control mechanism should have no or a little influence on the monitored service, so it must represent a low percentage of resource usage compared to it. Indeed, if the micro-service is too small, the processing done by the control mechanism can be in the same order

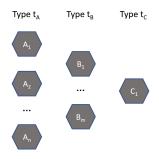


Figure 15: Three layers pyramidal decreasing layout.

760 of magnitude.

775

In conclusion, as expected, increasing the number of layers (the number of service types) increases the number of combinations. The number of layers should not exceed 9 (10 different service types) to obtain a processing time of a few seconds whatever the layout and constraints. The amount of used memory (100 MB in the worst case) is correct for a Raspberry Pi 3 Model B that has 1 GB available for all organizations. The number of assembled SCC can be higher (up to 1 million) depending on the chosen services arrangement and constraints. As we said, n choose 1 case and n choose n cases are commonly used and will present good results. In addition, computations could be distributed on all devices, which could significantly improve performance. The higher the number of SCCs, the greater the complexity, but the higher processing capabilities. Algorithms 1 and 2 can be easily parallelized since each startup service could do a part of the work. More complex distributions are possible.

6. Case Study: Medical Warning System

In this section, we deal with a case study related to the Internet of Things where our approach could prove useful. The idea is to show that our algorithm can be applied easily to use practical cases and can propose a suitable dynamically service assembly with a good processing time.

Due to the importance of observing the medical state of patients who are suffering from acute diseases, especially cardiovascular diseases, a continuous

remote patient monitoring is essential. The monitoring system can provide real-time analytical services, prediction and alerts in case of an emergency for users carrying out any activity, anywhere and anytime. It can recognize an abnormal user activity and can thereby detect falls on the ground or sickness. The user and the associated medical staff receive local notification in case of current or near-future predicted emergency. In addition, it reduces the response time of the system and increased his reliability. With the help of wearable wireless sensors, the system provides a continual access to medical parameters of a patient. IoT Gateways are located in every room in the house in a way to follow the patient. User sensors are always close to a gateway. They are equipped with computational capacity. They monitor the current state of the patient and provide a means to predict future medical condition via machine learning methods and artificial intelligence algorithms. They include the following capabilities: collect and aggregate data from the available sensors, implement data analysis and extract valuable information and knowledge from the incoming raw data, make decisions, detect emergency and make prediction, perform data compression and encryption and adapt data capture frequency. They are able to contact the rescue teams according to the type of emergency detected and to notify the nearest hospital of the arrival of a patient (Figure 16). These smart gateways make early recognition of deterioration more reliable by bringing the decision-making core and critical notifications closer to the patient. Due to the constant incoming data in a continuous medical monitoring, the system may encounter problems such as latency in system response, data transmission and computations related to data analytics. QoS has to be controlled from end to end. As previously shown, our SCC has been designed for that aim. As the user moves, service assembly is adapted in a way to follow him according to the requested QoS. The processing time of the services chain from end to end is thus controlled. This system takes advantage of local gateway processing to present notification and feedback with the minimum latency. The assembly is dynamic whenever the patient is moving. We compute an estimation of the processing time of the services assembly on a Raspberry Pi.

Each sensor is an SCC, each IoT gateway includes one or several SCCs and external services (hospital, rescue team) one or several SCCs too. To simplify, we consider only one SCC in an IoT gateway and in the external services.

815

The layout is here a pyramidal decreasing layout (Figure 15). It's often the case in IoT because data are processed and analyzed at each level. The result comes thus from aggregated data from several sources. The number of services at each level so decreases. These assembling layouts lead to well-known organizations like Edge [43, 44], Dew [45] or Fog computing [46, 47] [48] with three layers. Note that all calculations were done on a single raspberry pi. Calculations could be distributed on each device including SCC, which could considerably improve performance.

We build the first layer with 10 SCCs (wearable user sensors) of type tA, 9 SCCs for the IoT gateways (Second layer, type tB) and five SCCs for hospitals (third layer, type tC) and two SCCs for the rescue teams (third layer, type tD) (Figure 15). Sensors have to be connected to one gateway. The gateway sends notifications to one hospital and one rescue team. Threshold values, link QoS and SCC processing time are random. Each IoT gateway SCC is able to accept 10 sensors SCCs simultaneous requests. Results are: number of SCCs: 26, used memory: 280 KB, and processing time: 40 ms. Our algorithm is efficient for this use case. Assembly processing time is about 40 ms. So assembly can be done in a continuous way.

