Providing adaptive web services from MHs is a new approach in Mobile Web Services to cope resource scarcity of mobile network environment. This approach is explored through investigating mechanisms that are used to facilitate mobile web service adaptation. Offloading and migration mechanisms allow continuous and reliable provision of mobile services through distributing the execution of mobile web services and modeling the transfer of required location-based information. The distribution can be classified into Forward or Bounce offloading while the transfer modeling is based on either Frontend or Backend scheme. The paper provides an evaluation of two distinct prototypes with different distribution schemes using resource intensive applications.

Key words: Backend-Bounce Offload, Frontend-Forward Offload.

1. Introduction

Location-based applications can be hosted from mobile devices and has shown performance enhancement to companies who have employees deployed in the field[1]. Providing location-based service does often dependent on the current location of the service provider. For instance, providing the latest updated news and scene snapshots for a specific location in a predefined format requires portable devices with built-in GPS and cameras that are capable to move to the actual place of the event. Furthermore, it requires MHs that are aware of their location publish the event as a live feed and take latest information gathered at the current location. MHs allow processing of the gathered information and make them available, instantly, to clients. However, some of these services require non-interrupted provision from mobile devices to allow providing instant information before it becomes obsolete. Some other evolved services are complex and consume a large amount of the constrained resources. In spite of the fact that these constraints may be eliminated in the future and the resource capabilities might advance, the ideal performance and the minimum latency will always be the dominant requirements. In addition, the rapid growth of complex services and the increase of user demands, retain the scarcity of resources. Furthermore, battery life will always be a bottleneck. Consequently, a new trend in MWSs has appeared and attracted by researchers. Applying mechanisms that allow MWSs adaptation to compensate for the lack of resources has been explored by some researchers. For example, [2] have proposed a partitioning technique that allows the execution of complex large web services on MHs. However, this approach focuses on one resource only, which does not address the intermittent provision of MWSs. Continuous provisioning of MWSs in P2P network environment has been studied in [3] This is accomplished through service migrating to another surrogate mobile node when the MH’s battery power drains. However, this framework does not allow a lightweight process of complex mobile web services. Moreover, these approaches are for general services and don not address location-based services that require the existence of location-based information. Also, both approaches support only Simple Object Access Protocol (SOAP)-based web services that require heavy-weight parsers and large message payloads. Recent research studies focus on building resource-aware web service provisioning architecture that supports Representational State Transfer (RESTful)-based mobile web services [4] and compares between SOAP- and REST-based mobile service hosts. The results indicate that REST-based services are more suitable for mobile network environments. However, continuous provisioning of complex services is not explored in [4]. The scope of this work is to allow adaptive and reliable location-based MWSs. The remainder of this paper is organized as follows: Section II contains a brief overview of the technical approach. Section III presents a critical analysis of the results. Finally, Section IV summarizes the paper.

2. Technical Approach

The purpose of this research as mentioned above is to investigate mechanisms and technologies that facilitate web service adaptation. Offloading and migration mechanisms are used to achieve dynamic distribution of mobile web services. The execution logic of a web service is fragmented into partitions that are executed remotely on different Auxiliary MHs (AMH). This mechanism is called Offloading. Simultaneously, the method for transferring the required location-based data is modeled. This mechanism is called Migration. In this paper, Offloading is classified based on the methodology used for handling requests and responses into two main types: Bounce- and Forward-Offloading. Moreover, Migration is based on the type of data extraction process into Backend and Frontend. Hence, these classifications results in four different distribution strategies: Backend-Bounce Offload (BBO), Frontend-Bounce Offload (FBO), Backend-Forward Offload (BFO) and Frontend-Forward Offload (FFO). The scope of this work is constrained on investigating and analyzing BBO and FFO schemes. Fig. 1 illustrates BBO, where the MH sends the request back to the client associated with location address of the selected AMH; the required
parameters are invoked by the AMH from the original MH. On the other hand, FFO elaborated in Fig. 2 forwards the request from MH to AMH in conjunction with the required parameters. Two distinct types of MWS frameworks have been implemented each of these architectures represents a different strategy for achieving adaptive and distributed MWSs. In our experiments, RESTful-based MH Web Service Framework (MHWF) \cite{4} has been extended to implement the aforementioned architectures. This is because it has been shown in our preliminary work that RESTful- outperforms SOAP-based mobile web services. The architecture of the Extended MHWF (EMHWF) consists of seven main modules: WebService Servlet, HttpListener, Request Handler, Fuzzy Logic, Offloading Module, Message Parser and Response Composer. HTTP Listener’s, Request Handler, Parser Module and Response Composer are identical to those described in\cite{4}. The service hash table of WebServiceServlet module includes an additional field for defining the URI of the selected AMH. Fuzzy Logic is used for triggering distribution, monitoring resources and activating Offloading Module when MH is overloaded. The main task of Offloading Module depends on the offloading scheme. For example, in BBO, it changes the role acted by MH temporarily from server to client and forwards requests back to client. On the other hand, in FFO, it transfers the request and data from MH to AMH. Another important task is to select the AMH, which fulfills some predefined policies. In the next section we will investigate BBO and FFO schemes.

3. Key Results and Discussion

In this section we present and test two prototypes for providing, executing and deploying distributed MWSs. The architectures are fairly similar and there is no apparent difference in complexity. However, the main difference is in the offloading module. This difference causes distinctions in their performance and resource consumptions. The service used is a simple Image-Processing service. This location-based service is one of the most common resource intensive applications that can be provided from MHS, where clients send requests to the MH to process an image and adjust its dimensions (height, width or bit depth) to be compatible with the client’s screen size. In our experiments, we modify only the image’s height (l). This modification allows the client to vary the size of transferred message payloads. Thus the application is suitable for testing the performance of all the implementations. Each prototype consists of three mobile devices: MH, which is executed on a mobile device (Nokia N97m) running MIDP 2.1 over Symbian OS, AMH that is executed on an N97m and the client, which is executed on Nokia N80. The test is carried out for two different scenarios. In the first set of experiments, the level of internal resource consumptions is tested by measuring the average amount of memory capacity for 50 requests. In the second set of experiments, the level of external resource consumptions is examined by evaluation of the required bandwidth. Results presented in Fig. 2 show that: in FFO scheme, memory increases with increasing l, but in BBO, it remains almost steady state. This is expected because in Forward-Offload, MH allocates more memory for storing the increased response size before it is forwarded to the corresponding client. In addition, Backend-Migration requires more memory to open new socket connections with AMH and maintain input/output data streams than its corresponding Frontend scheme. Fig. 3 shows that as the input value l increases the size of the response message increases, which results in an increase in bandwidth. This increase is more obvious in FFO scheme over its corresponding BBO schemes. This raise in the bandwidth due to the two links traverse of response messages in FFO: The first link is from AMH to MH and the second is from MH to client. Moreover, BBO requires enquiry messages but there is no existence of enquiry messages in FFO.

4. Summary of the work, potential impact & Conclusion

Adaptive mobile web services can help to provide non-interruptible complex web services. Offloading and migration are the mechanisms used to facilitate the provisioning of adaptive mobile web services. We investigated two distinct mobile distribution service strategies: Backend-Bounce and Frontend-Forward schemes. The two implementations were tested and analyzed using a resource intensive application. The results presented show that basing distributed mobile hosted services on Backend Bounce Offload strategy is more suitable for mobile network environment.

Key References


Aircraft’s Datalink Approach for DTN
Applications and Fair Allocation of Resources Model
Mohammed Al Siyabi
Year 3 PhD Student, C.C.S.R., e-mail:m.alsiyabi@surrey.ac.uk

Abstract

DTN is an overlay network working over heterogenous networks in challenged environment. Due to the difficult working conditions of DTN, the physical transportations means such as buses and ferries have been used to carry data. Using aircrafts in scheduled routes for data transportation is proposed as a novel DTN carrier which can be a secondary industry business to Airline companies. Furthermore, Fair Allocation of Resources Model (FARM) is proposed which will be a complementary mechanism to many DTN routing algorithms. The simulation produced higher delivery probability and fairness results when applying FARM which enhances the DTN QoS.

Key words: Delay Tolerant Network (DTN), Aircraft, Quality of service (QoS) and Resource Management.

1. Introduction

Links and contacts in DTN environment are considered very limited and every means of possible transmissions might be used to transmit massages. Many types of transportations are used in the literature [1][2][3] such as ferries, buses and bicycle to carry messages. However, using aircraft in scheduled route as DTN bundle carrier is a novel trend for DTN applications. Aircrafts fly daily routes and pass over remote locations where the communication infrastructure is limited. They can be used to relay the user’s messages to their destinations.

As shown in Fig.1, commercial aircraft routes cover vast area of the world which makes them a good choice for carrying data as they pass by any interested users.

Fig.2 shows the basic Aircraft DTN architecture. The ground DTN node is located at remote area with limited communication infrastructure under a scheduled flying route of commercial airline company. The aircraft will store or deliver the message to the Air Traffic Control (ATC) then to its final destination through the backbone network.

Furthermore, the notion of Quality of Service (QoS) in Delay Tolerant Network (DTN) is difficult due to the challenging nature of DTN which might suffer from intermittent disconnections and long delays. The requirements for QoS in DTN and Fair Allocation of Resource Model (FARM) for DTN application are proposed which will be a complementary mechanism to many DTN routing algorithms. FARM is an admission control mechanism based on collecting local information of the node resources in order to avoid network congestion and enhance DTN QoS.

2. Technical Approach

Air transportation is the most popular, fastest, safest, widespread means of long distances transportation on earth. Commercial air transport represents one of the major aspects of the world economy and transporting around 1.8 billion passengers annually over thousands of flights. This gives indication of the potential application for this facility and might add a secondary service to airlines company business to act as DTN data carriers. This paper is discussing the utilization of commercial flying aircrafts to provide a data transmission service to ground users. Aircrafts do not have any terrain obstacles and can fly over all types of topography. There are many remote areas (such as oceans and deserts) where it is very expensive to lay the terrestrial network and there are many flights passing over some of these places on predictable time and location. Therefore, they can be used as mobile routers and data mules to carry DTN bundles along their flying routes.

Furthermore, Aircrafts mobility concept and FARM is configured in Opportunistic Network Environment (ONE) [4] which is a simulator for DTN application. The simulator produced good results and is indicating the advantage of aircrafts and FARM in DTN environment.
performance by best utilizing the available resources and applying the techniques which will provide this enhancement. Higher delivery probability will indicate better resource utilizations and higher fairness index will indicate better resource sharing. Therefore, QoS in DTN is more of a resources management. DTN has scarce resources which will lead to contention among users. Controlling the admission to these resources is useful to provide better control over these resources and eventually better QoS.

FARM will assign the whole bandwidth to users one after another. The selected nodes will be the one who satisfies the highest admission qualification metrics conditions. These metrics data are generated and updated by all interested nodes and used by the passing aircrafts to decide who to admit based on the highest weight of the qualification metrics. The metrics data for each node are the number of their past rejected requests, number of their past accepted request, their CoS and their FIFO orders. The history data is built with every passing aircraft and will be considered during the next admission evaluation process by other aircrafts. The results showed the effectiveness of FARM in DTN environments because it will enhance the DTN QoS by providing better delivery probability and higher fairness index compared with same scenario lacking it.

To analyses FARM behavior, Fig. 4 shows the delivery probability for a scenario of 1 aircrafts and 20 users per the three locations. The simulation is run with and without FARM and the results show improvements in the delivery probability vary between 10-23 %. These improvements will be limited at the first run of the simulation at 1000 s because at this stage, the history database of the model is at the data collection phase. Longer simulation time will cause more feedback data from the ground nodes to be recorded and as a result, the delivery probability will be enhanced with the longer simulation time.

3. Key Results and Discussion

The research scenario assumes three remote locations, A, B and C where each area has a group of users. These locations have limited communication infrastructure and they use three types of available transportation means; buses, ferries and aircrafts to exchange messages. Fig. 3 shows the comparison results and the obtained delivery probability for a scenario of 2 vehicles and 20 users per location when FARM model is enabled. It shows that aircrafts have higher delivery probability between 85-95 % compared with buses and ferries whose values range between 20 to 30 %. Overall, aircrafts are having up to 62 % improvements over buses and ferries. This is due to that all users are within the aircraft range; therefore, they all have better opportunities to send their data. On the other hand, buses and ferries have smaller range because they are at ground level; therefore, not all the users are in their range. Furthermore, the transmission availability time i.e. contact line of sight LOS, is larger with the aircrafts because they are at high altitude and therefore can stay longer time over the ground users to enable them to send more data, while the buses and ferries are staying shorter time and they become out of range. This shows the potential of using aircrafts in this special situation and its usefulness at remote locations.

4. Summary of the work, potential impact & Conclusion

Aircraft is proposed as data carrier which might add secondary business to airline companies. It is more appropriate for poor communication infrastructure areas, i.e. Africa and the sea, and less appropriate in countries with advanced infrastructure, i.e. N America and Europe. QoS in DTN is more of a resources management and allocation. Controlling the admission to these resources is useful to provide better QoS. FARM is proposed to provide higher delivery probability and fairness when applied as a complementary mechanism to DTN routing algorithms.

Key References

Feature Descriptions for Enabling Knowledge Derivation in Generic Applications

Hasini De Silva

Year 5 PhD Student, C.C.S.R., e-mail: H.DeSilva@surrey.ac.uk

Abstract

The popularity of highly personalised and context-aware applications has made the automated adaptation of generic applications and services in different environments challenging. This often requires the derivation of knowledge related to the target environment. For example, in a system where the location of an entity is described by the feature position, the location of a group of entities could be derived based on the set of position values of the members of that group. This paper proposes attaching a feature description with each concerned feature that specifies how to derive knowledge related to that particular feature.

Key words: knowledge derivation, feature description.

1. Introduction

Generic applications and services play a significant role in terms of their interoperability and reusability in different environments. The purpose of a generic application is to enable different target environments to re-use it in similar situations. However, the generic applications have to be adapted to the target environments accordingly. The increasing demand for such personalisation has made maintaining generality challenging. Therefore, it is important to find mechanisms to enable the automated adaptation of generic applications to variable environments without imposing generic structures for context definition that impairs the descriptiveness of these environments. This paper proposes a method for target environments to specify how knowledge is to be derived from a set of information as required by generic applications.

Using rules is a mechanism that is used to specify methods of derivation of knowledge based on existing information. Rule-based reasoning is effective in a nominal feature space, where derivations depend on logical relationships rather than arithmetic functions. However, defining conventional rules to aggregate a group of values is problematic since the number of predicates in the premise of a rule is fixed while the group size varies. A work around for this problem is found in [1], where they cluster the group members into a fixed number of clusters, so that a rule containing that number of predicates can be applied. This method however, restricts the functionality in cases where clustering is in applicable or when chosen number of clusters does not reflect the natural clustering.

A rule-based system is presented in [2], where rules are used to specify how to merge a set of (potentially inconsistent) structured news reports, which are data structures containing simple phrases with semantic information encoded as tags. They define ‘Fusion rules’ for the merging process along with a background knowledgebase. A number of approaches are presented for aggregation in terms of a set of aggregation predicates which cover a range of aggregation techniques. However, fusion rules can use only these pre-defined aggregation predicates, which limits the expressiveness of the rules.

This paper presents an approach for specifying how to derive knowledge. Section 2 describes the problem and presents the proposed feature description method while section 3 presents a prototype implementation that proves the concept. Finally, the paper is concluded in Section 4.

2. Feature Description

The knowledge derivation addressed in this work is based on a group of feature values. A feature \( f \) is used to describe an entity \( x \in X \) in the considered environment by means of a function \( v_f : X \rightarrow \mathbb{R} \), where \( f \) is the set of all possible values that can be taken by that feature. This feature could be numeric (e.g. age = 5), nominal (e.g. location = London) or a complex data type (e.g. position = (0.12, 1.23)). A group of \( n \) feature values for a particular feature refers to the values taken by a group of entities and is denoted by the \( n \)-tuple \( v^* \in V^n_f \), where \( V^n_f = V_f \times V_f \times \ldots \times V_f \) is the \( n \)-ary Cartesian product of \( V_f \). For example, if \( V_f = \{a, b\} \), a possible group of three feature values is \( v^* = (a, a, b) \). The derivation of knowledge from a group of feature values is defined in terms of the notion of metrics using a function \( k \) that maps a group of values \( v^* \in V^n_f \) to a metric \( m \in M \). The set \( M \) can be defined depending in the nature of the derived knowledge. For example, consider the group of values \( v^* = (v_1, v_2, \ldots, v_n) \) where each \( v_i \in \mathbb{R} \) (i.e. the set of real values). A metric average, can be defined by setting \( M = \mathbb{R} \), using the function \( k_{\text{avg}} : \mathbb{R}^n \rightarrow \mathbb{R} \). In this case the definition of the function would be as follows: \( k_{\text{avg}}(v_1, v_2, \ldots, v_n) = (v_1 + v_2 + \ldots + v_n) / n \).

The aim of this work is to define a mechanism that can be used to specify such definitions for knowledge derivation by means of metrics. For this purpose, it is proposed to attach a feature description with each concerned feature to enable target environments to specify how to derive knowledge required by generic applications. Describing a feature by...
means of a value space description and a set of metric inference rules is proposed. A base ontology, given in Error! Reference source not found., is proposed which contains the basic structure for describing a value space and metrics related to a feature. This can be extended to build a specific ontology that represents a particular feature.

As a means of proving the concept of feature description, a prototype system has been developed, using the Jess Rule Engine [5] that utilises feature descriptions in order to perform operations on a group of values in the value space of a particular feature. The system is able to derive knowledge, in terms of metrics, as defined by a particular feature description, for a given group of values in that feature value space. A query includes the feature name, metric name and the list of values in concern. The module first loads the corresponding feature description and the rules are activated. The input parameters of the query are then loaded to the knowledgebase, which generates the metric value in concern as the output, and is extracted by the main module.

Assume that the numeric feature user preference represents the preference of a user for a particular item by an integer value ranging from 0 to 10. The metric, group value, can be defined by the following rule with the use of a user-defined function size.

\[ \text{hasList}(?g, ?X) \rightarrow \text{hasGroupValue}(?g, \text{divide}((\text{sum}(?X)), \text{size}(?X))) \]

These rules and function definitions are transformed into Jess rules and function definitions. When a query is received, the input parameters are loaded to Jess’ working memory which fires the rules. When a query for the group value for a list of user preferences, say for example \{1,1,3,5,2,3\}, is loaded into the working memory, the rule execution asserts the relationship hasGroupValue(“userPrefGroup”,”2.5”) as a Jess fact. The result is then extracted and output as the result for the query.

4. Conclusion

This work presents a feature description method for the specification of methods of knowledge derivation, encoded as rules defined on a value space described by an ontology structure. The introduction of user-defined functions enhances the descriptiveness of the rules while the use of qualifiers to describe arguments of functions enables the use of functions with a variable number of arguments, which is often required when dealing with groups of values.

