A CoAP-based framework for collaborative sensing in the Semantic Web of Things

Michele Ruta*, Floriano Sciscio, Agnese Pinto, Filippo Gramegna, Saverio Ieva, Giuseppe Loseto, Eugenio Di Sciascio

Abstract

This paper proposes a novel Semantic Web of Things framework, enabling collaborative discovery of sensors and actuators in pervasive contexts. It is based on a backward-compatible extension of the Constrained Application Protocol (CoAP), supporting advanced semantic matchmaking via non-standard inference services. The framework also integrates efficient data stream mining to analyze raw data gathered from the environment and detect high-level events, annotating them with machine-understandable metadata. A case study about cooperative environmental risk monitoring and prevention in Hybrid Sensor and Vehicular Networks is presented and experimental performance results on a real testbed are provided.

Keywords: Semantic Web of Things, CoAP, Collaborative sensing, Resource discovery, Matchmaking, Data mining

1. Introduction and Motivation

The emerging Semantic Web of Things (SWoT) vision joins together the Semantic Web and the Internet of Things (IoT). It aims to enable new classes of smart applications and services by augmenting real-world objects, locations and events with semantically rich and machine-understandable information, conveyed through unobtrusive, inexpensive and often disposable micro-devices. Environmental monitoring is among the most relevant and challenging application scenarios. It requires coping with hard issues, such as: large-scale data and sensor management; volatility of resources, users and devices; heterogeneity of hardware/software platforms; dependence on context; strict computational resource constraints. The Constrained Application Protocol (CoAP) is becoming one of the most widely accepted application-layer protocols for things networks. Nevertheless, it currently allows only a basic data-oriented representation of resources and elementary retrieval procedures relying on string matching between requests and resource attributes, with just binary “yes/no” outcomes. Exact request/resource matches are very uncommon in real-world scenarios, with heterogeneous devices, sensors and actuators from several independent providers.

* Corresponding author. Tel.: +39-339-635-4949; fax: +39-080-596-3410
E-mail address: michele.ruta@poliba.it

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.05.425
The SWoT needs more effective resource discovery, supporting also approximate matches and possibly providing a relevance metric between each available resource and a request. By solving this problem, one would strongly promote interoperable collaboration in large, multi-party Wireless Sensor Network (WSN) deployments and federations. As a consequence, integrating smart objects and WSNs with Semantic Web technologies and infrastructures is a currently relevant research trend. Various solutions have been proposed, exploiting reference ontologies to annotate data, devices and services and sharing sensor data along the Linked Open Data guidelines through RESTful\textsuperscript{3,4} web services. The SPITFIRE\textsuperscript{5} service infrastructure for semantic applications leveraged Internet-connected sensors and lightweight protocols like CoAP. Sensors were described as RDF triples and service discovery was based on metadata such as device features or location. The use of semantics for automatic sensor composition was also exploited to enable a user-driven exploratory search aiming to select the most appropriate sensors for the particular problem\textsuperscript{6}. Unfortunately, those works allowed only basic queries in SPARQL fragments on RDF annotations. More advanced resource discovery features were not supported. Data and sensor management in mobile and pervasive contexts require techniques such as ontology-based complex event processing\textsuperscript{7} and semantic matchmaking\textsuperscript{8}. The latter in particular supports approximated matches and resource ranking with explanation of outcomes, by means of logic-based inference services. To reach this goal, this paper borrows technologies from the Semantic Web to define a comprehensive SWoT framework for fully decentralized cooperation. The approach manages high-level annotations of data streams, devices, events of interest and services, with a well-defined meaning w.r.t. a shared domain conceptualization (i.e., ontology). From a technological standpoint, the proposal integrates: (i) slight backward-compatible extensions to CoAP and CoRE Link Format (IETF CoRE Working Group RFC 6690, http://tools.ietf.org/html/rfc6690) resource discovery protocol; (ii) high-level event detection and annotation through resource-efficient data mining algorithms on raw data gathered by a Semantic Sensor Network (SSN, i.e., a semantic-enhanced WSN) using the SSN-XG ontology\textsuperscript{9} as reference vocabulary; (iii) non-standard inferences for semantic-based matchmaking\textsuperscript{8} for resource retrieval and ranking, supporting approximate matches besides full ones. A case study on collaborative environmental risk monitoring and management in Hybrid Sensor and Vehicular Networks (HSVNs) is presented to validate and explain the approach. A testbed was developed implementing the framework with real devices and experiments were executed.