Discussion: Implementation of such systems in real IoT scenarios faces significant/some challenges and constraints. We highlight some of them and give some relevant references.

As objects communicate with each other using an open channel, i.e., Internet, so security and privacy along with integrity of messages always remain a concern. Trust management among these objects is thus an open issue. [57] provide an elaborated view of trust management with a focus on Social IoT (SIoT) by comparing different existing trust management schemes based on the trust management process, parameters chosen for trust evaluation, characteristics of trust functions and objectives achieved by them.

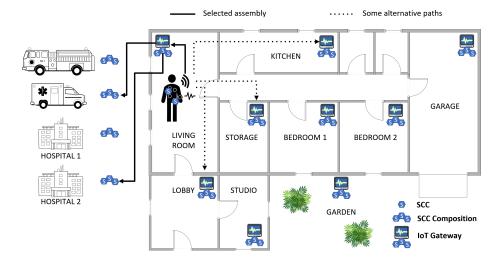


Figure 16: Self-assembly for a medical warning system.

Cognitive Internet of Things (CIoT) is another paradigm of IoT developed to enhance the capabilities of intelligence in IoT objects where these objects can take independent decisions in any environment. Unsolicited or undesirable electronic messages could disturb the internal cohesion of the assembly. The power of learning, thinking, and understanding by these objects, can make the information access more accurate and reliable but Web spam is one of the challenges while accessing information from the web. [58] present an intelligent cognitive spammer framework which eliminates the spam pages during the web page rank score calculation by search engines.

In medical scenarios, due to the potential remoteness of the services, the latency (link QoS) between them has to be controlled (what our solution does) but has to be minimal too. The need of health care systems like telediagnosis, or robotic telesurgery is increasing day-by-day. With these telesystems, doctors and patients need not to be present at the same location, but its requirements are fast response, ultra-reliable, and high availability. Traditional networks like 2G, 3G, and 4G-LTE are not able to fulfill these requirements, but [59] has shown that tactile internet and its applications in 5G era could solve this problem.

7. Conclusion

In this paper, we have proposed a self-assembling solution based on our self-controlled components that check the current behavior of the service and its conformity with the contract. This solution has several advantages. The dynamic assembling core algorithm can potentially be run at anytime and can potentially be used as self-adjusting, self-adapting, and self-healing mechanisms. For the user or architect, it simplifies fastidious manual assembling of a lot of devices. The service composition is made automatically without human intervention. The architect can use the algorithm to simulate different organizations before deploying and implementing them in the real world Our approach can build and maintain an assembly of services that, besides functional requirements, also fulfill global quality-of-service and structural requirements. Resource consumption and limitations of our approach have been reviewed. We have shown the feasibility of our approach on a case study. In a future work, we plan to deploy the proposed self-assembly algorithm on a full-scale real application scenario by taking into account the algorithm's energy consumption.

Acknowledgment

885

This work is supported by the European Telecommunications Standards Institute (ETSI) project entitled: User-centric approach in the digital ecosystem (Specialist Task Force: STF BM/543). The authors would like to thank Pr. Bernard Lemaire for his help in finalizing this paper.

[1] O. Khriyenko, V. Terziyan, O. Kaikova, User-assisted Semantic Interoperability in Internet of Things: Visuallyfacilitated Ontology Alignment through Visually-enriched Ontology and Thing Descriptions, in: In: Proceedings of the Sixth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2012, 2012, pp. 23–38.

- [2] IBM Cloud (2020).

 URL https://www.ibm.com/fr-fr/cloud
- [3] Amazon web services (2020).
 URL https://aws.amazon.com/fr

895

900

- [4] Microsoft azure (2020).URL https://azure.microsoft.com
- [5] B. Cheng, J. Yang, S. Wang, J. Chen, Adaptive Video Transmission Control System Based on Reinforcement Learning Approach Over Heterogeneous Networks, IEEE Transactions on Automation Science and Engineering 12 (3) (2015) 1104–1113. doi:10.1109/TASE.2014.2387212.
 URL http://ieeexplore.ieee.org/document/7015599/
- [6] J. Wu, B. Cheng, M. Wang, J. Chen, Energy-Efficient Bandwidth Aggregation for Delay-Constrained Video over Heterogeneous Wireless Networks, IEEE Journal on Selected Areas in Communications (2016) 1–1doi:10.1109/JSAC.2016.2632599.
 URL http://ieeexplore.ieee.org/document/7756386/
- [7] C. Simpkin, I. Taylor, G. A. Bent, G. de Mel, S. Rallapalli, L. Ma, M. Srivatsa, Constructing distributed time-critical applications using cognitive enabled services, Future Generation Computer Systems 100 (2019) 70–85. doi:10.1016/j.future.2019.04.010.
 - URL https://linkinghub.elsevier.com/retrieve/pii/S0167739X18311762
- [8] P. Štefanič, M. Cigale, A. C. Jones, L. Knight, I. Taylor, C. Istrate, G. Suciu, A. Ulisses, V. Stankovski, S. Taherizadeh, G. F. Salado, S. Koulouzis, P. Martin, Z. Zhao, SWITCH workbench: A novel approach for the development and deployment of time-critical microservice-based cloud-native applications, Future Generation Computer Systems 99 (2019) 197–212. doi:10.1016/j.future.2019.04.008.