References

Aligning the SensorWeb with the Internet of Things vision
Suparna De

Research Fellow, C.C.S.R., e-mail: S.De@surrey.ac.uk

Abstract

The Internet of Things concept envisions a future where sensors are engrafted in the fabric of our environment. This engenders two challenges: to provide semantics to sensor observation and measured values and to bring sensor services to the Web in a lightweight manner. This paper presents a RESTful approach that brings real-world services offered by sensors to the Web, by leveraging the semantic capabilities of the Resource Description Framework.

Key words: REST, RDF, Sensor Web.

1. Introduction

The vision of the Internet of Things (IoT) relies on the provisioning of real-world services, which are provided by sensors or embedded systems that are directly related to the physical world. The rapid development of sensor technology has involved diverse sensor types, both remote and in-situ, with diverse capabilities [1]. Moreover, the relevant Wireless Sensor Networks (WSNs) also feature heterogeneous protocols. However, it has been pointed out in [2] that the services offered by sensor nodes share significant similarities, such as atomic operations that read data from a sensor. This observation has lead to research initiatives that offer sensor measurement data services in a homogeneous and application-accessible way. Most of this research has been directed at applying Device Profile for Web Services (DPWS)-based implementations [3, 4] to sensor gateways. DPWS defines a limited set of WS-* standards for resource limited devices. Offering sensor data through web services offers easier access to the application space and offers the scope of interoperability with existing web services for composition. Another initiative in this direction has been the concept of the sensor web, which has been defined as “web accessible sensor networks and archived sensor data that can be discovered and accessed using standard protocols and application program interfaces” [1]. Standardization efforts for the sensor web have been driven by the Open Geospatial Consortium’s (OGC) [5] suite of XML schemas for describing sensors and measurement data and web service interfaces for requesting and retrieving observations. This paper identifies two challenges to align the sensor web with the vision of the Internet of Things: firstly, sensor capabilities and observation measurements need to be defined in a semantic way to provide contextual knowledge essential for situation awareness. Secondly, lightweight mechanisms need to be implemented so as to retrieve these descriptions and data directly from the sensors. This paper presents an approach to meet these challenges with a RESTful framework that manipulates semantically annotated sensor data.

2. Technical Approach

The approach presented here leverages the Semantic Web activity of the W3C and the REpresentational State Transfer (REST) architectural principles. The relevant components are described below.

Semantic Sensor Description

The Semantic Web activity aims to formally define the semantics, or meaning, of information on the Web. Resource Description Framework (RDF) is a key technology for representing information and exchanging knowledge on the Web. In the work presented here, the observation and measurement (O&M) values of sensors are annotated with metadata that provides information on how to interpret the value. The metadata consists of the sensor provider, the unit of measurement, data type of the value and a time stamp of the measurement. An example O&M description of a sensor that measures temperature values in degree Celsius is given below:

![Fig. 1 RDF model for sensor O&M](image)

Fig. 1 RDF model for sensor O&M
REST uses URLs for encapsulating and identifying services on the web. HTTP is used as the implementation protocol. Web service clients can access a particular representation of the service by transferring application content using a small set of globally defined remote methods that describe the action to be performed on the service. The supported HTTP methods include HTTP GET, POST, OPTIONS, PUT and DELETE. These methods offer a lightweight mechanism to perform atomic operations on sensor services.

The approach presented here makes use of the CCSR sensor testbed, where sensor O&M readings are stored in a MySQL database. The developed web service reads these values from the database and applies the semantic annotation as detailed in the previous section. The semantic O&M RDF envelope is then exposed through HTTP GET method. For instance, the following URL returns the RDF envelope that contains the current temperature value of the sensor in the testbed lab (node 3):

http://localhost:8188/rep/s/3/temp

Other supported REST methods include OPTIONS to know the sensor type.

3. Implementation and Discussion

This section presents the design and implementation of a RESTful sensor web service and client application that manipulates an actuator simulator. Figure 2 shows the design of the developed application.

![Diagram](image)

Fig. 2 RESTful sensor application

The RESTful sensor application provides semantically annotated sensor O&M values through a REST WS interface. This is retrieved by the REST client by employing the HTTP GET method. The semantic annotation allows the client to retrieve all sensor values from a certain sensor type (e.g., temperature sensor) from a desired location by parsing the returned RDF envelope. The received RDF data provides information on both the sensor type through the namespace:sensorType annotation (i.e., sensor:Temperature) and the location, through the provider mapping contained in the rdf:about tag. In this way, all the temperature sensor values in the same location are retrieved and then averaged. The REST client GUI offers a user the ability to dynamically change the threshold value during run-time. This is achieved through a HTTP PUT interface to the REST client, which accepts the threshold value as a parameter. In this implementation, if the calculated average temperature value is less than the current set threshold, an ‘On’ parameter is sent to the actuator module, which turns on the heater. The user can change the threshold value during application execution, and the new value is communicated to the processor module through its exposed HTTP PUT method.

The above implementation shows that REST services and clients are particularly suited for applications where atomic operations on sensors, such as reading sensor values, are required. It has also been pointed out in the state of the art that REST services cover a fair part of the basic services offered by embedded devices, since such devices usually offer rather simple and atomic functionalities [2]. Moreover, the lightweight nature of REST makes it an ideal candidate to build a universal web-based API for sensors and embedded devices, thus bringing the real-world services offered by these devices to the virtual world. This web enablement can also contribute to horizontal collaboration of sensor services with more traditional enterprise services in a seamless manner. Incorporating semantics in the sensor descriptions will also allow enhanced reasoning capabilities to applications.

4. Summary of the work, potential impact & Conclusion

The IoT concept embraces a future in which sensors are ubiquitous in the environment. The proposed semantic, RESTful approach to sensor services offers a way of bringing real-world services to the web, thus contributing to the realisation of the Internet of Things.

References

Smart Middleware for Wireless Sensor Networks

Frieder Ganz

Year 1 PhD Student, C.C.S.R., e-mail: F.Ganz@surrey.ac.uk

Abstract

Sensors will play an important role in our future daily life. Ongoing research areas such as Smart Homes, Healthcare and the Future Internet leverage the use of sensors in combination with Web data and services. Large scale networks can produce a large variety of information related to human users and their environment. Utilising and integrating this information are the key issues to enable future environment and user aware networks and applications. We introduce a middleware design which addresses the emerging issues by constructing a homogeneous framework to connect heterogeneous sensor networks and providing decision making mechanism to optimise and utilising resource constrained wireless sensor networks.

Key words: Wireless Sensor Network, Middleware, LTE Gateway, Sensor data and query processing

1. Introduction

In the next few years sensors become increasingly important in our daily life. Sensors and actuators monitor and control physical objects and their interaction to provide machine-readable information. There are several ongoing research areas trying to introduce sensors into our everyday life. In the area of medical healthcare, sensors are used to transmit health information measured by sensors attached to the body to medical staff. Another emerging area is the smart home initiative which can exploit sensors measuring the energy consumption to recommend saving opportunities or provide mechanism to control and/or interact with different objects and devices in homes. The current efforts in the wireless sensor and application/service areas lead to interconnected networks of devices, sensors and physical and logical world objects called the Internet of Things (IoT) [1]. The sensors and actuators and the gained information and processed knowledge is made accessible to the end-user or consumer applications and services in an IoT platform.

The heterogeneity of the sensors used and the variability of the information gained from different scenarios demands a homogeneous way to retrieve and exploit the observation and measurement data captured by these devices.

In this paper we introduce a middleware component which abstracts the underlying low-level networks and provides a unified access layer to the sensors and the information they provide. We will focus on the emerging challenges which arise by connecting sensors and especially wireless sensor networks (WSN) to a higher level network (e.g. IP-based) and discuss the possible solutions and mechanisms to address those.

2. Challenges and Design Principles

To design a successful middleware solution, challenges and issues related to WSN characteristics and high-level applications must be addressed [2]. Sensors often have different purposes and therefore are based on different hardware and software protocols [4]. IEEE 802.15.4, Bluetooth, ZigBee and 6lowWPAN are just some used standards in industry and research. To address the problem of the heterogeneity the middleware has to provide a unified connectivity layer to all different sensor islands [5]. This connectivity layer which acts as a gateway bridges the gap between underlying sensor networks and higher-level applications. The middleware has the overview and control over the whole network and is therefore responsible to distribute incoming queries over the network. Query Processing and Distribution is a difficult topic in WSN as sensors are energy- and processing-constrained devices and not available all the time or not capable to satisfy an incoming query due to limited hardware. The middleware has to select sensors which are capable to satisfy a query. Context information such as location, energy-status, available capabilities, Routing and Radio information from each connected sensor node has to be used to satisfy the incoming queries. If nodes are not available or data freshness is not needed, caches can be consulted to save energy and response time. Another way to compensate faulty nodes is to query similar nodes in the same spatial area with the same capabilities. A query processing mechanism and/or language has to be developed which is aware of the WSN characteristics and introduce parameters such as data freshness, quality and priority. Data freshness represents the need of data served directly by the sensor. Is no freshness required the data can be provided by a cache. If the query cannot be satisfied by a particular node and a similar node has to be consulted the quality of the provided information decreases if the compensation node has less specific capabilities or the spatial area does not exactly match the demanded one. Priority parameters declare how fast a query has to be answered. As the heterogeneous network contains less-capable and higher-capable devices, middleware functionality could be transferred to higher capable devices to aggregate or process queries. The overall architecture has to be lightweight, if high-capable nodes join the network, middleware should balance the...
overall stability and quality of the network by transferring functionality to different nodes.

3. Middleware for Wireless Sensor Networks

The different challenges from Section 2 will be addressed in our architecture in different layers which are shown in Figure 1. The Connectivity Layer tries to establish a connection between different sensor platforms. On each sensor, a management module is introduced which connects sensors via multi-hop connections to the sink node which acts as a base station. The management module on the sensor side is implemented either in the Operating system or in the protocol stack of the node. This allows a zero-configuration like management of nodes which is in large WSN environments mandatory required and often not provided by other middleware solutions. The Information Processing Layer manages the overall network. Incoming queries will be distributed over the underlying sensor networks and nearby middleware units which manage sensor islands in a different spatial area. Defect and not available nodes will be compensated by either introducing caching techniques or the discovery of similar capable nodes in the near spatial area. Information which is not available at the queried middleware unit will be relayed to other connected units via peer to peer or hierarchical connections. The Service Layer provides the knowledge gained from the network and introduces connectivity to RESTful and WS-* based Web services [3].

The layers are mapped to its counterpart in the OSI Layer model. The connectivity layer in our middleware connects devices from physical, data link and network layer to a unified access layer. The current stage of work contains the connectivity for IEEE 802.15.4 capable nodes which either have the support in the Operating system (SunSpot) or in the protocol stack (TinyOs, 6loWPAN). Modified Session enabled protocol can be designed here to support M2M communication over different networks such as LTE. Incoming Queries will be processed in the Information Processing Layer, either they will be distributed to the attached underlying sensor network or relayed to other middleware units in other spatial areas.

The Application Layer is represented in the Service Layer. Interfaces are provided to connect high-level applications to the underlying networks. Web services such as the WS-* stack can be used to represent the sensor network and orchestrate and integrate network into existing applications which are capable of the standards.

4. Conclusions, Future Work and Impact

In this paper we introduced an initial architecture to connect heterogeneous sensor networks to a unified framework. In the future work, query distribution and processing in an heterogeneous environment has to be examined. Prediction mechanisms have to be researched. Prediction not in the sense of forecasting data, but how the network could change or what other nodes and events can be checked to efficiently address the upcoming queries. Another issues is orchestration and communication between different gateway (middleware components); how different data can be retrieved from them and how queries have to be forwarded in the overall set-up. The goal is to have one platform connecting different sensor nodes connected to spatial distributed middleware components to integrate into the future internet.

Key References


A Fuzzy Reinforcement Learning Approach for Pre-Congestion Notification based Admission Control

Stylianos Georgoulas

Research Fellow, Centre for Communication Systems Research, e-mail: S.Georgoulas@surrey.ac.uk

Abstract

Admission control at routers in computer networks, such as the internet, aims to compensate for the inability of slow-changing network configurations to react rapidly to load fluctuations. Even though many admission control approaches exist, most of them suffer from the fact that they are based on rigid assumptions about the per-flow and aggregate underlying traffic models, requiring manual reconfiguration of their parameters in a “trial and error” fashion as soon as these original assumptions stop being valid. In this paper we present a fuzzy reinforcement learning admission control approach based on the increasingly popular Pre-Congestion Notification framework, in an attempt to minimize the need for such manual and frequent reconfigurations.

Key words: Admission Control, Pre-Congestion Notification, Fuzzy Logic, Reinforcement Learning.

1. Introduction

The dynamicity of future internet networks, where applications with different service requirements can appear, makes Quality of Service (QoS) provisioning and service continuity a challenging issue. Traditional traffic engineering approaches, usually based on offline optimizations (bandwidth provisioning), may not be able to address this efficiently. Towards this end, dynamic service management functions such as admission control can play a significant role with respect to supporting QoS for application flows during the service delivery time, helping to overcome the inability of slow-changing network configurations to react rapidly to load fluctuations in order to prevent QoS degradation.

Even though admission control is a well-studied subject, most of the existing schemes suffer from the fact that they are based on some very rigid assumptions about the per-flow and aggregate underlying traffic models and network characteristics. They require, therefore, manual reconfiguration of their parameters in a “trial and error” fashion as soon as the original assumptions stop being valid, in order to keep performing well. The idea of mechanisms able to self-adapt as the flow and network conditions change has been around for quite some time under the generic term autonomic management. Past and existing projects [1] have been working towards inducing self-x behaviour in internet communication mechanisms. Towards this end in this paper we propose a novel, autonomic admission control scheme based on the increasingly popular Pre-Congestion Notification (PCN) framework put forward to the IETF [2].

PCN, which targets core/fixed network segments, defines a new traffic class that receives preferred treatment by PCN-enabled nodes, similar to the expedited and assured forwarding per-hop behaviours in Differentiated Services [3]. PCN assumes that some end-to-end signalling protocol (e.g. RSVP or SIP) requests admission for a new flow and supports two distinct mechanisms; admission control (AC) and flow termination (FT). AC is a control function that decides on whether new flow requests should be admitted or rejected based on the current network conditions. FT is a control function that tears down already admitted flows in case of overload that can occur, in spite of AC, due to rerouted traffic in case of link failures and other unexpected events.

Our approach employs fuzzy logic [4] and reinforcement learning [5] in order to improve the original PCN AC mechanism. The aim is to induce autonomic behaviour in order to minimize the need for manual reconfigurations, as well as to enhance the robustness and adaptability of the mechanism to changing flow and network conditions.

The rest of this paper is organised as follows; in Section 2, first, we briefly present the concepts behind the original PCN approach and its limitations. Then we present our approach, how fuzzy logic and reinforcement learning are introduced into the scheme and what purpose they serve, and we also present an example case study to better illustrate these notions. Finally, in Section 3 we present the potential impact of our work and our conclusion.

2. Technical Approach

PCN introduces an admissible and a supportable rate threshold (AR(l), SR(l)) for each link l of the network, which create three different load regimes. If the PCN rate r(l) is below AR(l), there is no pre-congestion and further flows can be admitted. If the PCN traffic rate r(l) is above AR(l), the link is AR-pre-congested and no further flows should be admitted depending also on how much the rate exceeds AR(l). If the PCN rate r(l) is above SR(l), the link is SR-pre-congested and in this state some existing flows should also be terminated, depending on how much the rate exceeds SR(l).

Both the AC and FT mechanisms are triggered based on packet markings. That is, PCN nodes mark traffic accordingly...
depending on whether it exceeds $AR(l)$ or $SR(l)$. The egress nodes evaluate the packet markings and their essence is reported back to the AC and FT entities of the network so that they can take the appropriate action. This process is summarized in Fig. 1.

![Fig. 1. PCN rates and behaviour.](image)

The main issue with PCN is that even though the possible marking behaviours and the possible AC and FT mechanisms are described in detail [2, 3], the way to actually set the thresholds so as to achieve the desired QoS targets has not been addressed. This means that even if only the AC mechanism is considered, for a single path consisting of ten links there are ten distinct threshold values that need to be manually adjusted so that the combined marking behaviour along all these links, when used in the AC mechanism, guarantees the desired QoS targets. When network characteristics such as link capacities and/or flow characteristics change, these thresholds have to be manually readjusted. Another issue is that the original PCN makes the hard assumption that all flows require zero packet loss; therefore it does not account for flows (such as VoIP) that can tolerate a small amount of packet loss, meaning that in such cases the original PCN scheme will be overly conservative.

In our approach we aim to overcome these limitations by introducing rules to autonomically drive the threshold reconfigurations, by using fuzzy logic to enhance the robustness of such rule-based reasoners, and, finally, by applying reinforcement learning (Q-learning) to dynamically and “on the fly” redefine the rules in cases that the current rule-set fails to meet the designated QoS targets. The approach is fully distributed with each router employing a fuzzy Q-learning controller driving its packet marking behaviour.

To better illustrate these concepts we will use as an example case study the scenario where a certain Packet Loss Rate (PLR) target needs to be met through the PCN AC mechanism. Since, as it can be easily shown, PLR parameters are additive, given the overall PLR requirement of the PCN-controlled traffic, the target PLR to be supported by each link can be easily derived. For this case, since the marking behaviour needs to convey information about the PLR at each considered link, the rules that will drive the $AR(l)$ threshold reconfiguration need to reflect the current PLR at the link and also how this PLR changes (increases or decreases). For ease of presentation we will call the latter as the “trend of PLR”.

The first step towards employing a fuzzy reasoner is to define the Membership Functions (MFs) for the input and output variables. The input variables in this case are two; the PLR and the trend of PLR. The output variable is 1; the threshold adjustment. These are illustrated in Fig. 2.

![Fig. 2. Membership functions for inputs and output.](image)

The second step is to define the rules that define the association between input MFs and the output MF; in principle the “condition-action” clause. In this case, since there are two input variables, with each one taking three different “fuzzified” values, in total nine rules will compose the rule-set. These rules reflect the human reasoning. For example a rule can be “if PLR is LOW and trend of PLR is NEGATIVE then threshold adjustment is POSITIVE”. This rule states that if we are in a low PLR region compared to the link target PLR and the PLR is further decreasing, we can increase the $AR(l)$ threshold so that more flows are admitted.

Reinforcement learning can be introduced as a final step to further improve the robustness and adaptability of this approach. Its role is to cope with the fact that the initially defined rules or MFs may not be set in an optimal manner. To account for this, the example rule presented above becomes “if PLR is LOW and trend of PLR is NEGATIVE then threshold adjustment is: {POSITIVE (Qpos), NEUTRAL (Qneu), NEGATIVE (Qneg)} with the action with the highest Q-value being triggered each time the rule is invoked. Depending on how well or poorly a given action performs, its Q-value is further increased or decreased meaning that the reasoner is able to continuously redefine the rules on its own and adapt to changes in network and flow characteristics as well as to changes to the desired QoS targets.