2. Semantic Sensor Networks for advanced context extraction

The proposed reference architecture extends an earlier version of the CoAP-based framework\textsuperscript{10}. Sensors deployed in an area communicate with a local sink node, which acts as cluster head. Multiple sinks are connected to a gateway, interfacing the network toward the outside. Each sensor is characterized not only by data-oriented attributes, but also by a semantic annotation describing its features and functionalities. Sinks are able to: (i) register sensors along with their semantic descriptions as CoAP resources; (ii) support logic-based resource discovery on annotated metadata, leveraging a lightweight embedded matchmaker\textsuperscript{8}. For these purposes, sink nodes embed CoAP servers. They also gather and process data for event detection. When an event is recognized, it is annotated and a resource record is updated in the server. Beyond the semantic annotation, the record contains further extra-logical context parameters, such as geographic coordinates and a timestamp. The gateway waits for resource discovery requests from client applications searching for events in the area, and replies on behalf of connected sink nodes.

The two prototypical modules developed in the basic framework\textsuperscript{10} were improved to support a collaborative sensing process. Communication in SSNs was implemented using a modified version of Californium CoAP library (http://eclipse.org/californium/), enabling the semantic-based enhancements of the CoAP protocol\textsuperscript{10}.

**JOSM SSN plugin.** Figure 1 shows the prototype GUI of the SSN plugin for the Java OpenStreetMap (OSM) open source editor (http://josm.openstreetmap.de/). It can be used to perform the following tasks: (i) **SSN browsing**, showing on the map in (A) the available sensors and sink nodes registered on CoAP gateways; (ii) **Semantic-based sensor discovery**, for customizing a semantic-based CoAP request (by specifying reference location, maximum discovery range, inference task to perform and relevance threshold) visually through panel (B) and sending it to look for sensors in the area; (iii) **SSN scenario generation** to create random configurations for large-scale SSN simulations, through the panel (C) shown in Figure 1, which extends the *OSM to Rescue* plugin\textsuperscript{11} (http://kaspar.informatik.uni-freiburg.de/~osm/). Scenarios can be customized according to the parameters reported in Table 1.

**CoAP Mobile Node.** An Android-based client was developed using Android SDK Tools (Revision 21.1, corresponding to Android Platform version 4.2.2, API level 17) and tested on a Samsung GT-i9250 Galaxy Nexus smartphone.
Fig. 1: JOSM plugin for CoAP-based SSNs

Table 1: Parameters for scenario generation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>number of sink nodes</td>
<td>G</td>
<td>num. of CoAP gateways (GWs)</td>
</tr>
<tr>
<td>Dmin</td>
<td>min num. of CoAP sensors per sink</td>
<td>S min</td>
<td>min num. of sinks connected to a CoAP GW</td>
</tr>
<tr>
<td>Dmax</td>
<td>max num. of CoAP sensors per sink</td>
<td>S max</td>
<td>max num. of sinks connected to a CoAP GW</td>
</tr>
<tr>
<td>dMaxS</td>
<td>max distance in m between sink and sensors</td>
<td>dMaxG</td>
<td>max distance in m between two connected GWs</td>
</tr>
</tbody>
</table>

It was devised to support in-the-field communication with SSNs and to perform: (a) **SSN browsing and sensor discovery**, in which the user can select a gateway node and view all connected sensors or only devices retrieved after a semantic-based discovery. Each sensor can be also queried to retrieve data it measures; (b) **Collaborative sensing**: when a mobile node (e.g., an Android smartphone) queries a CoAP gateway, it can be also configured automatically as an information source, connected to the gateway temporarily. It can provide data coming from both embedded sensors (e.g., accelerometer, gyroscope) and external sensing peripherals available through wired or wireless connections. These data can further characterize the reference environment, enabling improved event detection. In this way, mobile nodes are enticed to share their perceptions with the rest of the SSN in order to obtain a more accurate feedback.