- URL https://linkinghub.elsevier.com/retrieve/pii/S0167739X1831094X
 - [9] A. Ochian, G. Suciu, O. Fratu, C. Voicu, V. Suciu, An overview of cloud middleware services for interconnection of healthcare platforms, in: 2014 10th International Conference on Communications (COMM), IEEE, Bucharest, Romania, 2014, pp. 1-4. doi:10.1109/ICComm.2014.6866753. URL http://ieeexplore.ieee.org/document/6866753/

- [10] C. I. Cook, Patent us 10,412,172 b2 internet of things (iot) self-organizing network 37.
- [11] P. Lalanda, J. A. McCann, A. Diaconescu, Autonomic Computing: Principles, Design and Implementation, Springer Publishing Company, Incorporated, 2013.
 - [12] M. Samaniego, R. Deters, Internet of Smart Things IoST: Using Blockchain and CLIPS to Make Things Autonomous, in: 2017 IEEE International Conference on Cognitive Computing (ICCC), 2017, pp. 9–16. doi:10.1109/IEEE.ICCC.2017.9.
 - [13] Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang, An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends, in: 2017 IEEE International Congress on Big Data (BigData Congress), 2017, pp. 557–564. doi:10.1109/BigDataCongress.2017.85.
- [14] T. Aubonnet, N. Simoni, Self-control cloud services, in: 2014 IEEE 13th International Symposium on Network Computing and Applications, NCA 2014, Cambridge, MA, USA, 21-23 August, 2014, 2014, pp. 282–286. doi: 10.1109/NCA.2014.48.
- [15] Tatiana Aubonnet and Noëmie Simoni, Service creation and selfmanagement mechanisms for mobile cloud computing, in: Wired/Wireless Internet Communication - 11th International Conference, WWIC 2013,

- St. Petersburg, Russia. Proceedings, 2013, pp. 43–55. doi:10.1007/978-3-642-38401-1_4.
- [16] The opencloudware project (2015).

 URL http://www.opencloudware.org/

960

- [17] T. Aubonnet, L. Henrio, S. Kessal, O. Kulankhina, F. Lemoine, E. Madelaine, C. Ruz, N. Simoni, Management of service composition based on self-controlled components, Journal of Internet Services and Applications 6 (15) (2015) 17. doi:10.1186/s13174-015-0031-7.
- [18] F. Lemoine, T. Aubonnet, N. Simoni, IoT composition based on self-controlled services, Journal of Ambient Intelligence and Humanized Computingdoi:10.1007/s12652-020-01831-4.
 URL https://doi.org/10.1007/s12652-020-01831-4
- [19] F. Lemoine, T. Aubonnet, L. Henrio, S. Kessal, E. Madelaine, N. Simoni, Monitoring as-a-service to drive more efficient future system design, EAI Endorsed Transactions on Cloud Systems 17 (9) (2017) 1–15. doi:10. 4108/eai.28-6-2017.152754.
 - [20] L. Cavallaro, E. D. Nitto, C. A. Furia, M. Pradella, A Tile-Based Approach for Self-Assembling Service Compositions, in: 2010 15th IEEE International Conference on Engineering of Complex Computer Systems, 2010, pp. 43–52. doi:10.1109/ICECCS.2010.6.
 - [21] D. Weyns, R. Haesevoets, A. Helleboogh, T. Holvoet, W. Joosen, The MA-CODO middleware for context-driven dynamic agent organizations, ACM Transactions on Autonomous and Adaptive Systems (TAAS) 5 (1) (2010) 3.
 - [22] D. Weyns, R. Haesevoets, A. Helleboogh, The MACODO organization model for context-driven dynamic agent organizations, ACM Transactions on Autonomous and Adaptive Systems (TAAS) 5 (4) (2010) 16.