3. Summary of the work, potential impact and conclusion

This approach, once fine-tuned, will allow for fully autonomous PCN AC mechanisms. Based on the QoS targets set by a network operator, these mechanisms will be able not to only adapt the $AR(l)$ thresholds accordingly to meet these QoS targets but to actually change their behaviour, continuously learning from the outcomes of the previous threshold reconfiguration actions.

References

Abstract

Autonomic Management is a hot topic in designing future networks architectures. The management is accomplished by participation of knowledge, governance and network components in a control loop. Stability is one of the important concerns in control loops, therefore a formal method is proposed for evaluation of the correctness of components’ design and stability of control loops within the self-organised network architecture. A typical control loop is modelled and an example rule is defined and the model is examined against the rule.

Key words: Network Architectures, Knowledge and Governance, Control Loop, Modelling, Acme

1. Introduction

Future networks are proposed to maintain capabilities such as Knowledge-based Governance of the network. Based on the knowledge, actions are taken to govern the network. Evaluation of the new status of the network updates the Knowledge and the loop becomes closed. There are many of these Control Loops in future networks that should be governed in a self-organised way. One of the important aspects of self-organised networks is stability; in any system, loops can be a potential source of instability. The interactions between different components of a loop should be carefully designed to avoid instabilities. Modelling the loop can help in detecting potential sources of instabilities. The modelling at this level is carried out by describing the architecture of the system by a proper modelling language. In this paper, we employ Architecture Description Languages (ADLs) [1] to enable formalization and verification of the architecture of systems and execution of preliminary analysis on them, aiming at identification and resolution of design problems in the early stages of development. In this work, Acme ADL [2], the language used for modelling the architecture in Acme Studio, is employed to validate a scenario modelling a typical control loop. The modelling in this way can help rectifying initial design problems at the early stages and is beneficial for the industries to avoid (at the early stages) costly corrections on the design at the development stage.

2. Technical Approach

To model the control loop of the future network architecture, Acme [2] was used to define the principles and properties of the control loop in terms of the components, properties and rules supported by this formal description language. In order to familiarise the terminology used in Acme Studio, an introduction to some of the terms used in Acme is provided below:

- **Families** are classes (in OOP terms) of objects that can specify objects having common properties.
- **Elements** are instances of the families (Object in OOP terms) defined in a particular scenario.
- **Components** represent the primary computational units of a system.
- **Connectors** represent and mediate interactions between components.
- **Ports** represent components’ external interfaces.
- **Roles** represent connectors’ external interfaces.
- **Properties** are annotations (data fields in OOP terms) that store additional information about elements (components, connectors, ports, roles, and systems).
- **Rules** specify constraints, and contextual cues to assist architects with the design and analysis of software architectures. This includes the specification of the rule itself, the policy for dealing with violations of the rule, and the scope over which the rule is enforced.

The first step in modelling the architecture was to define the Acme families. Different rules and properties were defined later, based on the definition of the families.

Figure 1 depicts the information flow between the Knowledge, Governance and network resources. From the figure, it can be seen that the information between different components is mainly directed from the Knowledge- to the Governance- / the Governance- to the network resources and the network resources to the Knowledge- components. (The exception is the signalling messages directed in the opposite direction, shown with black thin arrows).
In this model, SSP (Service Sharing Point) is of type Port. To recognise the type of the port, a property is defined to specify the direction of the port. The property direction is an enumeration of the SSP port direction (GatewayDirection), and, for example, can be of the values input or output.

The middle part of Figure 2 graphically depicts the model of the control loop shown in Figure 1, in which three components interact with each other.

By expanding this simple scenario, many other rules can be defined to check the compatibility of the architecture with design rules, including the rules for stability. However, this solution cannot be used for performance analysis, scalability evaluation and network simulations.

4. Summary of the work, potential impact & Conclusion

In this section, it was shown how some of the formalities for describing control loop architecture, that were expressed in text were interpreted into a proper predicate language in a formal format. This was done to show that a network architect could verify validity of a scenario based on the components of proposed network architecture, and check the fulfillment of different principles and properties in terms of the rules of the particular formal language. The formal language used here was Acme and the language for checking the rules was Armani predicate language.

The research on this topic has been carried out under FP7-4WARD project and reflected in [3].

5. Key References


(http://www.cs.cmu.edu/~acme/AcmeStudio/index.html).

Asynchronous Multi-channel MAC Protocol for VANETs

Chong Han
Year 3 PhD Student, C.C.S.R., e-mail: chong.han@surrey.ac.uk

Abstract

Large scale vehicular networks in terms of large number of vehicles are common in urban area especially in a traffic jam or in the coverage of a multi-layer traffic architecture. This paper presents a new multi-channel MAC protocol for VANETs, namely, Asynchronous Multi-Rendezvous Multi-Channel MAC (AMCMAC) to improve network performance in large scale networks. The AMCMAC supports simultaneous multi-rendezvous on different service channels (SCHs), as well as, allowing other nodes to make rendezvous with their provider/receiver or broadcast emergency messages on the Control channel (CCH); and mitigates both multichannel hidden terminal and missing receiver problems. Comparison with other multi-channel MACs shows AMCMAC outperforms the IEEE 1609.4 and AMCP.

Key words: Vehicular Ad Hoc Network (VANET), Analytical model, IEEE 802.11p, Multi-channel MAC.

1. Introduction

Vehicular Ad hoc Networks enable communications among nearby vehicles and between vehicles and nearby fixed infrastructures. These communication facilities are expected to be used for a variety of applications to improve safety of the future transport systems and provide many industrial and entertainment services. VANETs enable vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications using wireless local area network (WLAN) technologies. However, as a single channel MAC protocol, the original IEEE 802.11p standard [4] based MAC sub-layer demonstrates a poor performance in dense VANETs, in provisioning QoS for different categories of applications [1]. This seems to be a particularly important shortcoming for reliable deployment of safety related applications. Thus, multi-channel extension of the IEEE 802.11p standard, namely, the IEEE 1609.4 [2] has been proposed to help improve the service differentiation capability of the 802.11p standard. In a multi-channel system, both non-safety and safety related applications could be provided on different channels. This could help improve QoS support for different application types by allocating them to different channels.

Although the IEEE 1609.4 standard helps provide a better QoS support, its MAC sub-layer suffers from a few shortcomings: 1) Synchronous MAC operation results in several problems. Strict time synchronization is demanded by each node. In addition, channel switching latency is not as low as assumptive values. Hence, the guard intervals have to be set larger than the assumptions, which further wastes the channel resources. 2) Allocation of fixed time intervals to the CCH and SCH can result in poor resource utilisation. 3) The lack of the agreement on the start time of transmissions results in possibilities of collisions on service channels. 4) Imbalance use of service channels leads to unsuccessful transmissions and high collision probabilities on some service channels. 5) Imbalance allocation of SCHs to transmit data stay with CCH, and listen to the CCH if the service channels is always taken into account.

The paper presents a new multi-channel MAC protocol for VANETs, namely, Asynchronous Multi-Rendezvous Multi-Channel MAC (AMCMAC). The AMCMAC supports simultaneous multi-rendezvous on different SCHs, as well as, allowing other nodes to make rendezvous with their provider/receiver or broadcast emergency messages on the CCH. We compare the performance of the proposed protocol with that of IEEE 1609.4 and AMCP, in terms of network throughput, channel occupation time, and the penetration rate of successfully broadcast emergency messages.

2. Asynchronous Multi-Rendezvous Multi-Channel MAC

AMCMAC protocol aims to overcome the weaknesses in existing multi-channel protocols such as time synchronization difficulty, missing receiver problem, and multi-channel hidden terminal problem. Meantime, the aim is to improve the performance of multi-channel operation.

Fig. 1 shows the basic channel access mechanism of the proposed MAC protocol. All the nodes which are not using SCHs to transmit data stay with CCH, and listen to the CCH if
they do not have packets to transmit in their queues. Once a node needs to broadcast/unicast a packet, it competes with other stations for the access of the CCH. If the channel is sensed idle, after a backoff time the sender broadcasts an RTS packet through CCH. Inside the RTS packet, a list of available SCHs for the sender is added. The destination node obtains the list from the RTS and checks the local SCHs entry table, which records the status of all the SCHs and the duration it will be busy for. If the destination finds an available channel which also is available for the sender, it replies a CTS after a period of SIFS with the channel information that they will use for the transmission. After a successful handshake, the two nodes will switch to the SCH they agree to finish the data transmission. Transmissions can simultaneously proceed on different service channels, as a result more transmission could be accomplished simultaneously in the network.

The destination node decides which SCH will be used; hence, no more renegotiation is needed, unlike AMCP. Due to the fast movement of the nodes in VANETs, the shorter a handshake takes, the higher probability of successful transmission will be achieved. In a multi-channel MAC, in some occasions, all the available SCHs for the sender may seem unavailable for the receiver node. Then, the receiver node replies a CTS that indicates CCH as the channel to be used. This allows the nodes to occasionally use the CCH for data communications.

The AMCMAC aims to solve the multi-channel hidden problem: when nodes switch to an SCH, first, they listen to the service channel for a short period of time before transmission. If the channel is not idle, i.e., another pair is currently using that particular channel, the nodes record that SCH as busy for a certain duration and return to the CCH to renegotiate for another available SCH. This measure helps to mitigate multi-channel hidden terminal problem.

A shorter NAV timer is adopted to reduce the channel access time due to the unsuccessful handshakes. Once receiving an RTS packet, nodes on CCH (excluding the sender and destination nodes) set a timeout period, TimeoutCTS, given by (1).

\[
\text{Timeout}_{\text{CTS}} = \delta + \text{SIFS} + \delta + (\text{Node ID}) \mod 31. \quad (1)
\]

After this short timeout, if the receiver does not send a CTS message, the other nodes could start their own backoff procedures, immediately. A random duration, the 4th term in the right side of (1), is introduced to avoid RTS collisions on the CCH.

3. Simulation Results

We evaluate AMCMAC using the well-known simulation tool, NS-2. The simulation scenario considers a single hop reference area that comprises a straight road of 10 km without Road Side Units. A “moderate” scenario is used to evaluate the network performance, which shows a situation where AC1, AC2 and AC3 flows are saturated but the interval of AC0 flow is 1s. This scenario is practically useful as emergency messages do not frequently occur in the network. The rest of the major simulation parameters are chosen from the IEEE 1609.4 standard.

4. Summary of the work, potential impact & Conclusion

The proposed AMCMAC protocol outperforms the IEEE 1609.4 and AMCP scheme in terms of system throughput, channel occupation time and the penetration rate of successfully broadcast emergency messages. AMCMAC is more suitable for large scale or heavy loaded vehicular networks.

Key References


Abstract
The work described herein seeks to develop sensing and emulation hardware for use in the Internet of Things and Persuasive Technologies experimentation. This hardware will be utilised to ascertain a user’s energy consumption patterns whilst also providing a context of how that energy is being used. Once inefficient energy usage patterns have been identified, persuasive technologies may be employed to affect a change in the user’s behaviour. The proposed hardware platform should in itself be energy efficient whilst also providing a useful degree of resolution of the user’s local environment so that the correct energy usage context can be ascertained.

Key words: Internet of Things, Wireless Sensor Networks, Persuasive Technologies

1. Introduction
The Internet of Things (IoT) has the ability to completely transform the world as we now know it. The amount of information that one could glean from having the IoT in place is almost unimaginable. Practically every item that one would come into contact with in his or her daily lives would not only be uniquely identifiable, but it would also allow one to search for and possibly interact with and obtain information on a myriad of objects throughout the world. Such technology, however, does not come without its costs. One cost in particular is the energy consumption required to interact with all of these objects.

Research into electronic components that use ever decreasing amounts of energy or that use energy more efficiently has certainly assisted in improving the outlook on the energy consumption required to have truly ubiquitous IoT. However it may not be enough considering the astonishing number of items that would require some degree of electronic intelligence. Thus it may be necessary to take a more proactive approach by persuading users of IoT technology to be more efficient in their energy consumption. This concept is the basis for the proposed hardware platform currently being developed.

A Persuasive Energy Node (PEN) is currently being developed which will not only capture a user’s energy usage patterns, but it will also seek to create a context of how the user is utilising that energy. For instance, if the user has left his lights, computer and desk fan on but is not his desk, the energy the user is consuming is essentially wasted. The user can then be notified of his behaviour (e.g. he is sent an automated text message stating that he left his computer on), or, if the users allows it, the system that is monitoring the situation can turn the items off in his stead. The methods with which to persuade users to modify their behaviour is one of the key points of REDUCE, an EPSRC funded project which will directly utilise the PEN’s1.

2. Technical Approach
As shown in Figure 1, the envisioned PEN consists of two parts, the energy monitoring unit (Plogg) and the sensor unit which is connected to the top of the Plogg. The PEN has been designed so that the sensor unit can be detached from the Plogg so as to allow for the optimal placement of the sensor unit without having to be constrained by the location of nearby power points. The detached sensor unit would then be connected through the use of a 9-pin serial umbilical cable.

![Figure 1. The Persuasive Energy Node (PEN)](image)

The Plogg has the ability to measure the electrical characteristics (e.g. energy, power, RMS voltage, RMS current, phase, etc.) of any electrical appliance that is plugged into the front of it. It also has built in actuators so as to allow for the possibility of turning off and on whatever is plugged into it.

The second part of the PEN is the sensor unit and contains five sensors for detecting light levels, temperature, motion, noise and vibration. The unit also has an LED lamp installed on the top of the unit which will be used to alert a nearby user

1 FP7 projects such as IoT-A (http://www.iot-a.eu/public) and SmartSantander (http://www.smartsantander.eu/) may also utilize these devices in their research.
that the sensors are actively acquiring data. The five sensors were carefully chosen to provide the greatest amount of information resolution whilst also being cost effective.

The light sensor has a spectral range of 330-720 nm with a peak in its spectral response at 580 nm. This sensor will be used primarily for determining if any room lighting is being used at the time energy usage patterns are being generated. As these sensors detect all forms of visible light, some calibration may be necessary to distinguish room light from natural light.

The temperature sensor used has a functional range of -40 to +125°C with a scale factor of 10mV/°C. It has an accuracy of ±2°C and a ±0.5°C typical linearity over the functional range. This particular device was chosen because of its cost. There are more accurate devices on the market but they are considerably more expensive and hence not very practical for large-scale deployments.

The motion sensor is a Passive InfraRed (PIR) device. The sensor is capable of detecting lateral movements of approximately 20cm, within a distance of 2m from the front of the sensor unit across (over a 110° horizontal arc across the front of the sensor). The sole purpose of this sensor is to determine the presence of an individual in the proximity of where energy consumption patterns are being generated.

Because it is possible that the motion sensors can give a false positive sensor event (e.g. someone walks by the desk of a user who is not there) a noise sensing circuit has been incorporated into the sensor unit. It is likely that this sensor may at best be interrogated every few seconds. Thus an electronic circuit was developed so that when a noise is detected, the circuit creates a voltage signal whose amplitude is proportional to the volume of that noise. The circuit would then slowly decay (decay time of about ten seconds) so that any noise made in the area, no matter how short the duration, can still be measured by the sensor unit.

A vibration sensor was installed to aid in ensuring the validity of the acquired data. It is plausible that under certain experimental conditions a user may attempt to falsify their context data by covering or moving the sensor unit. Thus if the sensor unit is tampered with, a notification will be sent to the network administrator alerting him to the fact that the sensor unit has moved. A command can then be sent by the administrator to reset the sensor once the sensor unit’s status has been investigated.

Housed within the Plogg unit are several more circuit boards. The first of these is a TelosB mote which provides wireless network capability (see Figure 2) and is responsible for interrogating the various sensors. An optocoupler circuit (also in Figure 2) has been added on the communications bus between the Plogg electronics and the TelosB board in order to protect each board from potentially damaging voltage spikes that may arise on the data bus.

The last circuit board consists of a 4-port USB hub and a Peripheral Interface Controller (PIC). The USB hub has been added to aid in the communications between the TelosB and PIC and the rest of the gateways used in the proposed network. The primary function of the PIC is to provide a means for creating a virtual sensor by using its Pulse Width Modulation (PWM) output in combination with a simple RC filter. The PIC’s PWM can then be programmed with a desired response so that this virtual sensor in effect emulates the response of a real one. This will aid in the characterisation of the network by providing each PEN with a predetermined sensor event.

The PIC is also used to interrogate a high-side current shunt monitor circuit. This circuit provides a means with which to measure the electrical current passing through the PEN whilst it is in different states of functionality (e.g. acquiring sensor data, transmitting data, etc.). It is also used in resetting the vibration sensor circuit.

3. Key Results and Discussion

The PEN is currently in its final stages of development with some of the circuit boards being mass produced. Thus a minimal amount of key results have been obtained thus far. Early versions of the PEN and prototypes of the circuits therein have been used in demonstrating the proof-of-concept of the PEN. They have also been used characterise the amount of electrical current required for the unit. The sensor unit draws approximately 20 mA at 3.3VDC thereby giving a typical power consumption of approximately 66 mW. The 4- port USB hub, PIC and supporting electronics, TelosB and optocoupler circuits utilise approximately 60-80mA of current (depending on whether the mote is transmitting data wirelessly) at 3.3 to 5VDC thus requiring approximately 300-400 mW of power to function.

4. Summary of the work, potential impact & Conclusion

It is expected that the PEN will have far-reaching applications. In both the REDUCE project and other European projects as well. Work is currently underway to rollout approximately 150 of these units in the CCSR. Future, large-scale deployments are being planned as well where these units will be installed in other university buildings and student accommodation sites. Whilst the unit currently being developed as part of the REDUCE project, such devices can be used in nearly all aspects of daily life so that wasteful energy usage can be minimised as much as possible thereby reducing the impact of ubiquitous technologies such as the IoT.
Efficient Pairing non-interactive Key Distribution Protocol for the Internet of Things

Dan He

Abstract—Key distribution is a challenge in the Internet of Things (IoT) due to the limited resource of networked devices. Traditional approaches relied on public key cryptography that is not always practical in some networks such as wireless ad hoc networks without fixed infrastructure. Existing pairing based key distribution protocols are neither efficient nor formally secure in the sense. This paper proposes a light-weight pairing based non-interactive key distribution protocol with provable security, particularly suitable for wireless sensor network applications.