The proposed SSN framework includes a simple yet effective data mining method, devised to extract relevant information from sensor readings and annotate it, consisting of the following steps. (i) Read and collect data from the sensors embedded in the device, as well as from external sensors in the field through standard CoAP requests. (ii) Compute average, variance and standard deviation for the current time window, so as to assess the variability of collected information within the monitored area. (iii) Compute incremental ratio of the above indexes w.r.t. previous time windows, in order to highlight trends and significant changes. (iv) For every data collection, define a (binary or multiple) classifier, to detect relevant events when given conditions occur. In the case study in Section 3, mining and event detection are executed at gateway level after sensors send data via standard CoAP frames. (v) The output of each classifier is a logic-based expression constructed according to knowledge modeled in a reference ontology, which formalizes a conceptualization of the sensing domain.

CoAP adopts the **CoRE Link Format** specification for resource discovery. This protocol only allows a syntactic string-matching of attributes, lacking an explicit and formal characterization of the resource semantics. To overcome this limitation, semantic-based CoAP protocol enhancements allow to exploit non-standard inferences for automated semantic sensor discovery and composition. **Concept Covering** is particularly useful in SWoT scenarios, such as sensor networks, where data must be gathered from specific and different types of sensors to infer proper events. A client application composes a discovery request and queries a SSN gateway to find a set of most suitable sensors, among those managed by sinks directly connected to the gateway, having a semantic description expressed in OWL (Web Ontology Language, http://www.w3.org/TR/owl2-primer) w.r.t. a shared ontology. The gateway carries out semantic matchmaking by solving a Concept Covering Problem (CCoP), in order to find the set of resources which together satisfy the request to the maximum extent. In case of a partial cover, the response can also include both the semantic description of the uncovered part (H) of the request and the percentage of request not covered. This value is obtained comparing H w.r.t. the starting request by means of semantic-based ranking algorithms. Furthermore, exploiting the proxy support built into CoAP, the gateway has the possibility to forward the uncovered part as a new request towards other SSN nodes in the area of interest, searching for more resources to satisfy missing features. In this way, each semantic-enabled gateway can start a collaborative and multi-hop resource discovery.
3. Case study: collaborative environmental monitoring

In order to clarify the proposed approach and show its benefits, a case study in cooperative environmental monitoring is reported. It focuses on Hybrid Sensor and Vehicular Networks (HSVNs), which merge Vehicular Ad-Hoc Networks (VANETs) and WSNs. In HSVNs, sensors are distributed along roads to monitor and gather information about the environmental conditions of a given area. Furthermore, vehicles receive safety warnings and traffic information from deployed Road-Side Units (RSUs) through Vehicle-to-Infrastructure (V2I) wireless communication technologies. Each RSU is a CoAP gateway and periodically queries sinks in its range. Sinks perform Concept Covering for semantic-based discovery to find suitable sensors and return the most appropriate device set to the RSU. The latter can now start obtaining raw data from sensors and detects weather events via data mining, as described in Section 2. Event annotations are then exposed to warn vehicles about current driving risk factors. Extending the SSN-XG ontology along the Stimulus-Sensor-Observation design pattern, both observed parameters (e.g., temperature, humidity, wind speed) and sensor measurement capabilities (e.g., accuracy, resolution, frequency) were defined.