[23] S. Schuhmann, K. Herrmann, K. Rothermel, Y. Boshmaf, Adaptive Composition of Distributed Pervasive Applications in Heterogeneous Environments, ACM Transactions on Autonomous and Adaptive Systems 8 (2) (2013) 1–21. doi:10.1145/2491465.2491469.

970

- [24] D. Sykes, J. Magee, J. Kramer, Flashmob: distributed adaptive self-assembly, in: Proceedings of the 6th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, ACM, 2011, pp. 100–109.
- [25] C. Perera, P. P. Jayaraman, A. Zaslavsky, P. Christen, D. Georgakopoulos, MOSDEN: An Internet of Things Middleware for Resource Constrained Mobile Devices, in: 2014 47th Hawaii International Conference on System Sciences, 2014, pp. 1053–1062. doi:10.1109/HICSS.2014.137.
- [26] C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, Sensing as a service model for smart cities supported by Internet of Things, Transactions on Emerging Telecommunications Technologies 25 (1) (2014) 81–93. doi: 10.1002/ett.2704.
- [27] V. Terziyan, O. Kaykova, D. Zhovtobryukh, UbiRoad: Semantic Middle-ware for Context-Aware Smart Road Environments, in: 2010 Fifth International Conference on Internet and Web Applications and Services, 2010, pp. 295–302. doi:10.1109/ICIW.2010.50.
- [28] Calvin (2018).

 URL https://www.ericsson.com/research-blog/
 open-source-calvin/
 - [29] P. Persson, O. Angelsmark, Calvin Merging Cloud and IoT, Procedia Computer Science 52 (2015) 210–217. doi:10.1016/j.procs.2015.05. 059.
- 995 [30] Paraimpu (2020).

 URL http://www.paraimpu.com/

- [31] V. Tsiatsis, A. Gluhak, T. Bauge, F. Montagut, J. Bernat, M. Bauer, C. Villalonga, P. Barnaghi, S. Krco, The SENSEI Real World Internet Architecture, "Towards the Future Internet - Emerging Trends from European Research", IOS Press, 2010, pp. 247–256.
- [32] Node-red (2020).

 URL https://nodered.org/

1005

1010

- [33] M. Autili, P. Inverardi, M. Tivoli, CHOREOS: Large scale choreographies for the future internet, in: 2014 Software Evolution Week - IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering (CSMR-WCRE), 2014, pp. 391–394. doi:10.1109/CSMR-WCRE.2014. 6747202.
 - [34] P. Evensen, H. Meling, SenseWrap: A service oriented middleware with sensor virtualization and self-configuration, in: 2009 International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2009, pp. 261–266. doi:10.1109/ISSNIP.2009.5416827.
 - [35] Zero configuration networking (2020).
 URL http://www.zeroconf.org/
- [36] D. Guinard, V. Trifa, S. Karnouskos, P. Spiess, D. Savio, Interacting with the SOA-Based Internet of Things: Discovery, Query, Selection, and On-Demand Provisioning of Web Services, IEEE Transactions on Services Computing 3 (3) (2010) 223–235. doi:10.1109/TSC.2010.3.
 - [37] A. Palade, C. Cabrera, G. White, M. A. Razzaque, S. Clarke, Middleware for Internet of Things: A quantitative evaluation in small scale, in: 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2017, pp. 1–6. doi:10.1109/WoWMoM. 2017.7974340.
 - [38] Global Sensor Networks (GSN) (2014).

URL

1025

1030

1040

global-sensor-networks-gsn/

- [39] J.-P. Calbimonte, S. Sarni, J. Eberle, K. Aberer, XGSN: An Open-source Semantic Sensing Middleware for the Web of Things (2014) 16.
- [40] I. Corredor, J. F. Martinez, M. S. Familiar, L. Lopez, Knowledge-Aware and Service-Oriented Middleware for deploying pervasive services, Journal of Network and Computer Applications 35 (2) (2012) 562–576. doi:10. 1016/j.jnca.2011.05.009.
- [41] S. Cheshire, M. Krochmal, DNS-Based Service Discovery, RFC 6763 (Feb. 2013). doi:10.17487/RFC6763.
- [42] I.-I. O. f. Standardization, ISO International Organization for Standard ization (Sep. 2020).
 URL http://standards.iso.org/ittf/PubliclyAvailableStandards
 - [43] A. Jain, P. Singhal, Fog computing: Driving force behind the emergence of edge computing, in: 2016 International Conference System Modeling Advancement in Research Trends (SMART), 2016, pp. 294–297. doi:10. 1109/SYSMART.2016.7894538.
 - [44] E. Ahmed, M. H. Rehmani, Mobile Edge Computing: Opportunities, solutions, and challenges, Future Generation Computer Systems 70 (2017) 59–63. doi:10.1016/j.future.2016.09.015.
- [45] Y. Wang, Definition and Categorization of Dew Computing, Open Journal of Cloud Computing (OJCC) 3 (1) (2016) 1–7. doi:10.19210/1002.3.1.1.
 - [46] S. K. Datta, C. Bonnet, J. Haerri, Fog Computing architecture to enable consumer centric Internet of Things services, in: 2015 International Symposium on Consumer Electronics (ISCE), 2015, pp. 1–2. doi:10.1109/ ISCE.2015.7177778.