I. INTRODUCTION

Key distribution protocols have faced two challenges in the Internet of Things (IoT). The networked device restricts the use of advanced cryptographic algorithms [1] while network topology diversity and bandwidth limit further make public key infrastructure (PKI) inapplicable. PKI-based key exchange requires several round of message exchanges before security association established. This wastes bandwidth and consume the node energy. Non-interactive key distribution protocols (NIKDP) [2] have been studied as alternative way to creating a shared security association between peers. In this model, a trust third party called private key generator (PKG), hold a master key and will issue private keys to their legitimate holders. Therefore, two peers can communicate with a PKG asynchronously. The identity-based cryptography (IBC) was first proposed by Shamir [3] while the drawback of IBC requires a trust authority (PKG) in charge of escrowing privacy key’s generation. With bootstrapping process in wireless sensor networks (WSN), a node’s private key can be preloaded into nodes before deployment stage. This loses the dynamic key distribution property. Recently, the implementation of Oliveira et [4] work has shown progress in practical Tate Pairing key distribution scheme but it remains non-interactive key distribution untouched with requiring identity certificate. This paper extends the efficiency of (NIKDP) by pairing cryptography without loss of security features.

II. PROBLEM STATEMENT AND PAIRINGS

Reducing the key escrowing and weaken a trust authority are key issues in IBC-based key distribution protocol. Pairings provides secure primitives based on the hardness of Bilinear Diffie-Hellman Problem (BDHP).

Let \( G_1 \) and \( G_2 \) be a cyclic additive group of prime order \( q \) and \( G_T \) is a multiplicative group of the same order \( q \). A bilinear pairing \( e : G_1 \times G_2 \rightarrow G_T \) is a mapping satisfying:

1. Bilinearity: \( e(aP, Q) = e(P, Q)^a \), where \( P \in G_1, Q \in G_2, a \in Z_q \)
2. Non-degeneracy: \( e(P, P) \neq 1_{G_1}, e(Q, Q) \neq 1_{G_2} \)
3. Computability: existing an efficient algorithm to compute \( e(P, Q) \) where \( P \in G_1, Q \in G_2 \)

Definition GBDHP A generalized Bilinear Diffie-Hellman Problem (GBDHP) defines: Given \( (P, Q, aP, bQ, cP, cQ) \), compute \( e(P, Q)^{abc} \). An efficient algorithm A is said to \( (t, \epsilon) \) -solve GBDHP in \( (G_1, G_2, G_T, e) \) within time slot \( t \) then the problem under random oracle can be solved with probability at least \( \epsilon \). i.e.

\[ Pr\{ (A(P, Q, aP, bP, cP, cQ)) = e(P, Q)^{abc} \} \geq \epsilon. \] (1)

Definition CDH (Computational Diffie-Hellman (CDH) problem) Let \( P \in G_1 \); Given \( (P, aP, bP) \) with uniformly random choices of \( a, b \in Z_q^{∗} \), compute \( abP \in G_1 \).

III. PAIRING NON-INTERACTIVE KEY DISTRIBUTION PROTOCOL

Here we use type 1 pairings [1], our protocol contains the following procedure:

A. Setup

Let \( H : \{0, 1\}^{∗} \rightarrow G_1 \) be cryptographic has functions. Public known parameters are \( (G_1, G_2, G_T, q, e, H) \)

B. Private key distribution

PKG randomly chooses a master-key \( s \in Z_q^{∗} \). It issues node’s private key \( S_a = sH(ID_A) \) to each sensor node (BS can be the PKG), and sends it to user A via secure channel.

C. Shared Secret Establishment

Let users A and B wish to create secret key. A can compute B’s public key based on B’s ID: \( P_B = H(ID_B) \) and B can compute A’s public key based on A’s ID: \( P_A = H(ID_A) \).

Then A and B can compute their shared key by:

\[ K_{ab} = e(S_a, P_B) \] (2)

\[ K_{ba} = e(P_A, S_b) \] (3)

It is easy to verify based on e’s bilinearity. This protocol is simplified version of Dupont without loss security property.

IV. SECURITY PROOF

A. Attacker models

Two kind of attackers are assumed.

- Type1 Adversary: Adversary A cannot access the master key, but A can replace the public keys of any entity with a value of his choice, since there is no certificate involved.
- Type2 Adversary: Adversary A has access to the master key, but cannot replace any user’s public key.
1) Extraction queries: A can extract node $ID_i$ to the challenger, who replies the private key $sH(ID_i)$ back to A.

2) Guess: Adversary A can collect information, and picks any two identities $ID_A$ and $ID_B$, publishes a quadruple $(ID_A, ID_B, \zeta, \eta)$.

The adversary A’s advantage can be defined as:

$$Adv(A) = Pr\{e(H(ID_A), H(ID_B))^x = \zeta\} + Pr\{e(H(ID_B), H(ID_A))^x = \eta\}$$

B. Security Proof

Proposition 4.1: [2] if the GBDHP in setting $(G_1, G_2, G_T, q, e)$ can be $(t, \epsilon)$-solved, then key distribution protocol in the setting $(G_1, G_2, G_T, q, e, H)$ can be $(t + \delta, \epsilon)$-broken, where $\delta$ is the computation time to hash function.

Theorem 4.2: With hash function H, the adversary A who breaks the protocol with parameters $(G_1, G_2, G_T, q, e, H)$ with query upper bound $q_u$, then a simulator S can have $(t', \epsilon/e^2)$-solves the GBDHP for $(G_1, G_2, G_T, q, e)$, where $e$ is the time consumed in scalar multiplication in $G_1, G_2, G_T$ plus looking for queries order list $q_u$.

Proof: The simulator S takes random uniform distribution to $(P, Q, P_a, Q_b, P_c, Q_c)$. To find the solution $e(P, Q)^{abc}$ in type II model, the simulator S controls hash function H. The query probability of S can try up to $q_u$ times of private keys of A. Each S query for hash value of some bit string X, S will build up an list L started with empty and look for existing entry in L and store missing tuple $(X, R, h, u)$. for missing entry in L, B can do

- randomly choose a $h \in [1, q - 1]$
- select $u \in \{0, 1\}$, assign probability $Pr(u = 0) = \theta$. if $u = 0$, $R = hP$, otherwise sets $R = hP_a$.
- stores $(X, h, R, u)$ to L.

Now the simulator S will deal with the answer from random oracle A with string ID.

- query H and record the tuple $(ID, R, h, u)$
- verify if $u = 1$, report failure.
- compute $hPc$ and send it to random oracle A.

Once receiving the guess $(ID_A, ID_B, \zeta, \eta)$ from A, the simulator S performs

- record the tuple $(ID_i, R_i, h_i, u_i)$ where $i \in \{A, B\}$ in L
- randomly select a $t \in \{0, 1\}$, if $t = 0$, and $u_a = 1$, the record a guess with probability $\zeta^\frac{1}{\eta}$, otherwise failure
- if $t = 1$, and $u_b = 1$, the output probability of a guess is $\eta^\frac{1}{\zeta}$

the simulator will try and the output probability

$$\theta = e(H(ID_A), H(ID_B))^c$$

$$\theta^{1/h_a} = e(ah_A P, bH_b Q)^c^{1/h_a} = e(P, Q)^{abc}$$

when $t = 1$, the analysis is the same and the equal probability to verify A and B that the simulator can correctly guess the private key is $Adv(A)/2$ when it doesn’t fail. Due to the upper bound queries $q_u$, if each query the non-abortion guess probability is $\mu$, the simulator has the overall probability of non-abandon during guess is $\mu^2(1 - \mu)$. Minimizing this function, we found the optimal solution for $\mu = 1$ so the non-abortion probability at least $\frac{1}{2}$, therefore, the simulator S correctly solve the GBDHP instance is at least $\theta/e$. Compared with Dupont’s protocol, we found every hash function increased in the protocol, the simulator S will decrease the correct solution probability by the order of 2.

V. IMPLEMENTATION AND EVALUATION

This protocol has been implemented in stanford PBC-based library. We have evaluated the execution time cost on Intel M620 2.67GHz CPU 4GB ram laptop as showing below Table I:

To be secure, generic discrete log algorithms must be infeasible in groups of order $r$, and finite field discrete log algorithms must be infeasible in finite fields of order $q^2$. The experiments are performed 10 times for different system parameters in pairing cryptography. $ts$ is execution time on tested laptop. From the results, we found our key distribution protocol is extremely efficient although we haven’t implemented it on 8-bit processor as reported in [4].

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>EXECUTION TIME (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>320</td>
</tr>
<tr>
<td>$q$</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>512</td>
</tr>
<tr>
<td></td>
<td>1024</td>
</tr>
<tr>
<td>$ts$</td>
<td>7.45</td>
</tr>
<tr>
<td></td>
<td>18.1</td>
</tr>
<tr>
<td></td>
<td>60.7</td>
</tr>
<tr>
<td></td>
<td>292</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

We have shown an efficient pairing based non-interactive key distribution protocol proposed for the IoT. In this work, the security in WSN can be implemented through non-interactive key distribution protocol that can save bandwidth. Our experiments confirm the pairing cryptography efficiency and secure. The future work could be implementation of this protocol on real WSN testbed.

REFERENCES


Resource-Friendly Authentication Scheme for Disruption Tolerant Networks

Enyenihi Johnson

Year 4 PhD Student, C.C.S.R., e-mail: E.Johnson@surrey.ac.uk

Note: A submission exceeding two pages will immediately be returned to the author for Revision and Resubmission. Please include only Key References (≤ 5)

Abstract

The abnormal link condition in disruption tolerant networks requires an authentication scheme that facilitates offline processing and does not depend on round trip for security. The frequent disruption in these networks is at conflict with internet assumption of constant connectivity. This accounts for the poor behaviour of internet protocols and inapplicability of existing authentication schemes. In this paper, we proposed an authentication scheme that employs Delay/Disruption Tolerant Networking (DTN) concept to address authentication issues in these networks while taking into consideration the capabilities of different network entities. Its comparison with existing schemes is made and its comparative advantages established.

Key words: Network, Security, Authentication, Resources.

1. Introduction

Disruption tolerant networks are networks with intermittent connectivity and high packet loss rate caused by high node mobility, short range of radio frequency, low battery power and no line of sight amongst others. Existing authentication schemes perform poorly in these networks due to their interactive and resource exhaustive nature, design principle or operational complexity. DTN [1] an overlay network designed with operating assumptions to accommodate these limitations requires a new concept in authentication.

In this paper we propose an authentication scheme that employs DTN concept of custody transfer and hop-by-hop communication. It is designed for a structured wireless communication environment with a common overlay platform (DTN) where communication is facilitated through satellite. The networking environment comprises localised nodes (with higher capability) and mobile nodes (with relative capability) conveying messages from one localised node to another. Our proposed scheme is discussed, its comparison with existence schemes made and its comparative advantage established. The industrial value of the research includes deployment of authentication schemes that are resource efficient and affordable.

2. Technical Approach

The three-party authentication model was found suitable because of its scalability in large network and the presence of trusted third party to establish trust between two communicating parties with different network characteristics. The need for strong pre-authentication is also required to facilitate distributed authentication needed in disruptive environment. While the Kerberos protocol [2] concept of three servers and a client station was considered applicable, the Kerberos protocol itself is not implementable because it was designed for a connected environment and also involves much keying materials with increase in network size. Our proposed authentication scheme is shown below.

![Fig. 1 Proposed Authentication Scheme](image)

Registration: Once Prior to Network Membership

1. Alice → RA: Pb_RA{devIDA, rtj DTN network}
2a. RA → Alice: Pb_A{devIDA, secInfoA, IDNA, secInfoNA, PbNA}
2b. RA → DTNNA: PbNA{nfa, devIDA, secInfoA}

Authentication: Once Per Network Membership

3. Alice → DTNNA: PbNA{rtj, devIDA, secInfoA, PbA}
4. DTNNA → Alice: PbA{auConf, IDNA, secInfoNA, Kdtn, APassA}

Data Exchange: Anytime while Member of the Network

5. Alice → Bob: {PbB{Hello} | {APassA, NA} | {Tstmp, IDA, IDB}}·hmac
5b. Alice → Bob: [{KR(Bulk message), PbB(KR)} | {Tstmp, IDA, IDB} | {APassA, NA}}·hmac
RA and DTNNA denote Registration Authority and DTN Network Administrator; devID_A denotes device ID of Alice; ID_A, ID_B and ID_DTNNA are the network identifiers of Alice, Bob and DTNNA; secInfo_A and secInfo_DTNNA the secret information of Alice and DTN Network Administrator; PbRA_A, PbNA_A, PbB_A and PbB_B the public keys of RA, DTNNA, Alice and Bob; APass_A and APass_B the authentication mechanisms (proposed in [3]) of Alice and Bob; nfa, rtj and auConf denotes notification for authentication, request to join and authentication confirmation respectively; Kdm the network-wide shared secret key, Kg the randomly generated symmetric key, Tstamp the timestamp, N_A the random nonce value generated by Alice. In the case bulk encryption message 5b is used instead of 5.

The scheme involves three phases of registration, authentication and data exchange where Registration Authority (RA) provides pre-authentication service, network administrator (DTNNA) provides actual authentication into the network while Bob (recipient) the last server carries out administrator (DTNNA) provides actual authentication into authentication and data exchange where Registration device for DTN network with initial key pair and Pb RA while during the data exchange phase. We assumed a customised authentication phases at different times to communicate respectively; Kdm the network-wide shared secret key, Kg the randomly generated symmetric key, Tstamp the timestamp, N_A the random nonce value generated by Alice. In the case bulk encryption message 5b is used instead of 5.

The scheme involves three phases of registration, authentication and data exchange where Registration Authority (RA) provides pre-authentication service, network administrator (DTNNA) provides actual authentication into the network while Bob (recipient) the last server carries out the action requested by Alice (sender). Alice and Bob are network entities that must pass through the registration and authentication phases at different times to communicate during the data exchange phase. We assumed: a customised device for DTN network with initial key pair and PbRA_A while devID_A and PbA was made available to RA by the vendor. Each of the entities generates their public/private keys; RA generates the secret information (secInfo) needed for authentication phase and DTNNA generates the credentials needed for data exchange phase. We assumed that the source node is in custody of the database of all authentic members prior to sending a bundle in the data exchange phase.

The RA and DTNNA in our scheme are similar to the Authentication Server (AS) and Ticket Granting Server (TGS) in Kerberos. Communication between the client and AS/TGS in Kerberos is symmetric based while communication between the client (Alice) and RA/DTNNA in our proposed scheme is asymmetric based. This takes place once at different times prior to the data exchange phase while DTNNA communicates occasionally with entities after authentication in the form of updates. The frequency of operation makes it suitable for relatively computationally weak mobile nodes. Our preference for asymmetric cryptography is to facilitate secured online exchange of security credentials. The data exchange phase is both symmetric and asymmetric based. This is designed to accommodate gateways and mobile nodes since communication here will be frequent.

Our scheme uses Hash Message Authentication Code (HMAC) and asymmetric mechanism for source authentication. HMAC is computed using Kdm (change during updates by DTNNA) and secInfo_DTNNA by all nodes including the intermediate node. A message is authenticated when the computed hmac matches appended hmac. The security issues associated with this concept is taken care of by the asymmetric mechanism. APass was proposed in [3] because of the heavy weight nature of digital certificate and its associated issues. It is designed to expire and does not need revocation list storage. The custody transfer together with location and knowledge based routing implemented in our simulation tool accommodates the high packet loss.

3. Key Results and Discussion

Our proposed authentication scheme will be validated using simulation and compare with some existing works like [2], the public key version in [4] and a traditional cryptography based protocol in [5]. A simulator is being developed in C++ using Microsoft Visual Studio 2008 with security provided using Crypto++ Cryptographic Library. The framework is modelled to work well in both connected and disrupted environment. The identified protocols will be modelled and validated using our framework to analyse their behaviour in both connected and disrupted environment.

We assumed: a connected environment with one intermediate node between the source and destination; the use of certificate in [4] issued by a trusted third party; acknowledgement of all received messages; the entities have been authenticated by the third party to communicate in the network and the encrypted message is not bulk. The comparative analysis of our proposed protocol with existing protocols [2], [4] and [5] using some resource related metrics is shown in fig 1 below:

<table>
<thead>
<tr>
<th>Analysis Metrics</th>
<th>Authentication Protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[2]</td>
</tr>
<tr>
<td></td>
<td>Our Protocol</td>
</tr>
<tr>
<td>Total number of exchanges</td>
<td>8</td>
</tr>
<tr>
<td>Number of contacts with the server by intermediate and destination nodes</td>
<td>1</td>
</tr>
<tr>
<td>Number of cryptographic operations by the server during communication</td>
<td>10</td>
</tr>
<tr>
<td>Number of cryptographic operations by intermediate node</td>
<td>6</td>
</tr>
<tr>
<td>Number of keys used</td>
<td>7</td>
</tr>
</tbody>
</table>

Fig. 1 Comparative analysis with existing protocols

4. Summary of the work, potential impact & Conclusion

A traditional cryptography based authentication scheme has been proposed to address authentication issues in delay tolerant networks. From the comparative analysis, our protocol is found to be more resource efficient than the referenced existing protocols.

Key References (≤ 5)


Performance study of a Green Light Optimized Speed Advisory (GLOSA) Application Using an Integrated Cooperative ITS Simulation Platform

Konstantinos Katsaros

Year 1 PhD Student, C.C.S.R., e-mail: K.Katsaros@surrey.ac.uk

Abstract

This paper proposes a Green Light Optimized Speed Advisory (GLOSA) application implementation in a typical reference area, and presents the results of its performance analysis using an integrated cooperative ITS simulation platform. Our interest was to monitor the impacts of GLOSA on fuel and traffic efficiency by introducing metrics for average fuel consumption and average stop time behind a traffic light, respectively. The simulations are varied for different penetration rates of GLOSA-equipped vehicles and traffic density and indicate that GLOSA systems could improve fuel consumption and reduce traffic congestion in junctions.

Key words: vehicular communications, traffic light advisory, fuel consumption, traffic congestion.

1. Introduction

Advances in wireless communications and in particular vehicular communications have led to the advent of cooperative Intelligent Transportation Systems (ITS). These systems employ vehicular communication technologies such as 802.11p to enable deployment of applications that could potentially improve road safety, traffic efficiency, and introduce new entertainment and business applications [1]. Exploitation of ITS for traffic congestion control in urban areas as well as fuel consumption reduction are among the most promising applications according to transportation authorities [2]. This can be achieved by vehicle-to-vehicle (V2V) or infrastructure-to-vehicle (I2V) communications, intelligently advising individual drivers about traffic events, such as the traffic light phases and beyond.

The potential of V2V communication for improving fuel efficiency has been demonstrated in [3], showing that vehicular communication can assist to reduce average fuel consumption especially under high traffic density and long traffic light cycles. Other projects have investigated the impacts on fuel efficiency when using V2V or I2V communications employing different algorithms to smoothly slow down at a red traffic light or to reach it at the next green phase. Depending on the used consumption models, different results can be observed. One important aspect is the dependency of the results on different penetration rates of the communication enabled vehicles.