It is morning. A car is travelling on SS16, a highway near Bari, Italy. The road has low-density traffic with 90 vehicles flowing per hour. Possible risks are due to crossroads. This environmental monitoring scenario was simulated with an SSN randomly generated by the JOSM plugin described in Section 2. Figure 2 depicts three RSUs, eight sinks and fourteen sensors in the network. The car (blue icon in the picture) is driving near the RSU1 gateway, which composes a discovery request D, using concepts defined in the domain ontology, as reported in Figure 3 in OWL 2 Manchester Syntax. The CoAP request also includes: (i) the RSU reference location P, defined through the attributes latitude and longitude; (ii) maximum distance md; (iii) minimum covering threshold sr for resource retrieval. In particular, RSU1 looks for devices located near SS16 with a maximum distance of 3000 m from P and a coverage threshold value of 90%. After a distance-based pre-filtering, RSU1 solves the CCoP applied to sensors. Figure 3 reports concept expressions for some of the sensors inside the measurement area in Figure 2 and connected to gateway node RSU1. Connected sinks retrieve a covering set \( S_c(\text{RSU1}) \) composed of LM70Sensor, BMP085Sensor and FS11Sensor. Nevertheless, this set does not fully cover the request: an uncovered part \( U_{RSU1} \) is returned, corresponding to 37% of \( D \). In detail, no anemometer or humidity sensor has been retrieved, and LM70Sensor does not completely satisfy the required temperature measurement capabilities. Consequently, RSU1 sends a CoAP semantic request to the reachable gateway RSU2, forwarding \( U_{RSU1} \) to discover remaining sensors. \( S_c(\text{RSU2}) \) is composed of Hts2030Sensor, while \( U_{RSU2} = 14\% \). Similarly, RSU2 forwards \( U_{RSU2} \) to RSU3, obtaining BitLineBLVSensor. Finally, RSU2 returns the discovered sensor set to RSU1, also specifying the final uncovered part \( U_{RSU3} \), corresponding to 5% of the original \( D \).

Now RSU1 can query and observe sensors to acquire data for weather event detection. Using the process described in Section 2 (computation details not shown due to lack of space), the classifier identifies Fog and Rain events in the example. The corresponding semantic annotations become resources for a further matchmaking process carried out for vehicle safety. RSU1 waits for vehicles equipped with a wireless interface entering its radio range. Let us suppose that the vehicles described in Figure 4 drive nearby RSU1 and are equipped with a prototypical system for real-time monitoring and driving assistance. Therefore, each of them is able to interpret data extracted from On-Board Diagnostics (OBD-II) car interface and smartphone sensors, integrating locally detected environmental information and potential risk factors into the request. Consequently, the RSU can use the provided information to further enrich event annotation, e.g., traffic level, road pavement conditions, and so on. RSU1 will perform matchmaking between vehicle descriptions and weather events, reported in Figure 4, each annotated in terms of safety requirements a car must implement to limit risks. Finally, for each (vehicle, event) pair, RSU1 exploits Concept Abduction inference service to detect risk level. The X5 is the safest vehicle, because it is equipped with snow tires (also useful in case of rain), fog lamps, ABS and ESP. The A3 has higher risk levels due to its medium-high speed, despite the activated ABS and fog lamps. A high speed and inadequate safety features make the 600 absolutely unsuitable.

4. Experiments

To prove the feasibility of the proposed framework, a performance evaluation was carried out in a testbed with real devices. Tests aimed to measure the amount of data exchanged between network nodes and the time spent by RSU1 to obtain the list of sensors useful for monitoring environmental conditions. Semantic-enabled CoAP servers for RSU1 and the three sinks it manages were executed on different Raspberry Pi Model B (specifications:
Request ≡ Sensor and (hasTemperature only (LowRes. and LowAcc. and HighLatent.)) and (hasVisibility only (LowAcc. and LowFreq.)) and (hasOperatingRange only LowMedAltit.) and (hasPressure only (LowAcc. and MediumRes.)) and (hasWindSpeed only (MediumRes. and LowAcc. and LowPrec.)) and (hasHumidity only (MediumAcc. and MediumRes. and HighFreq.))

L70Sensor (Si) ≡ TemperatureSensor and (hasTemperature only (LowAcc. and HighRange and MediumFreq.)) and (hasOperatingRange only LowMedAltit.)

BMP085Sensor (Si) ≡ Barometer and (hasPressure only (LowAcc. and MediumRes. and LowRange and LowPrec.))

FS11Sensor (Si) ≡ VisibilitySensor and (hasVisibility only (LowAcc. and LowRange and LowFreq.))

Hts2030Sensor (Si) ≡ HumiditySensor and (hasHumidity only (MediumAcc. and MediumRes. and HighRange and HighFreq.))

BitlineBLYSensor (Si) ≡ AnemometerSensor and (hasWindSpeed only (MediumAcc. and LowRes. and MiddleRange and LowPrec.))