- [47] S. Sarkar, S. Chatterjee, S. Misra, Assessment of the Suitability of Fog Computing in the Context of Internet of Things, IEEE Transactions on Cloud Computing PP (99) (2015) 1–1. doi:10.1109/TCC.2015.2485206.
 - [48] K. Skala, D. Davidovic, E. Afgan, I. Sovic, Z. Sojat, Scalable Distributed Computing Hierarchy: Cloud, Fog and Dew Computing, Open Journal of Cloud Computing (OJCC) 2 (1) (2015) 16–24. doi:10.19210/1002.2.1. 16.
 - [49] Raspberry Pi (2020).
 URL https://fr.wikipedia.org/wiki/Raspberry_Pi
- [50] Android Things (2020).

 URL https://developer.android.com/things/

- [51] Raspberry Pi Touch Display (2020).

 URL https://www.raspberrypi.org/products/
 raspberry-pi-touch-display/
- [52] A. M. Lacorte, E. P. Chavez, Analysis on the Use of A* and Dijkstra's Algorithms for Intelligent School Transport Route Optimization System, in: Proceedings of the 4th International Conference on Human-Computer Interaction and User Experience in Indonesia, CHIuXiD '18, CHIuXiD '18, ACM, New York, NY, USA, 2018, pp. 12–17. doi:10.1145/3205946.3205948.
- [53] A. Schambers, M. Eavis-O'Quinn, V. Roberge, M. Tarbouchi, Route planning for electric vehicle efficiency using the Bellman-Ford algorithm on an embedded GPU, in: 2018 4th International Conference on Optimization and Applications (ICOA), 2018, pp. 1–6. doi:10.1109/ICOA.2018.8370584.
- [54] Z. Ramadhan, A. Putera Utama Siahaan, M. Mesran, Prim and Floyd-Warshall Comparative Algorithms in Shortest Path Problem, in: Proceedings of the Joint Workshop KO2PI and The 1st International Con-

- ference on Advance & Scientific Innovation, EAI, Medan, Indonesia, 2018. doi:10.4108/eai.23-4-2018.2277598.
- [55] S. Anitha, B. M. Ramesh, Network Reconfiguration for Loss Minimization by Using Johnson's Algorithm, in: 2018 4th International Conference on Electrical Energy Systems (ICEES), 2018, pp. 680–684. doi:10.1109/ ICEES.2018.8442416.
- [56] Y. Zhou, An Efficient Routing and Interface Assignment Algorithm for Multi-Channel Multi-Interface (MCMI) Ad Hoc Networks, in: Y. Zhou, T. Kunz (Eds.), Ad Hoc Networks, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer International Publishing, 2018, pp. 131–142.
- [57] R. K. Chahal, N. Kumar, S. Batra, Trust management in social Internet of Things: A taxonomy, open issues, and challenges, Computer Communications 150 (2020) 13-46. doi:10.1016/j.comcom.2019.10.034.

 URL https://linkinghub.elsevier.com/retrieve/pii/S0140366419310953
- [58] A. Makkar, N. Kumar, Cognitive spammer: A Framework for PageRank analysis with Split by Over-sampling and Train by Underfitting, Future Generation Computer Systems 90 (2019) 381-404. doi:10.1016/j.future.2018.07.046. URL https://linkinghub.elsevier.com/retrieve/pii/ S0167739X18305703
- [59] R. Gupta, S. Tanwar, S. Tyagi, N. Kumar, Tactile internet and its applications in 5G era: A comprehensive review, International Journal of Communication Systems 32 (14) (2019) e3981. doi:10.1002/dac.3981.
 URL http://doi.wiley.com/10.1002/dac.3981