This paper aims to present the implementation of a GLOSA system to reduce traffic congestion by decreasing the average stop time at traffic lights while reducing fuel consumption and CO₂ emissions. The GLOSA application provides the advantage of timely and accurate information about traffic lights cycles and traffic lights position information through infrastructure-to-vehicle (I2V) communication, and advises drivers with an optimized speed guiding them with a more constant speed and with less stopping time through traffic lights. The main challenges in achieving this include the modelling of the vehicle traffic, the communications between traffic lights and vehicles and finally the driver’s behaviour. Individual research has been performed for each one of these areas, but complete simulations by taking into account the dynamics of all parameters are scarce. Thus, for our implementation of GLOSA, we used an integrated simulation tool based on the Fraunhofer VSimRTI, which enables online two-way coupling of different simulators (communications and vehicle traffic) for monitoring the influence of GLOSA application on the traffic and fuel consumption. Our results show reduction up to 7% in average fuel consumption and up to 89% in average stop time.

2. Technical Approach

We designed and tested GLOSA in an integrated simulation platform. The first step was to define a reference scenario which is depicted in Figure 1. In this reference area we placed a traffic light. Vehicles will enter following a common arrival process such as Poisson distribution. In the scenario the equipped vehicles rate is varied, since we want to investigate the influence of the penetration rate and find out which is the minimum percentage of equipped vehicles for GLOSA to have a positive impact on the traffic efficiency and fuel consumption. Road traffic will be modelled using the microscopic Stefan Krauss (SK) model, a car-following model with two basic rules. First, vehicles in free motion have a target speed and try to cruise at it. Second, when a vehicle senses the distance to the vehicle ahead to be less than a certain threshold, it slows down keeping a safe distance. The speed of the vehicles is limited within a certain range \([V_{\text{min}}, V_{\text{max}}]\) where \(V_{\text{min}}\) is the minimum speed that vehicles can cruise without causing further traffic congestion and \(V_{\text{max}}\) is the maximum speed that is forced by the speed limit of the road. Acceleration is also bounded and asymmetric - higher deceleration than acceleration - for more realistic simulations. The SK model is integrated in SUMO traffic simulator that we used in our work.
The GLOSA algorithm has been implemented to support the aforementioned simulation approach and is presented in Figure 2. First, vehicles enter the reference area according to the Poisson distribution as mentioned before. The RSU (Road Side Unit) attached to a traffic light broadcasts periodically CAM’s (Co-operative Awareness Messages) including the position, timing information and additional data for the traffic light. When OBU receives a CAM, it runs the GLOSA algorithm and provides an advisory speed to the driver.

```
1: Find the most relevant traffic light (closer on route)
2: Calculate Distance and Time to Traffic Light (TTL)
3: Check phase at (now + TTL)
4: if GREEN then
5: Continue Trip
6: Target Speed (Ut) = Umax
7: else if RED then
8: Calculate remaining Red Time (Tred)
9: Calculate target speed for (now + Tred + TTL) = Ut
10: else if YELLOW then
11: Calculate remaining Yellow Time (Tyellow)
12: Check for possible acceleration
13: Calculate target speed for (now + Tyellow + Tred + TTL) = Ut
14: end if
15: Advisory speed = MAX (Ut, Umin) & MIN (Ut, Umax)
```

Figure 2 GLOSA Algorithm

3. Key Results and Discussion

For the evaluation of the GLOSA application a series of simulations were conducted. We measured the influence that GLOSA penetration rate has on the two performance metrics (fuel consumption in Figure 3 and stop time in Figure 4) and how the non-equipped vehicles are affected. The simulations were conducted in a high traffic density environment. The most interesting outcome from these figures is that even the non-equipped vehicles are affected in a beneficial way from the GLOSA equipped vehicles and this is due to the SK model. They follow the leading vehicle which - if equipped with GLOSA - forces them to adjust their speeds accordingly since we assume that there is no overtaking in our simulations. The second notice is that the average stop time is reduced even when the penetration rate is small, but in order to see positive results in fuel efficiency we need at least 50% equipped vehicles. The observed average maximum reduction in fuel consumption is 7% which is slightly higher than the average maximum fuel savings in [5] for their scaled up scenario.

4. Summary of the work, potential impact & Conclusion

The results suggest that the GLOSA application has a positive effect on both performance metrics. Thus, such an application can be incorporated into future vehicles to reduce their carbon footprint by reducing fuel consumption and accommodate the drivers with a smoother trip. For a more extended version of the paper see [6].

Key References

Abstract

This article describes a service front-ends concept in form of eXtended User Interfaces (xUIs) that evolved from research on intelligent user interface adaptation in ubiquitous multi-device environments. The general goal of xUIs is to intelligently incorporate real-world objects, such as digital appliances, networks of displays, user interaction devices, etc. for ubiquitous services delivery.

Real-world objects are bound to a particular service during the adaptation process, directly by the user, automatically by the system on behalf of the user, or by service definition. Since this is a conceptual work, the article introduces the xUIs concept, discusses the related research challenges and provides links to selected research outcomes.

Key words: Multimodal User Interfaces, Context-Awareness, Internet of Things, Ubiquitous Computing

1. Introduction

Traditionally research on user interaction is facilitated in the area of human computer interaction (HCI). Multimodal user interfaces, as part of this research area, offers the possibility to “... process two or more combined user input modes – such as speech, pen, touch, manual gestures, gaze, and head and body movements – in a coordinated manner with multimedia system output.” [1]. This allows users to interact more natural with computer systems and applications as they are used to in everyday human-human interactions.

With the advent of ubiquitous/pervasive computing the meaning of user interfaces widens. According to M. Weiser [2], the several interaction devices in the user’s vicinity, such as tabs, pad, and board sized surfaces, constructed for a particular computing task should be smartly integrated for a combined and consistent user experience. More recent work, such as presented by Costa et. al. [3] supports the earlier statements, and provides a good overview on ubiquitous computing challenges including the notion of transparent user interaction in order to minimise integration efforts.

The introduction of such technologies at current is mainly pushed forward by multimedia systems (e.g., UPnP and DLNA in home and mobile devices), games systems (e.g., Nintendo Wii, Sony Playstation Move and Microsoft Kinect) and partly mobile applications. Specialized systems, such as the multi-room music system SONOS, also generate an integrated user experience addressing music consumers.

Most modern services, such as social networks, media portals and online shopping, are accessible via a Web browser on a personal device (e.g., PC, laptop, pad, mobile terminal).

From the paradigm of accessing services from a single point of entry or specialized systems it should be possible in the future that services are completely ‘surrounding’ the user, using context information seamlessly through sensors as deployed by the Internet of Things and present their service offerings in the best possible presentation in the current environment to the user. Further users should be supported in retrieving relevant information of their environment and in navigating through it. This will break the boundaries of services only existing through the web or on mobile terminals, but rather allowing them to exist in the real-world. This paradigm shift demands for new service front-end systems that are flexible and intelligent enough to follow the user and always support the user with the best device configuration for their subscribed real-world services.

Following this introduction the article discusses xUIs as the conceptual basis for service front-end systems in Section 2, provides current key results and discussions in Section 3 and gives a summary with outlook for future work in Section 4.

2. xUIs concept

xUIs as envisioned in this article are distributed user interfaces consisting of a set of real-world objects, which have been bound to a particular service, directly by the user, automatically by the system on behalf of the user, or by the service definition (provider). xUIs may include objects based on traditional Web browsers, home services (IPTV, telephony), digital appliances, gesture-based devices, (networks of) displays, and so on. Multiple users can share a xUI bound to one particular service and at the same time multiple services can operate xUIs in the same environment and even the same real-world object but using different semantics. In order to assure the complex situation based user interface adaptation for the xUIs, front-end systems have to be designed and developed, which allow users and applications to easily and flexibly “compose” a xUI for a particular service and to dynamically recompose it when necessary. The front-end system visualizes how users can relate devices and objects to a particular xUI and in what way they can use them for that particular service (semantics). The front-end system personalises this information based on learning mechanisms, for instance based on the interaction and rendering devices someone has used for a particular xUI in the past. During runtime of a service, the front-end system assures continuous multimodal interaction and service delivery with session management functionality. This takes into account the changes in the environment of the user through object event handling and user activities tracking and evaluation (e.g.
commands or user movements). According user interface description languages matching the requirements imposed by xUIs have to be defined. An illustration of the xUIs and related concepts is depicted in Fig. 1.

In order for a xUI to function an underlying layer that allows the consistent incorporation of heterogeneous devices and network protocols has to be defined. With the current developments in the Internet of Things towards interoperable solutions for sensors and devices, such an underlay could be provided. Reasoning about optimal device configuration for a particular service is supported by the front-end systems.

3. Current Key Results and Discussion

The presented xUIs concept is an evolution from previous research work on mobile multimodal user interfaces extending the way users can interact and communicate with and through real-world objects that are connected to the Internet of Things. Several aspects have been investigated in earlier demonstrators with a research target on mobile multimedia applications. The topics researched in this respect are explained in the following.

**User mobility:** relates to personalisation of a multimedia application based on user context (i.e., location). Users experience dynamic adaptation behaviour suitable to their situation (e.g., on foot, in car, at home). Mechanisms included automatic media transformation (i.e., in a car situation news text is played as audio after text-to-speech transformation) and session continuity (i.e., a video played on mobile device is continued on larger display, once the display was in the vicinity of the user). User interaction has been extended to available devices (e.g., steering wheel buttons in the car in order to control scrolling through news items on mobile device).

**Complex multimedia presentation delivery:** The distribution of multimedia content (i.e., a video presentation that includes additional scene information in form of text and pictures) across available devices has been demonstrated. Main focus was given to the design of a Context-Aware adaptation process and quality matching of devices capabilities to multimedia content capabilities, as described in [4].

**User Interface personalisation through learning:** An association rule mining approach was investigated, in order to design a recommendation mechanism that takes into account device usage during several device configurations. This is an approach described in [5].

4. Summary of the work, potential impact & Conclusion

This paper presented the concept of eXtended User Interfaces (xUIs) in the context of ubiquitous environments and summarised related research demonstrators. Through future research and an implementation for service front-ends based on the xUIs concept, user interaction through customized services with real-world objects could be enabled. New applications are envisioned and existing ones would benefit through better integration with the users’ everyday surroundings.

**Key References**


Assessing and improving an approach to delay-tolerant networking

Lloyd Wood

Research Fellow, Centre for Communication Systems Research at the University of Surrey, e-mail: L.Wood@surrey.ac.uk

Abstract

Delay-tolerant networking (DTN) is a term invented to describe and encompass all types of long-delay, disconnected, disrupted or intermittently-connected networks. This term is also used to refer to the Bundle Protocol, which has been proposed as a unifying solution to address disparate DTN networking scenarios. We have evaluated the Bundle Protocol as part of our work conducted to test the Bundle Protocol in a space environment. We have found architectural weaknesses in the Bundle Protocol that may prevent engineering deployment of this protocol in realistic delay-tolerant networking scenarios, and have proposed approaches to address these weaknesses.

Key words: delay-tolerant networking, DTN, Bundle Protocol, end-to-end principle.

1. Introduction

Work on delay-tolerant networking attempts to address many difficult networking problems, including mobility, disruption, delay, naming, routing, and more. The Bundle Protocol is a proposed approach to these problems.

Delay-tolerant networking was originally proposed as a generalisation of NASA-led work to move to packetized networking for its spacecraft. This is called the ‘Interplanetary Internet,’ where long propagation delays between spacecraft scheduled contacts dominate communications. Extending the scope of the problem space to also include addressing very different disrupted terrestrial ad-hoc networks, including military networks, significantly increased interest in and funding for this new approach to networking, and has led to the development of the Bundle Protocol suite. The Bundle Protocol attempts to encompass a large number of environments and use cases, even though individual scenarios can be quite different [figure 1].

2. Technical Approach

The Bundle Protocol is intended to embody a new architectural approach to networking. It is not by itself directly compatible with other networking protocols such as the Internet Protocol suite, and cannot be as it attempts to introduce new approaches to identification, addressing and routing. However, the Bundle Protocol can be layered over these other networks using ‘convergence layer adapters’ [figure 2]. Given the prevalence of IP networking, most Bundle Protocol development has been with convergence layer adapters for the existing TCP/IP suite, although there has been some work over other protocol suites, including CCSDS for space agencies and AX.25 for ham radio use.

The Bundle Protocol specifies a new way of transmitting data in a complex protocol format that is assembled from different blocks for different purposes. Blocks and header information can be inserted, removed and modified by intermediate nodes.

Emphasis on security has been a focus of the design of the Bundle Protocol from an early stage, with a complex security architecture that provides authentication of messages and encryption of data delivered. This is a deliberate change from earlier architectures; security was famously deliberately left out of the Internet’s TCP/IP suite, and had to be retrofitted later with IPsec, HAIPE, SSL and other protocols. However, this focus on security has neglected protocol reliability.

Transmissions and memory storage do not always produce perfect copies (although we may wish to believe so) and do have non-zero error rates. Any introduced errors must be detected with deliberate checks. A good network protocol will sanity-check its headers to make sure that the information it is exchanging was received reliably before being processed. It may also sanity-check its payload data, but that is best left up to the highest networking layer, in accordance with the well-known end-to-end principle [1].

In practice, responsibility for end-to-end reliability usually falls to the protocol providing end-to-end transport – here, the Bundle Protocol. When payload data is encrypted or authenticated, a reliability check comes as a side-effect of the security check.

Figure 1 – Comparing different DTN scenarios
However, the ability to permit block information to be deliberately altered en route in the Bundle Protocol, without security checks on that in-the-clear information, because any alteration would be viewed as an attack, weakens overall reliability. Errors can be introduced and can go undetected.

The Bundle Protocol also includes node-to-node authentication. This can provide a lower-level reliability check, again as a side-effect of a security mechanism. Alternatively, the Bundle Protocol can rely on the reliability checks in convergence layers and lower protocols. In either case, relying on layers underneath the Bundle Protocol to guarantee correctness of data sent by the Bundle Protocol is hoping for the best in violation of the end-to-end principle, and leads to a more complex protocol stack.

The approach adopted by the Bundle Protocol requires high-complexity, processor-intensive, security mechanisms to be implemented just to provide an approximation of functionality of a lightweight checksum, as a side-effect of the authentication and encryption that the security mechanisms provide. The threat model for the environment may not require the level of security offered by the security architecture, but will require reliability checks in accordance with the end-to-end principle. As a result, the security mechanisms are now required to be implemented to gain an assurance of reliability. This is an added cost to deploying the Bundle Protocol.

3. Results

Some known deployments of the Bundle Protocol have run without security being implemented. Two in-space tests of the protocol for the Interplanetary Internet – on the UK-DMC satellite and onboard the Deep Impact/EPOXI comet probe – chose not to implement bundle security [2]. Not doing so can be attributed to a number of different reasons, including reliance on lower layers for ‘probably good enough’ reliability, lack of security code readiness, lack of available memory to store and run code, lack of any threat to be worth mitigating against, and security not being required to be able to demonstrate the protocol running in space. The complexity of the Bundle Protocol is one argument against placing it in low-end embedded systems, and processing hardware for space is often low-end and unable to execute modern cryptographic algorithms rapidly. Non-essential processing is not done.

Recent EU trials in a remote area of Sweden also did not implement bundle security, and so are also running without high-level transport reliability checks as a side-effect of not having security [3]. The risks to data of doing so are well-known, and are described in the end-to-end literature [1].

The Bundle Protocol also presumes that all communicating nodes have a shared understanding of UTC time, with synchronised clocks. Bundles expire and are discarded after a set clock time. Bundles sent from nodes with misset or drifting clocks may be expired at the next node simply because their timestamp is in the far past or distant future. If you don’t know the time, you can’t ask for the time by using the Bundle Protocol. A bundle age extension block has since been proposed for when UTC is not available, but still presumes working clocks are available.

We completed the first in-space tests of the Bundle Protocol for the Interplanetary Internet on the UK-DMC satellite [2], and have combined this practical experience with theoretical analysis to provide a detailed consideration of technical aspects of the Bundle Protocol, which includes the reliability and timing issues highlighted here [4]. The design of the Bundle Protocol is such that we suggest adding support for lightweight reliability checking within the imposed limits of the existing security framework [5]. Unfortunately, that workaround still requires the security architecture to be implemented, and is unlikely to see widespread adoption.

4. Summary of the work, potential impact and conclusion

We have evaluated the Bundle Protocol, highlighted architectural problems in its design, and proposed workarounds where possible. Our work shows that the basic design of the Bundle Protocol neglects architectural issues. We expect this to limit its adoption and deployment.

References

1. Introduction

SaVi, the Satellite Visualization tool [1], is a computer program for visualizing and animating the movement of satellites and their coverage. SaVi was originally developed at the Geometry Center at the University of Minnesota, but became homeless when that research centre was closed due to lack of ongoing funding. Maintenance of the software was taken over by Lloyd Wood, who had found it useful during his doctoral work on satellite constellations. SaVi has been maintained at the University of Surrey since then.

2. Technical Approach

SaVi exists as a standalone program that can also be run as a ‘module’ that interfaces with and controls the Geomview program [2]. Geomview is a general-purpose rendering program useful to mathematicians; SaVi leverages Geomview for simple three-dimensional rendering and OpenGL texturemapping, while ignoring Geomview’s ability to render higher dimensions of interest to mathematicians.

SaVi is implemented as a satellite orbit simulator, written in ANSI C, which is driven by commands added to the higher-level Tool Command Language (Tcl). This two-pronged approach allows SaVi to be scriptable. Simple, short, Tcl scripts generating satellite constellations and driving the underlying simulator are written in a similar manner to the scripts of the network simulator ns-2, which also relies on Tcl. Many scripts simulating, illustrating and animating proposed and existing satellite constellations are included with SaVi.

SaVi’s user interface is presented in Tcl’s Toolkit, Tk, which complements Tcl and allows for relatively straightforward creation of a graphical dialog- and window-driven system [fig. 1]. Running a program to see and animate a complex satellite constellation is as simple as clicking the Constellations menu and selecting, say, the Iridium system [fig. 2]. Graphical output can be recorded and saved. Satellite and constellation properties can be edited.

As it relies only on Tcl/Tk and standard Unix POSIX libraries, SaVi is portable across a wide range of Unix-compatible systems, including Linux, FreeBSD, and Mac OS X.

SaVi can also be run under Microsoft Windows using Cygwin or a virtualisation environment such as VirtualBox.

As a popular community-driven effort, SaVi is in the top 1% of projects on the SourceForge site for open and free software. SaVi is also available as an installable Debian package for Ubuntu users.

SaVi’s portability and popularity is maintained by users reporting bugs and requesting features, or providing fixes for problems encountered with specific platforms or with new compilers, so that after over fifteen years of life, SaVi remains compatible with modern systems.