Table 3: Basic CoAP vs semantic CoAP

<table>
<thead>
<tr>
<th>Time (ms)</th>
<th>Data (byte)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RX (byte)</td>
<td>TX (byte)</td>
</tr>
<tr>
<td>Basic CoAP</td>
<td>150</td>
</tr>
<tr>
<td>Semantic CoAP</td>
<td>21</td>
</tr>
<tr>
<td>RSU1</td>
<td>1232</td>
</tr>
<tr>
<td>RSU2</td>
<td>7</td>
</tr>
<tr>
<td>RSU3</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 2: Network performance evaluation

https://www.raspberrypi.org/products/model-b/) boards, configured with Wheezy Raspbian (http://www.raspbian.org) operating system. Gateways RSU2 and RSU3, their sinks and sensors were simulated via the JOSM plugin running on a workstation (HP Z820 with Xeon E5-2643 quad-core CPU at 3.3 GHz, 16 GB RAM, 1 TB hard disk) Table 2 shows the results on time and data exchanged within the simulated scenario. In particular, for each RSU the following parameters were measured: wait time (Tt); the total amount of bytes transmitted and received (BTx, BTx) with its sinks to obtain sensors descriptions; the amount of time spent (Te) for executing the Concept Covering process locally; the time (Tf) and bytes exchanged (BFx, BFx) with neighboring RSUs after forwarding the uncovered part of the request. For times, five runs of each test were executed and the average of the last four runs was taken. Table 2 also highlights the performance gap between the real node (RSU1) and the simulated ones (RSU2, RSU3), due to memory and processing constraints on the Raspberry Pi. However, turnaround time result appears acceptable (= 1.7 s). The most time-consuming step is Ti, since it includes the data structures setup by the embedded reasoner running on the board. As found in5 for other reasoning tasks, Te in RSU1 exhibits a similar trend w.r.t. RSU2 and RSU3; time is roughly an order of magnitude higher on the Raspberry Pi node. Concerning data exchanges, there is a significant difference between the size of BTx and the other ones (BTx, BFx, BFx). Indeed, the packets each RSU sends to connected sinks to obtain the list of semantic resources contain only standard CoAP queries. Instead, replies from sinks and packets exchanged between each RSU include the new query parameters, defined in10, including the compressed annotations.

A performance comparison was executed between a standard and a semantic CoAP request/response session, in a network composed only by real nodes. In this test the semantic request included only the temperature and operating range measurement capabilities detailed in Figure 3, so as to require a single resource category via Concept Abduction resolution10. For standard CoAP, the request is characterized only by the query parameter rt="Temperature" as a straightforward string searching for generic temperature sensors and the output produced is a simple list of resources managed by the RSU1 with no detailed information. Conversely, although heavier in terms of processing time and packet size as depicted in Table 3, the semantic-enhanced CoAP protocol allows to specify a more precise query and obtain an accurate and smaller response, listing resources in relevance order with related metadata.

Benefits of the proposed approach w.r.t. the state of the art were assessed in a comparison with works cited in Section 1. Table 4 shows only the proposed approach combines fitness for resource-constrained environments (by
using CoAP and a distributed search strategy), expressiveness of sensor modeling (by exploiting OWL 2) and support for both exact and approximated matches, with formally grounded resource ranking and composition.

5. Conclusion

The paper described an advanced Semantic Sensor Network framework. It exploits backward-compatible CoAP extensions for semantic-based resource description, management and discovery. Efficient data stream mining and collaborative sensing are further notable features of the proposal. A case study in a HSVN scenario and experimental tests on a real testbed allowed to evaluate both feasibility and usefulness of the approach. Albeit processing times and network load are higher than in standard CoAP, the improvement in the quality of discovery justifies the proposed approach in complex scenarios like large-scale distributed environmental monitoring.

Future work includes the combination of machine learning algorithms with semantic-based sensor data management for more flexible context mining, as well as the integration of specialized compression algorithms for Semantic Web languages\(^\text{13}\) to reduce storage and network load.

Acknowledgements. The authors acknowledge partial support of Apulia region cluster project PERSON.

References