SaVi shows satellite coverage areas on a number of different map projections. A fisheye view of the sky is also available to examine how satellites pass over different points on the Earth. SaVi shows satellite coverage as either minimum elevation angle or as half-angle beamwidth, and indicates how that coverage moves over time. Multiple spotbeams on a satellite, communication channel properties, and precise orbital motion with complex precession are not yet simulated; the University has other, custom, tools for simulating these in far more detail.
3. Results

While SaVi lacks the large number of features present in far richer commercial offerings such as the Satellite Toolkit (STK), it has the advantage of being entirely free to use, which makes SaVi immediately attractive to the academic community. Its output has appeared in a number of research papers and articles [3].

SaVi has also been found useful for teaching purposes at the MSc level and on short courses at a number of institutions, when constellations and orbital movement of non-geostationary systems must be explained clearly. SaVi has been used in lectures at the University of Surrey, the International Space University, SUPAERO, and elsewhere.

SaVi enables quick and easy explanation of the features of satellite geometry. This can include showing the differences between rosette and star constellations, by contrasting the seamed Iridium and seamless Globalstar systems, comparing diversity, overlapping coverage, and the large number of satellites seen in the sky for navigation constellations, or demonstrating repeating-groundtrack designs such as Molnya [Fig. 3].

This educational focus has driven recent development, with attention given to built-in help explaining what each of the systems simulated is, and to providing pointers to further information for the autodidactic.

SaVi has also been picked up by industry companies designing satellite constellations for communication, and has been used publically to illustrate the designs of their systems [4, 5].

4. Summary of the work, potential impact and conclusion

Although SaVi only provides a relatively simple degree of satellite simulation functionality when compared to more full-featured commercial packages, its open codebase and contributions from around the world have led to a long-lived, robust, portable, cross-platform, tool that has attracted a wide degree of interest. SaVi appears to have gained a useful educational role in introducing and explaining the properties of satellite constellations.

References

Planning of IoT-Resource Usage upon Service Request
Stefan Meissner

Year 3 part time PhD Student, C.C.S.R., e-mail: s.meissner@surrey.ac.uk

Abstract

Assuming an Internet of Things in which billions of sensors are deployed it becomes a challenge for lookup systems to find services that are suitable to user’s requests. In this paper an approach is described that reduces the search space containing services provided by IoT resources based on type of information or by clustering services by location. It is assumed that IoT users are mostly interested in services related to the location the users are in at the time of service request. It will be shown that clustering by location can significantly reduce the search space for IoT related services.

Key words: Internet of Things, Service Discovery, Query Processing, Semantic Web

1. Introduction

Imagine in an Internet of Things (IoT) widely deployed sensors provide information about things of any interest for people. Things like the current weather parameters at a particular location could be an example for information of interest. The idea behind the IoT is that every user can retrieve information about real world phenomena over the internet in order to use that information for applications [1]. This work addresses the problem of finding the service in the IoT that is most suitable for the application needs of an arbitrary user. The service discovery problem in IoT emerges as the number of IoT devices providing services increases exponentially. It is assumed that IoT services are provided by a number of makers using heterogeneous technologies. Different technologies used for services required different solutions for service discovery before. To overcome the technology barrier in internet services the Semantic Web [2] has been introduced. The Semantic Web provides a common vocabulary about real world objects and phenomena that can be shared among service providers and service users. This common knowledge allows a match-making between semantically expressed service requests made by users and semantically annotated services offered by providers. Those service annotations are discoverable by users through a lookup query interface. SPARQL [3] is a semantic query language widely used for information retrieval in Semantic Web. Every user query leads to scanning of wide search space of service annotations. If the search criteria taken from user’s query match a service annotation this service is a candidate to be returned to the user. A query engine separates relevant from non-relevant information. Currently IoT services are deployed for test and research purposes mainly. When IoT enabled devices are deployed in an industrial scale the number of available services will increase by a serious amount and will significantly widen the search space for query engines. Response times for queries will be unacceptable by then. A sample scenario could be as follows: Let the number of available services may be 1000 in a test IoT environment covering a real world urban area. Let the number of relevant services being able to serve user’s need may be 10. Then the ratio between relevant and non-relevant services will be 10/1000 = 0.01. Assuming the user sticks to its need, but the number of available services is now 1 million. Then the ratio would be 10^-5. In other words 999,990 services are non-relevant to the user but they are still part of the search space during service lookup. This paper presents an approach to limit the search space so that the time for scanning the search space can be reduced significantly.

2. Technical Approach

In this work following approach for reducing the service lookup search space is used: the space is divided into clusters and only the relevant clusters are scanned. For clustering the search space several criteria can be used. The following section will describe two experimental calculations for determining a suitable criterion for clustering.

3. Key Results and Discussion

One obvious criterion is to organise the space by phenomenon measured by sensors. In previous research [4] ten different sensor types have been used to classify information IoT services can provide. Assuming the sensors are distributed equally by type each cluster would contain a tenth of the entire search space then. Given a total search space of 10^9 a cluster would contain 10^8 services to be scanned. Also given that the targeted number of relevant services to be returned to the user should not exceed 10 the ratio would be 10^-7. Physical phenomena are valid globally so clustering by physical phenomena would cover all services that are deployed globally. Scaling up the search space to a global magnitude a number of 10^9 services is realistic. With this assumption the ratio between relevant and non-relevant services would be 10^-7 which means around 10^-10 services are non-relevant to the user, but need to be scanned during service lookup. If sensor technology improves in the future and a higher variety of measurement services will become...
available the number of clusters needs to be extended accordingly. An optimistic calculation is to multiply the number of clusters by 10 so that the total number of global clusters would be 100. This number applied to the previous scenario would lead to a ratio of 10^-6 meaning there would be a number of 10^-10 non-relevant services. This is still a large search space which needs time to be scanned.

An alternative criterion for clustering could be the location of the sensor providing the service the user needs. Locations are measured by geo coordinates characterising points on the Earth’s surface. The total surface of planet Earth is 510.072 million km², only 21.9% of it is covered by land. It is possible to deploy sensors in oceans, but this paper focuses on land area only which is an area of 148.94 million km² [5]. Users are normally interested in location based information so that a query can be scoped to typical locations. Below a list of the biggest urban areas (by population=potential users) of each continent together with their area is given.

<table>
<thead>
<tr>
<th>Urban area</th>
<th>Population</th>
<th>Area in km²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokyo-Yokohama</td>
<td>36.7 million</td>
<td>9,065</td>
</tr>
<tr>
<td>New York City</td>
<td>20.7 million</td>
<td>11,264</td>
</tr>
<tr>
<td>Cairo</td>
<td>17.6 million</td>
<td>1,709</td>
</tr>
<tr>
<td>Moscow</td>
<td>13.7 million</td>
<td>4,533</td>
</tr>
<tr>
<td>Sydney</td>
<td>3.8 million</td>
<td>1,788</td>
</tr>
<tr>
<td>Average</td>
<td>18.5 million</td>
<td>5,676</td>
</tr>
</tbody>
</table>

According to previous calculation an average urban area (metropolitan area in this case) would cover 0.0038 percent of Earth’s landmass. Clustering the service lookup search space over areas of that size would lead to a number of 26,240 clusters all over the world. Assuming the IoT resources providing services are distributed equally over the clusters the search space would be divided into fragments of the size 3.8*10^-5 of the global search space. Based on the future vision of having a billion of services available such a fragment of the search space contains 38110 services to be scanned during lookup.

In Figure 1 a comparison is shown between clustering by type of sensors (or more general by observed phenomenon) and clustering by location. Clustering by type is shown with 10 clusters as well as 100 clusters as explained before. The horizontal axis shows the total number of available services the vertical axis is the estimated size of search space assuming services are distributed equally over all clusters. It is shown that clustering by location leads to the highest reduction of service lookup search space. The advantage of geographic clusters is more significant for a large amount of services deployed globally. Clustering by location is to be preferred for another reason. The Earth’s surface is a finite number so that the number of location clusters is fixed. The number of sensor or information types is likely to be increased along the progress of sensor technology. A simplification has been made in this work with averaging the location clusters of IoT resources providing services all over the world. Further research is necessary to estimate differences of service distribution in metropolitan areas and agricultural areas or other areas of low population density. Furthermore it needs to be investigated more how many services users request within the cluster they are situated in and how many services users request which are outside their cluster.

![Fig. 1 search spaces in relation to number of available services](image)

### 4. Summary of the work, potential impact & Conclusion

In this work it has been shown that lookup of IoT services can be made more efficient by clustering the search space used for service lookup. Experiments have shown that clustering over geographic location is suitable for globally deployed IoT services.

### Key References


Secure Network Entry Process for WIMAX Mesh

Shahab Mirzadeh

Research Fellow, C.C.S.R., e-mail: S.Mirzadeh@surrey.ac.uk

Abstract

The current proposals for WiMAX mesh initial network entry process, link establishment and traffic encryption keys (TEK) exchange mechanisms are not secured enough and attacks such as man-in-the-middle (MitM) and denial-of-service attack are still possible. In this paper, we propose a secure network entry process, link establishment and traffic encryption keys (TEK) exchange mechanism by using short term mesh certificate for the mesh routers and using them in authentication process and also in establishing pairwise keys.

Key words: WiMAX, Mesh Networking, Security, Key Management

1. Introduction

In order to join the WiMAX mesh network, a new node as a Candidate Node (CN) has to perform the network entry process. In network entry process, the CN chooses one of the active or operational nodes as Sponsor Node (SN), get synchronized with network, negotiates its capabilities, authorizes itself and registers itself as member of network. Unfortunately in the current standard [1] with lack of sponsor node’s authentication, this important stage is not well secured and the adversary can impersonate valid sponsor nodes, lunch different attacks such as topological and DoS attack, result in weaker security links, and disturb registration process. To address these problems, we propose a secure network entry process, link establishment and traffic encryption keys (TEK) exchange mechanism. In our proposal we considered following assumptions:

- Every subscriber station (SS), CN or SN, is equipped to a X.509 certificate which binds its identity to its public key. The X.509 certificates are issued and signed by manufacturer certification authority (CA).
- Every node is also equipped to the operator’s authorization server’s public key at the time of deployment. Operator authorization server (AS) acts as a trusted certificate authority for the mesh network and issues mesh certificates for the public keys of the mesh nodes.
- We are proposing using short term mesh certificates for the mesh routers (mesh BS) instead of long term X.509 certificates. The rationale behind this decision is to free mesh subscriber nodes from holding and checking the certificate revocation list (CRL). This practice does not jeopardise the protocol security as the mesh routers’ certificates are short term certificates and the probability of being compromised is low. It should be mentioned that the mesh routers can afford renewing their certificates frequently as they are static nodes and usually have an established connection with the authorization server.
- UDP tunnel between sponsor node and authorization server is assumed to be secured from eavesdropping, message replay and message modifications attacks.

2. Secure network entry process

Figure 1 shows our modification to network entry process. In channel opening stage, we propose that sponsor nodes sign their “mesh network entry open” messages (MSH-NCFG: NetEntryOpen) with their public keys certified by mesh certificates and attach their mesh certificates to the message. We chose not to sign the “MSH-NCFG: Network Descriptor” message to save mesh network’s traffic from sending regular mesh certificates and also to reduce candidate node’s overhead from verifying unnecessary signatures. Upon reception of “mesh network entry open” message, the candidate node verifies the sponsor node’s mesh certificate, extracts its public key and verifies its signature. If both verifications are successful, it assumes the sponsor node as a valid node and sends its acknowledgment message.

It should be mentioned that the candidate nodes still authenticate themselves to the sponsor node as defined in standard by sending the operator authentication value in “MSH-NENT: NetEntryRequest” message. Operator authentication value, given to the device at the time of deployment, is HMAC value of node’s MAC address and its serial number which is calculated with the authentication key (AK). The sponsor nodes, which got a copy of AK from the authorization server, can verify the CNs’ operator authentication values.

To thwart SN’s malicious behaviour in node authorization stage, we suggest that CN and AS sign the authorization request (PKM-REQ: Auth-REQ) and the authorization reply (PKM-RSP: Auth-RSP) respectively. This practice prevents the malicious SN from impersonating AS by fabricating valid authorization reply, commencing rollback attack by changing the CN’s security capabilities in authorization request, or launching DoS attack by changing the security parameters in authorization reply. The authorization reply message includes the CN’s mesh certificate. The AS issues mesh certificates on CNs’ public keys and basically binds them to the mesh network after verifying their X.509. Besides explained use of mesh certificates in SN authentication process of sponsor channel opening stage, they are also used in neighbor authentication and traffic encryption key (TEK) exchange.
In the current node registration process, the message authentication between the CN and registration server (RS) is based on derived keys from the authentication key (AK). As Jabri and Azzouz stated in [2], AK is not well protected in the current standard and malicious sponsor nodes can easily retrieve it. With having AK, a malicious SN can tamper with registration request/response or even impersonate registration server. To address this problem, we suggest that both CN and RS sign their messages (REG-REQ and REG-RSP) with their server. To address this problem, we suggest that both CN and RS sign their messages (REG-REQ and REG-RSP) with their server. To address this problem, we suggest that both CN and RS sign their messages (REG-REQ and REG-RSP) with their server.

3. Secure link establishment and traffic encryption key exchange

The main problem with current link establishment procedure and TEK negotiation mechanism is in using operator shared secret (OSS) key in authenticating the neighbouring nodes and also protecting the integrity of the exchanged messages. The OSS key is not well secured in the current standard, especially in the authorization stage and can be disclosed to the adversary. With knowing OSS key, the adversary can impersonate a valid mesh node in the neighbourhood of victim node and hijack its traffic or change TEK in “PKM-RSP: Key Reply” message and launch DoS attack. In addition, the proposed challenge-response protocol for the neighbour authentication is flawed and adversary can impersonate neighbouring nodes even without knowing the OSS key.

To address these vulnerabilities, we proposed to use mesh certificates in authenticating neighbouring nodes and establishing pairwise keys. Our protocol is different from Zhou and Fang proposal [3] in that it improves both security and efficiency. Firstly, our challenge-response mechanism does not suffer from the impersonation attack. Secondly, we do not need the extra initial message for transferring B’s public key for pairwise keys establishment. Figure 2 shows the proposed link establishment protocol. In this figure, two neighbors A and B authenticate each other based on their mesh certificates and establish their secure link key. The procedure is as follow:

- Node A sends its challenge to node B containing its mesh certificate, node B’s identity, and its signature on both frame numbers (A and B’s) and B’s identity. Node B’s frame number is the last known frame number of node B’s MSH-NCFG message. The Node A’s frame number is the frame number of A’s MSH-NCFG message that carries the challenge.

- Upon reception the challenge, node B checks its identity and verifies A’s mesh certificate and A’s signature. If verifications are successful, node B sends its challenge response containing its mesh certificate, node A’s identity, its randomly selected link identity, randomly selected link key K_AB encrypted with A’s public key, and its signature on both frame numbers, A’s identity, its link identity and encrypted link key.

- Upon reception the challenge response, node A similarly verifies it and if verification is successful, it sends a signed acceptance message with its randomly selected and unused link ID.

To exchange and update the traffic encryption key (TEK), we propose to use the established pairwise keys instead of OSS keys in authenticating key request and reply messages.

4. Conclusion

In this work we highlighted the vulnerabilities of current standard for WiMAX mesh and proposed some modification to its network entry process, link establishment and traffic encryption keys (TEK) exchange mechanisms. We proposed a novel technique of using short term mesh certificate for the mesh routers and used the mesh certificates to provide SN’s authentication to CNs and also in authenticating neighbouring nodes and establishing pairwise keys. We also exploited the public key techniques to augment the node authorization and registration process.

References


Towards Reliable Routing for 6LoWPAN Wireless Sensor Networks

Michele Nati

Research Fellow, Centre for Communication System Research, e-mail: M.Nati@surrey.ac.uk

Abstract

Over the past few years, due to the continuous evolution of Internet and the development of the Internet of Things paradigm, the original concept of Wireless Sensor Networks has started to change. Overcoming the view of an autonomous network, nowadays, the Wireless Sensor Networks are considered as active part of the Internet. This requires the implementation of IPv6 stack on every mote composing the network. Whilst support is present for applications, a routing protocol is needed. This paper introduces and shows the good performance of ROME, a routing protocol, explicitly designed in order to fulfill this gap.

Key words: Internet, 6LoWPAN, Routing, Wireless Sensor Network (WSNs).

1. Introduction

Since their first definition and introduction, a decade ago, Wireless Sensor Networks (WSNs) have been considered as autonomous networks, only occasionally connected and inter-operating with other networks. For this reason, most of the developed routing and application protocols are represented only by custom and non-standard solutions [1], able to only achieve some very specific collection tasks.

In order to export the collected data to the outside world, some times these autonomous islands have been also connected to a dedicated Gateway [2], in charge of forwarding the packets received from a given sink to a user resident on a different network. Conversely, in the last few years, following the continuous growth of Internet together with the development of the IPv6 protocol and the emerging need to connect to Internet the everyday objects (Internet of Things), the view of WSNs is currently changing. WSNs are now seen as a natural extension of the traditional Internet, where every sensor represents a single and distinct element, interacting with other objects and realizing a Machine-to-Machine communication based on protocol, infrastructure and architecture characterizing the Internet domain.

After initial scepticism about the possibility to implement on sensor mote complex protocols such as IPv6, nowadays a couple of different IPv6 stacks (6LoWPAN) have been implemented [3]. While these implementations provide protocols such as UDP, ICMPv6 and Neighbours Discovery to the WSNs world, the discussion about the routing protocol to implement and adopt as standard is still missing in order to complete the transition between standalone and IPv6 sensor networks. Quickly fulfilling this gap will then permit the industry to focus on the development of useful services, by exploiting a robust and reliable underlying standardized stack.

2. Technical Approach

ROME is a cross-layer approach combining an energy-efficient MAC layer with a greedy forwarding routing and endowed with a mechanism for escaping from connectivity holes and for dealing with node mobility and failures. Nodes operating ROME require position information, but this does not represent a constraint in the current IoT devices. It has also been shown that ROME is resilient to localization errors, so no precise and costly localization mechanisms are required.

The Data Collection protocol of ROME works as follows [5]. In order to save energy, nodes follow an awake/asleep schedules (duty cycles). A node does not know its neighbors and their duty cycles. When a node has a packet to transmit it initiates a contention phase, looking for a relay between its awake neighbors (eligible relays). The nodes do not need to know the position of any other sensors but their own and the sink since any other information is piggybacked using a RTS/CTS transmission scheme. The way in which a relay agent is selected involves the computation of two priority indexes. These two indexes measure the effectiveness of forwarding as perceived by the relay. The first index uses information about the local position of the relay agent with respect to the sink in order to select nodes, which minimize the distance to it. The second index, on other hand, takes into account parameters such as message buffer availability and burst size in order to capture how fast a packet could be further advanced if sent through that neighbour. The aim of finding a relay is to advance the packet toward the sink. Let $x$ be a node engaged in packet forwarding, and let $F(x)$ be the portion of node $x$ transmission area where relays offering positive advancement toward the sink are located. To advance the packet toward the sink node $x$ selects a relay among the awake neighbors in $F(x)$ (greedy forwarding). This advancement is always possible when there is at least a
neighbor closer to the sink. There are times, however, when this simple greedy mechanism fails because of dead ends, i.e., nodes that have no neighbors in the direction of the sink. To avoid dead ends the protocol uses a colouring scheme algorithm, assigning different colours to neighbouring nodes, in order to backtrack packets until there is a path made by nodes offering a positive advancement toward the sink. A node initially assigns itself a RED label (bad node) if it fails in forwarding a packet after a sufficiently high number of attempts. RED nodes refrain from accepting packets from nodes with better colour (lower colour id); conversely they start to send to the better coloured nodes the packets generated, in a direction that allow the packet to escape the dead end. This process can be then generalized for an arbitrary number of colours in order to provide a loop-free geographic routing of the packets towards the sink.

Moving mobile nodes. A node that is moving has no colour and does not reply to any RTS message. Whenever the node has to send a message it searches for a relay among those composing the static backbone of the network. A process for quickly updating its colour is then started when the node stops moving [5].

3. Key Results and Discussion

Extensive simulations have been performed ([5]) in order to show the good performance of ROME with respect to previously proposed approaches. Moreover, in order to assess the adaptability of the proposed approach as a routing protocol for the 6LoWPAN stack and its limited code and memory footprint (< 9KB of ROM), an implementation on real sensor-node has been provided. Experimental results in a forty TelosB nodes outdoor testbed are provided below. The nodes are deployed in an 80 × 200 meters grid. X-axis shows the IDs of the nodes in each column. Y-axis represents the position of each row of nodes. The sink is placed in the middle of the short edge of the topology deployment (X-axis in Figure 1 and 2). Each bar represents a node. Apart from the nodes near to the sink (represented with null bars), all the nodes generate a periodic report. The experiments have been performed for different duty cycle \(d\)=[0.05, 0.1, 0.3] and different traffic load \(l\)=[1pkt/s to 4pkt/s] that represent typical values for WSNs applications. Each experiment lasted for 1 hour. In this scenario the nodes are static. Figure 1 and 2 show the end-to-end latency (in seconds) of each source node for the scenario with duty-cycle \(d=0.3\) and traffic equal to \(l=1\)pkt/s. In the first case a simple greedy geo-forwarding is considered, while in the second, the ROME mechanism is implemented.

The results show how effective is ROME in recovering from network dynamics. Node 2 (orange bar) is connected to the others, but due to the harsh environment it cannot reach any nodes closer to the sink then itself. For this reason the greedy forwarding fails (latency bar is equal to zero, because no-one of its packets reaches the sink). Conversely, the colouring mechanism adopted by ROME can quickly adapt to this situation and find alternative paths towards the sink, modifying the colour of the node 2 and relaying the packets through its neighbours further from the sink. The latency values confirm that there is no degradation when the ROME approach is used. The slight increase in the end-to-end latency in Figure 2 is due only to the higher data traffic managed by the network that correctly sends also all the packets generated by node 2.

![Figure 1](image1.png)  
**Figure 1 E2E Latency – Greedy geo-forwarding**

![Figure 2](image2.png)  
**Figure 2 E2E Latency – ROME forwarding [5]**

4. Summary of the work, potential impact & Conclusion

ROME has been shown to be a very promising approach for routing in static and mobile WSNs with lossy link and network dynamicity. Moreover, due to its small code and memory footprint it can be easily integrated in the current 6LoWPAN stack. Therefore, in order to fulfill all the needs of an IPv6 network, also the ability for point-to-point communication needs to be supported. This will be part of future extensions of the protocol.

References

Abstract

The CCSR European Union and the research community as a whole are actively involved in various researches around the Future Internet theme. However, there is a demand within CCSR to converge all FI related outputs via a unified demonstration platform. The paper provides a high level overview on the requirements and approach aimed at building the CCSR FI platform.

Key words: future internet, Cloud computing, testbeds, management framework.

1. Introduction

As experimentally driven research plays a key role in the development of Future Internet (FI) technologies, the CCSR Future Internet Platform aims to consolidate all existing and future testbeds within CCSR into a single dynamically reconfigurable platform that can support the R&D of future Internet technologies.

The aim of this paper is not to describe each of the FI technologies in detail but to propose the requirements for developing a modular platform that is easily extensible to support the research of existing and future FI technologies. Most importantly, the FI platform aims to provide an API driven interface to external users to dynamically re-programme or reconfigure the FI platform via well-known EU FI federated testbed platforms such as PlanetLab and OneLab. Other research groups, SMEs and companies that are not connected to PlanetLab and OneLab [1] can also call these APIs over the Internet.

A high-level description of the platform structure is provided in Section 2 highlighting the key blocks required in achieving a modular FI platform. In Section 3, some technical requirements are proposed for connectivity, networking and application deployment within the FI platform. This describes key issues like interconnectivity between JANET/GEANT and the University networks while also looking at key protocols for reconfiguring the platform. Finally, in Section 4 some conclusions are provided.

2. Overview of the FI Platform

One of the key aims of the FI platform is modularity, as this will allow for the easier integration of new and future technologies with existing technologies. Also, it is important to have a structured way for both external and internal users for using the platform. Although there are far more requirements that can be considered, Figure 1 shows a proposed approach on how the key blocks can be connected in a modular way allowing for the addition of new blocks as needed. The JANET/GEANT connectivity block provides all the connectivity functionalities required as the University is already connected to the JANET/GEANT academic network. However, there is a need for some modification to current network topology, which will be described in more detail in Section 3. Most importantly there is also the need to connect securely to the University network infrastructure via wireless or wired connectivity in order to deploy various experimental devices all around the University campus. This will enable geographically related experiments in the areas of access and sensor networks technologies. More details on this are also provided in Section 3. For now, this is represented by the UniS network block in Figure 1.

![Figure 1: A High Level Overview of the FI Platform](image_url)
computing will be provided for the deployment of nodes and web services for SOAP and HTTP protocols to mention a few. As described earlier on, the main aim is modularity and new building blocks can be easily added at various level. In the next section, we provide more details on the technological requirements required.

3. Technical Approach

In this section various high level technical requirements are proposed in the area of the area most of the properties described in Sections 1 and 2. The technical requirements are subject to revisions based on additional information from various technical groups within CCSR and the University IT services. The proposals in this paper provide the foundational architecture that new ideas can be built upon.

3.1. Connectivity

It is important for the FI platform to be allocated its own subnet directly from the JANET/GEANT network. This will allow for unfiltered traffic destined for the FI platforms to be routed directly to the FI platform via the University router. This is of great importance because the FI platform may need to offer some experimental services that may be filtered out by the University as they could be seen security threats to the University. Also, one of the key criteria of joining most federated FI testbed is to have nodes in de-militarised zones (DMZs) where traffic are not filtered. The security of the FI platform will be handled by a firewall, which will be managed only by CCSR in order to provide flexibility and easy configuration.

The most critical issue with this proposal is the deployment of experimental devices around the University campus. Even though the FI platform is proposed to be an isolated network from the University networks, there is a need to provide a secured connection between the FI platform and the University network so that experimental devices can reuse existing wired and wireless infrastructure already deployed and managed by the University. There are currently various discussions with the University to address this issue.

3.2. Topology

A proposed approach towards implementing the FI platform infrastructure is shown in Figure 3. At the moment the platform below provides all necessary resources required to run most network and service based experiments.

The core routers (CR1, 2, 3, 4) and switches (SW1, 2, 3, 4) are connected with redundant links to allow for any topology to be easily deployed without any human intervention. This approach will allow for automated scripts to automatically configure various scenarios using standardized protocols with no intervention of a technician. The router and switches are currently separated, as this is a logical diagram. When deployed, we could either have four Layer 3 switches with each providing both routing and switching functionalities or a single high end switch that allows the virtualisation of any many routers or switches as needed.

The network environment is connected to the clouding platform that allows the dynamic deployment and management of virtual servers and clients on the fly. This will allow for the automated creation, powering and management of virtual servers and clients. Ready-made templates of servers and clients will be made available to reduce the deployment times.

4. Conclusion

The Future Internet platform will provide new opportunities for CCSR and the University as whole as it will a the first research driven future internet platform that allows for third party researchers, developers, and companies to carry out experiments and trials using real environments within a diverse community such as the University Campus.

Most importantly, the resources needed to develop the platform are already available within CCSR. However, there needs to awareness that all outputs and outcomes within all CCSR projects should be driven towards the FI platform.

5. References


Figure 2: FI Network Topology

For example, there could be templates for preconfigured IMS servers based on either Ericsson Services Delivery Platform [2] or FOKUS IMS platform. As most of these services are open source, there will be no issues deploying them in our environment. Also as, virtual servers have the ability to use as many physical resources and network interfaces as possible, they can also be used on the deployment of virtual routers and switches to test new network protocols and services.
HIP based Mobility Management Scheme for 3GPP Evolved Packet System
Meng Wang

Year 6 PhD Student, C.C.S.R., e-mail: M.Wang@surrey.ac.uk

Abstract
The Third Generation Partnership Project (3GPP) Evolved Packet System (EPS) adapts Proxy Mobile IPv6 (PMIPv6) and GPRS Tunnelling Protocol (GTP) based mobility management schemes to support the seamless mobility in the core network. However, the hierarchical management character of PMIPv6 and GTP is not the best solution for the flat IP architecture. In this paper, a Host Identity Protocol (HIP) based mobility management is proposed to perform a distributed mobility management; a context caching based handover optimization is also proposed to utilise the X2 interface of EPS to reduce the signalling exchanges among the core network nodes. Result shows the signalling load is reduced more than 22.6% and handover latency is reduced more than 27.9% comparing with the original 3GPP design.

Key words: Handover Cost, Mobility Management, 3GPP SAE, Host Identity Protocol.

1. Introduction
With the exponential growth of mobile network users and the rapid introduction of new services from the popularisation of iPhone, android and other smart phones, mobility management in the 3GPP EPS is expected to become more and more complex due to the composite signalling exchange processes related to different layers. Based on the new architecture and requirements for mobility management of the 3GPP standardisation, two different mobility management schemes have been proposed in 3GPP SA working groups, namely GTP plus GTP based mobility management scheme [1] and GTP plus PMIPv6 based mobility management scheme [2]. Both kinds of schemes follow a hierarchical architecture from the legacy 3GPP system and based on the GTP. This results in large signalling overheads, handover latency and restricts the scalability of the network.

The 3GPP EPS introduced the flat IP architecture to simplify the network management and packet overhead, however its 3GPP proposed GTP based IP based mobility management schemes relies on the hierarchical architecture, and this would not take full advantage to increase the scalability of the network. On the other hand, there’s an increasing growth of User Equipment (UE) to support multi-air interfaces, e.g. Universal Mobile Telecommunications System (UMTS), Long Term Evolution (LTE), Wireless LAN (WLAN), however the current EPS mobility management schemes do not support multi-homing. Furthermore, peer-to-peer is becoming one of the hot applications in mobile networks; however current EPS mobility management protocols do not support it. HIP [3] developed by Internet Engineering Task Force (IETF)’s HIP working group, integrates mobility, security, multi-homing and multi-access, and also supports peer-to-peer networking. However, as HIP is primarily designed to target global communication changes and hence for mobility macro-mobility management (UE moving across domains), in a micro-mobility scenario (UE moving within a single domain) it generates unnecessary signalling message between itself and the Corresponding Node (CN). HIP was hence not chosen as a candidate by 3GPP for mobile networking as based on the architecture most of the UE handovers are intra-domain.

In this paper, HIP is adapted to propose a HIP based micro mobility management scheme for the 3GPP EPS architecture. A “mobility context cache” based mechanism is also proposed to further optimise the handover performance. The performance evaluation is based on OPNET simulator. The evaluation results show the reduction of handover signalling load and overall delay as it is crucial to the mobile network’s scalability and the UE experience during the active session.

2. Technical Approach
The proposed HIP based mobility management scheme for 3GPP EPS includes three parts:

(1) HIP based Distributed Mobility Management Architecture
Based on the S1-flex interface of the EPS architecture, we propose a HIP based Distributed Mobility Management Architecture which in shown in Figure 1. In this architecture, for the level between the PDN Gateway (PDN GW) and the Serving Gateway (Serving GW) it still follow a hierarchical management; but the level between Serving GWs and eNBs, multiple Serving GWs and Mobility Management Entities (MME) form different pool area to perform distributed mobility management for the evolved NodeBs (eNB) and attached UEs in the corresponding eNB pool area. The HIP is used for mobility management and Host Identity Layer is added between the network and the transport layer to map the IP addresses and the Host Identity Tags (HITs). We propose to use the UE’s International Mobile Subscriber Identity (IMSI) as its Host Identity since the IMSI is a unique global number associated with the 3GPP network mobile phone users.

(2) Proposed HIP based Mobility Management Scheme
When moving across the EPS, a UE may trigger different levels of handover, as shown in Figure 1. If the handover process involves the Serving GW changes across the Serving GW pools, the UE moves between local administrative domain and triggers macro mobility management, where the HIP Update process is executed between the UE and the CN
and the security association between UE and CN is updated with the rekeying process. When a UE moving within an eNB pool shown in Figure 1, which triggers the micro mobility management. The UE stays with the same Serving GW; it keeps its present IP address, so there is no need for the HIP update process. However, the MME needs to update the forwarding table of the Serving GW to the target eNB to continue the ongoing session.

![Fig. 1 HIP based Distributed Mobility Management Architecture](image)

(3) Context Caching based handover Optimisation
Propose the eNBs cache the mobility management related context from MMEs, and transfer the mobility context from source eNB to target eNB on X2 interface during handover preparation phase, to enable eNBs to be able to directly send the control signalling to Serving GW for user plane update without going through MMEs. This mechanism can reduce the signalling exchanges during handover completion and also reduce the handover interruption time.

3. Key Results and Discussion
In order to simulate two mobility management schemes in the EPS architecture and investigate their performance, an EPS simulator is developed by using OPNET modeller 14.5. The simulator is designed to follow the EPS topology and built up a network scenario that consists total of 63 UEs, 63 eNB Cells, 9 Serving GWs, 3 MMEs, 1 PDN GW and 1 HSS; all the nodes are distributed in a geographical area with the cell radius set to 5km. The simulator simulates the UE movement across the network by random walk model and the UE speed is configurable from 0 Km/h to 350 Km/h. The simulation duration is configured be 2 hours. In order to analysis the handover delay, the delay parameter assumption is made based on the 3GPP TR25.912 [4].

Figure 2 shows the total network side signalling load comparison which is the total signalling of eNBs, MMEs, Serving GWs and PDN GW. . The enhanced HIP scheme has 28.5% less signalling than the GTP plus GTP scheme and 22.6% less than the GTP plus PMIP scheme. The optimized enhanced HIP scheme has 1.0% less total network side signalling compared with the original enhanced HIP scheme. The proposed schemes all reduced signalling load compared with original schemes, therefore offer better scalability to the network.

Figure 3 shows the break down of total handover signalling delay. It can be discovered that the optimised schemes shift the delay during handover completion to post-handover process. Therefore, the handover interruption (consists of handover execution and handover completion) is reduced 27.9% for original GTP plus GTP and GTP plus PMIP schemes and 30.6% for original enhanced HIP scheme, which is important for the user experience of on-going application.

![Fig. 2 Total Network Side Signalling Load](image)

![Fig. 3 Break Down of Average Handover Signalling Delay](image)

4. Summary of the work, potential impact & Conclusion
This paper describes a proposed HIP based micro mobility management scheme for 3GPP EPS. The result shows improvement on both signalling load and handover latency, which implies the improvement of the core network scalability and user experience during handover.

Reference:
Assessing and improving an approach to delay-tolerant networking

Lloyd Wood
Research Fellow, Centre for Communication Systems Research at the University of Surrey, e-mail: L.Wood@surrey.ac.uk

Abstract

Delay-tolerant networking (DTN) is a term invented to describe and encompass all types of long-delay, disconnected, disrupted or intermittently-connected networks, where mobility and outages or scheduled contacts may be experienced. ‘DTN’ is also used to refer to the Bundle Protocol, which has been proposed as the one unifying solution for disparate DTN networking scenarios, after originally being designed solely for use in deep space for the ‘Interplanetary Internet.’ We evaluated the Bundle Protocol by testing it in space and on the ground. We have found architectural weaknesses in the Bundle Protocol that may prevent engineering deployment of this protocol in realistic delay-tolerant networking scenarios, and have proposed approaches to address these weaknesses.

Key words: delay-tolerant networking, DTN, Bundle Protocol, end-to-end principle.

1. Introduction

Delay-tolerant networking was originally proposed as a generalisation of NASA-led work to move to packetized networking for its spacecraft. That work was named the ‘Interplanetary Internet.’ There, long propagation delays between spacecraft and scheduled, planned, contacts dominate communications. Extending the scope of the problem space to also include addressing very different, disrupted, terrestrial ad-hoc networks, including military networks, significantly increased interest in and funding for this new approach to networking, and has led to further development of the Bundle Protocol suite. The Bundle Protocol attempts to encompass many environments and use cases beyond its original deep space scenario, even though those other cases can be very different in connectivity and networking requirements [fig. 1].

We completed the first in-space tests of the Bundle Protocol for the Interplanetary Internet on the UK-DMC satellite [1], and have combined our practical experience with theoretical analysis to provide a detailed consideration of many technical aspects of the Bundle Protocol. The design of the Bundle Protocol ignores the reliability concerns that led to the development of the well-known ‘end-to-end principle,’ [2] and also raises other technical issues. The issues that we have uncovered include the important reliability and timing problems that we highlight here [3].

2. Technical Approach

The Bundle Protocol specifies a new way of transmitting data in a complex protocol format that is assembled from different blocks for different purposes. Blocks and header information can be inserted, removed and modified by intermediate nodes.

Emphasis on security has been a focus of the design of the Bundle Protocol from an early stage, with a complex security architecture that provides authentication of messages and encryption of data delivered. This is a deliberate change from earlier architectures; security was famously deliberately left out of the Internet’s TCP/IP suite, and had to be retrofitted later with IPsec, HAIPE, SSL and other protocols. However, this focus on security has neglected protocol reliability.

Transmissions and memory storage do not always produce perfect copies (although we may wish to believe so) and do have non-zero error rates. Any introduced errors must be detected with deliberate checks. A well-designed network protocol will sanity-check its headers to make sure that the information it is exchanging was received reliably before being processed. It may also sanity-check its payload data. Checking payload data must also be done in any case by the highest networking layer handling the payload, in accordance with the end-to-end principle, to detect introduced errors.

![Figure 1 – Comparing different DTN scenarios](image-url)
3. Results

Known deployments of the Bundle Protocol have run without any security being implemented. Three in-space tests of the protocol for the Interplanetary Internet – in our UK-DMC satellite tests and later on the Deep Impact/EPOXI comet probe [1] and on the International Space Station – chose not to implement bundle security. Not doing so can be attributed to a number of different reasons, including reliance on lower layers for ‘probably good enough’ reliability, lack of security code and specification readiness, lack of available memory to store and run code, lack of any threat to be worth mitigating against, and security not being required to be able to demonstrate the protocol running in space. The complexity of the Bundle Protocol is one argument against placing it in low-end embedded systems, and processing hardware for space is often low-end and unable to execute modern cryptographic algorithms rapidly. Non-essential processing is not done.

Recent EU trials in a remote area of Sweden also did not implement bundle security, and so are also running without high-level transport reliability checks as a side-effect of not having security [4]. The risks to data of doing so are well-known, and are described in the end-to-end literature [2].

The design of the Bundle Protocol is such that we suggest adding support for lightweight reliability checking within the imposed limits of the existing security framework [5]. Unfortunately, that workaround uses the complex security architecture and requires it to be implemented, so this is unlikely to see widespread adoption in embedded systems.

The Bundle Protocol also expects that all communicating nodes have a shared understanding of UTC time and its leap seconds, with synchronised clocks. Bundles expire after a set clock time and are discarded. Bundles sent from nodes with misset or drifting clocks may be expired at the next node simply because their timestamps are in the far past or distant future. If you don’t know the time, you can’t ask for the time by using the Bundle Protocol. A bundle age extension block has since been proposed for when UTC time is not known, but setting the age still needs working, reliable, clocks.

4. Summary of the work, potential impact and conclusion

We have evaluated the Bundle Protocol, highlighted architectural problems in its design, and proposed a workaround to give reliability. Our work shows that the basic design of the Bundle Protocol neglects important architectural issues. We expect this to limit its adoption and deployment.

References


SaVi: satellite constellation visualization

Lloyd Wood

Research Fellow, Centre for Communication Systems Research at the University of Surrey, e-mail: L.Wood@surrey.ac.uk

Abstract

SaVi, a program for visualizing satellite orbits, movement, and coverage, is maintained at the University of Surrey. This tool has been used for research in academic papers, and by industry companies designing and intending to deploy satellite constellations. It has also proven useful for demonstrating aspects of satellite constellations and their geometry, coverage and movement for educational and teaching purposes. SaVi is introduced and described briefly here.

Key words: satellite, orbit, coverage, 3D rendering, Unix.

1. Introduction

SaVi, the Satellite Visualization tool [1], is a computer program for visualizing and animating the movement of satellites and their coverage. SaVi was originally developed by Worfolk et al. at the Geometry Center at the University of Minnesota, but became homeless when that was closed due to lack of ongoing funding. Maintenance of the software was taken over by Lloyd Wood, who had found SaVi useful during his doctoral work on satellite constellations. SaVi has been maintained at the University of Surrey since then.

2. Technical Approach

SaVi exists as a standalone program that can also be run as a ‘module’ that interfaces with and controls the Geomview program [2]. Geomview is a general-purpose rendering program useful to mathematicians; SaVi leverages Geomview for simple three-dimensional (3D) rendering and OpenGL texturemapping, while ignoring Geomview’s ability to render higher dimensions of interest to mathematicians.

SaVi is implemented as a satellite orbit simulator, written in ANSI C, which is driven by commands added to the higher-level Tool Command Language (Tcl). This two-pronged approach allows SaVi to be scriptable. Simple, short, Tcl scripts generating satellite constellations and driving the underlying simulator are written in a similar manner to the scripts of the network simulator ns-2, which also relies on Tcl. Many scripts simulating, illustrating and animating proposed and existing satellite constellations are included with SaVi.

SaVi’s user interface is presented in Tcl’s Toolkit, Tk, which complements Tcl and allows for relatively straightforward creation of a graphical dialog- and window-driven system [fig. 1]. Seeing and animating a complex satellite constellation is as simple as clicking the Constellations menu and selecting, say, the Iridium system to run the associated script [fig. 2].

As SaVi relies only on Tcl/Tk and standard Unix POSIX libraries, with continued maintenance it remains portable across a wide range of Unix-compatible systems, including Linux, FreeBSD, and Mac OS X. It comes as an easily-installable Debian package for Ubuntu users.

SaVi can also be run under Microsoft Windows, using Cygwin or a virtualisation environment such as VirtualBox.

As a popular community-driven effort, SaVi is in the top 1% of projects on the SourceForge site for open and free software. SaVi’s portability and popularity is maintained by users reporting bugs and requesting features, or providing fixes for problems encountered with new compilers or with specific platforms. As a result, after over fifteen years of life, SaVi remains compatible with modern systems.

SaVi shows satellite coverage areas on a number of different map projections. A fisheye view of the sky is also available to examine how satellites pass over different points on the Earth. SaVi shows satellite coverage as either minimum elevation angle or as half-angle beamwidth, and indicates how that coverage moves over time. Graphical output can be recorded and saved. Satellite and constellation properties can be edited.

Multiple spotbeams on a satellite, communication channel properties, and precise orbital motion with complex precession are not yet simulated; the University has other, custom, tools for simulating these in far more detail.
3. Results

Since writing constellation scripts that shipped with SaVi 1.0 and adopting SaVi, Lloyd has added many new features, including many more scripts, new map projections, resizable coverage and fisheye displays, viewing of high-diversity constellations and coverage movement over time, moving coverage texturemaps, constellation generation tools, and an educational help system that explains each constellation.

While SaVi lacks the large number of features present in far richer commercial offerings such as the Satellite Toolkit (STK), it is entirely free to use, which makes SaVi immediately attractive to the academic community. Its output has appeared in over twenty research papers and articles [3].

SaVi enables quick and easy explanation of the basic features of orbital motion and satellite geometry. This can include showing the differences between rosette and star constellations, by contrasting the seamed Iridium and seamless Globalstar systems, comparing diversity, overlapping coverage, and the large number of satellites seen in the sky for navigation constellations, or demonstrating repeating-groundtrack designs such as Molnya [fig. 3]. SaVi has been found useful for teaching purposes at the MSc level and on short courses at a number of institutions, such as the International Space University and SUPAERO, when constellations and orbital movement of non-geostationary systems must be explained.

SaVi has also been picked up by industry companies designing satellite constellations for communication, and has been used publically to illustrate the designs of their systems [4, 5].

4. Summary of the work, potential impact and conclusion

Although SaVi only provides a relatively simple degree of satellite simulation functionality when compared to more full-featured commercial packages, its open codebase and contributions from around the world have led to a long-lived, robust, portable, cross-platform tool that has attracted a wide degree of interest. SaVi appears to have gained a useful educational role in introducing and explaining the properties of satellite constellations.

References

Abstract

A novel distributed MAC scheduling protocol with reservation mechanism is proposed for enhancing the QoS performance of IEEE 802.11e-based wireless mesh networks. The proposed mechanism guarantees resources in pre-configured contention-free period (CFP) for real-time sessions (RTSNs) in the network in a distributed manner. Here, resource reservation (RR) for multiple sessions can be made without any collision through an effective signalling process which coordinates the RR for RTSNs. Distributed admission control (AC) ensures that the existing RTSNs are not violated from QoS guarantees. In addition, a concurrent transmission (CT) mechanism is implemented within the pre-configured CFP. This is to further improve the bandwidth utilization by synchronizing the reserved transmissions among different nodes which do not interfere with each other. Simulation results indicate that the proposed scheduling mechanism can address the problems posed by inter- and intra-flow interference and can achieve guaranteed QoS for admitted RTSNs while providing fairness for the other sessions in the network.

Key words: IEEE 802.11e, multi-hop, quality of service, resource reservation

1. Introduction

Nowadays, IEEE 802.11-based technology has been widely employed in various applications such as Wi-Fi hotspots, city wide mesh networks, etc. However, guaranteeing quality of service (QoS) for time critical sessions is still a big issue.

In order to improve QoS in 802.11-based wireless networks, numerous solutions such as admission control (AC) algorithms, QoS scheduling mechanisms [1], QoS routing protocols have already been proposed. Moreover, IEEE 802.11e has been standardized for offering prioritized QoS in distributed networks by implementing enhanced distributed channel access (EDCA) [2], as well as guaranteeing QoS in centralized networks through the support of hybrid coordination function controlled channel access (HCCA). Although HCCA can inherently guarantee QoS for RTSNs by assigning sufficient resources (i.e. bandwidth) to them, it can not be employed in infrastructure-less networks, like ad-hoc networks. In addition, its polling mechanism is too complicated. In contrast to HCCA, EDCA can be simply implemented without a centralized controller. Thus, it can be a preferred choice for distributed networks despite of its limited QoS enhancement feature.

So far, many contributions have been made for addressing QoS provisioning issues in EDCA-based distributed wireless networks. Some solutions provide prioritized approaches based on end-to-end delay measurement or proper queuing algorithm. They try to assign adequate resource to real-time packets which urgently need to access the channel for meeting their delay requirement. However, these approaches merely guarantee the channel access for RTSNs as deferral and back-off mechanisms are still affecting the performance. Meanwhile, collisions among different nodes are inevitable in the face of multi-hop communications with multiple sessions which may interfere with each other. In order to deal with these deficiencies, novel scheduling schemes with dedicated resource reservation (RR) can be utilized.

In this paper, we propose a distributed resource reservation scheme for IEEE 802.11-based wireless mesh networks (WMNs). The proposed distributed RR scheme can provide RR for distinct RTSNs within interference range of each other, making them obtain dedicated bandwidth and transmit without collisions. For implementing this distributed multi-hop RR scheme, an explicit signalling process is utilized for coordinating and disseminating the state information of reservation among the nodes in the networks. This supports the QoS MAC scheduler to make multiple reservations without conflicts and/or collisions. Logical clusters (LCs) are created to alleviate the overhead introduced by the signalling process. Moreover, a concurrent transmission (CT) mechanism is utilized for scheduling simultaneous transmissions from non-interfering nodes on the route. This further enhances the bandwidth utilization in contention-free period (CFP).

2. Technical Approach

The proposed MAC scheduling mechanism is called EDCA with distributed resource reservation (EDCA-DRR). The purpose of EDCA-DRR is to provide satisfactory QoS for RTSNs by reserving dedicated resources. The proposed protocol utilizes service intervals (SIs) which consist of contention-free period (CFP) and contention access period (CAP). A RTSN can secure transmission opportunities (TXOPs), during which multiple data frames can be transmitted provided that the residual time left in the TXOP is still able to afford data transmission. To provide fairness, CAP will be consumed by other types of traffic sessions as well as non-QoS guaranteed RTSNs which still follow the legacy EDCA. The proposed mechanism can successfully schedule the TXOPs among different routes by deploying a supportive signalling process. Concurrent transmission (CT) mechanism is used for further optimizing the bandwidth utilization within the CFP. When a RTSN wishes to reserve dedicated bandwidth, the corresponding sender first implements the add
traffic stream (ADDTS) signalling process in order to check whether the session’s reservation request can be fulfilled. If the reservation can be met, the source node will continue the signalling process for making each participating and/or interfering member confirm and follow the RR. Moreover, CT partners are identified and synchronized during the signalling process. Each CT group contains the CT partners that can transmit concurrently. They will get their own specific transmission duration within the CFP so that the bandwidth utilization can be improved while avoiding the collisions.

The aforementioned ADDTS signalling process can be split into 3 sub-processes, namely ADDTS request process, ADDTS response process, and ADDTS synchronization process. ADDTS request process initiates and traverses from the source node to the destination when a RTSN wishes to reserve bandwidth for guaranteed QoS. This request process is able to identify interference relationship within the contending region of the traffic route. It also establishes logical clusters (LCs) along the route in order to mitigate the signalling overhead. The ADDTS response process takes responsibility for recognizing CT relationship and updating CT information across each LC. Eventually, ADDTS synchronization process is implemented for disseminating state information of reservation to all the nodes along the route as well as the off-route nodes located within the contending region of the tagged QoS route. Note that enlarged transmission power (extend as much as the original CS range, here we assume CS range is at least about twice the transmission range) is applied in the ADDTS request and synchronization processes for detecting interference and announcing the reservation information to the off-route nodes. The interference probing can support the CT recognition process. On the other hand, each off-route node that has received the information will align to the reservation scheduling so that collision can be avoided. Fig. 1 shows an example of reservation-based transmission of a RTSN 1 for a route from node 1 to node 9.

3. Results
In this section, we evaluate the effectiveness of the distributed RR scheme for WMNs with multiple-flows. The cross network topology is used in simulations conducted by network simulator - 2. In the simulation scenario, one RTSN as well as a NRTSN transmit from node A to node B. Meanwhile, another RTSN processes data transmission from node C to node D. The number of hops across the two routes is varied from 2 to 16. The network topology is given in Fig. 2.

Fig. 3 shows the delay performance of the RTSNs. With the increment of number of hops, the delay of RTSNs dramatically increases with EDCA. This attributes to the contention based channel access in EDCA, which incurs deferral, back-off as well as collisions and thus affects the QoS performance of RTSNs. Moreover, available bandwidth along the route can not be efficiently utilized with EDCA. This is because no coordination (i.e. synchronized transmissions) can be made for the transmissions from different nodes along the route. EDCA-DRR achieves enhanced performance, which can be seen in Fig. 3. This attributes to its CS interference avoidance mechanism. On the other hand, the CT mechanism further improves the efficiency of bandwidth utilization of RTSNs.

4. Summary of the work, potential impact & Conclusion
EDCA-DRR is an effective MAC layer protocol which can guarantee the QoS for multimedia traffics while considering the fairness to other type of traffics. We expect the future application of EDCA-DRR in wireless mesh networks for enhancing the services for clients.

References
An Agent-based Scheme for Supporting Service and Resource Management in Wireless Cloud

Yanbo Zhou

Final Year PhD Student, C.C.S.R., e-mail: Y.Zhou@surrey.ac.uk

Abstract

The wireless cloud is a service-oriented architecture that opens various service models for network access and service providers along with developments in cloud computing and emerging wireless technologies. The growing demand of wireless mobile data services and the delivery of those data services over wireless and mobile networks introduce severe bottleneck at the wireless link and the core network of underlying infrastructure. This paper proposes an agent-based scheme with an optimization resource selection strategy to reduce this bottleneck and improve quality of service (QoS) for internet data service over wireless mobile network.

Key words: wireless cloud, agent-based, QoS.

1. Introduction

Wireless cloud is the new trend in cloud computing which supporting the wireless data service of mobile wireless internet access demanded [1]. It is traditional grid computing evolved into wireless and mobile environments that seamlessly integrating the existing wireless access networks (e.g. WLAN, 3G/4G net) and therefore to support comprehensive and personalized services, providing stable system performance and Quality of Services (QoS) [1]. In practice, clouds offer services that can be grouped into three categories: software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). The delivery of those services over wireless and mobile networks introduces severe bottleneck at the wireless link and the core network of underlying infrastructure. Regard the nature of wireless access network in terms of coverage, bandwidth capacity, cost and so on, it makes a challenge of emerging wireless access networks as cloud virtual resources to support those service with the dynamic inter-operation, integration and disintegration of collaborative processing at different scales of network use in wireless cloud environments.

Wireless cloud implies a service-oriented architecture that leverages developments in cloud computing and emerging wireless mobile technologies, such as software radio and remote radio head technology. It enables cost savings through infrastructure sharing between different wireless operators and separation of software development [2]. Existing research efforts are focused on the discover service and an agent based model to provides computational service only by discovery, brokering and allocation of cost effective resources [3]. Most of wireless resource management is trying to provider wireless users being always best connected [4]. Obviously many of these existing efforts cannot enable seamless utilization, interoperability and composition of the distributed resources and services to supporting the demand of wireless data service in wireless platform. In this paper, we present a novel agent-based scheme for supporting wireless data service and resource management based on user requirements and preferences within wireless cloud environment.

2. Proposed Agent-Based Scheme

Development in mobile and wireless access network technologies and increase in the demand of user wireless access application are driving the deployment of service provisioning and efficiency resource management within wireless cloud system. Figure 1 shows the outline wireless cloud system with an agent-based scheme for supporting cloud service to wireless platform users.

![Fig. 1 Proposed agent-based scheme](image-url)

In the proposed architecture includes different properties: wireless AN resources, user WT is equipped with a wireless user agent (WUA), agent-based server (ABS) acts as a gateway entity and that enable to discover the requested cloud service from the application server providers and selects and allocates optimal resources to supporting and maintaining the...
user application. Otherwise, all of the wireless ANs are virtualized network resources and they are coordinated by ABS as virtual resources. We give a brief overview on the involved agent-based scheme and interoperation/interaction between its several components in the architecture. The following top level roles can be envisaged:

(i) Wireless User Agent (WUA): It will be present in every WT (mobile, laptop, PDA). The WUA manages the application requirement and monitors the delivered QoS of wireless data service. Specifically, WUA simply issues one single application request message (capable of carrying QoS constraint of application) to multicast to reached ABS through ANs. One most importance task of the WUA is making decision of selection of the request service and AN resource by using optimization resource selection strategy which equipped with it. Any time the WT user moves (by changing the new AN) the WUA monitor will update the status of service QoS. The service monitor gathers necessary up-to-date information on application server and underlying network conditions (i.e. wireless link) for supporting optimized service. If the service QoS is degraded, WUA will send new requirement of update application server.

(ii) Agent-based Server (ABS): It allows the WT to connect to the wired infrastructure through wireless links by the presence of ANs. It enable to acquire the service information from application server provider; monitor the updated status of all the AN resources within the system and process and serve the requirements received from WT users. While it receives a request from users: first, it discovers the availability of service from application server; second, service management works out the QoS information of service provision. Third step, allocates and delivery the service to supporting and maintaining the user’s application. ABS collaboratively resolves the service in the request, in server domain-by domain manner, towards the desired application service provider. Once ABS receive the decision of requested service from WUA, the selected application server will deliver the application service to the WT user by allocated AN via wireless link.

(iii) Application Server Provider: The system consists of an ABS in each edge components and some non-agent application server components like application server or storage server. It offers application content to be accessed and WT user across the internet. They are responsible for dealing with the service discovery requests from ABS and WUA and the actual delivery of application service with QoS awareness.

Such a paradigm, wireless users requesting wireless data service that simply issues one single service request message by connect to reachable the agent-based server through AN via wireless links. Agent-based Server (ABS) deal with the problems: such as how to discover wireless data service meet the demand of users and efficiency utilize and manage cloud resources (data, storage, network, etc), how to guarantee quality of service and select and allocate resources fairly among users, and finally, the delivery of wireless data service at Internet scale. A Wireless User Agent (WUA) resides in the user’s Wireless Terminal (WT) that supports optimized resource selection and monitors the service QoS covering the entire lifecycle of the service provision.

3. Key Results and Discussion

In this section, simulation and analytical results show that the proposed agent-based optimization scheme achieves better resource utilization and decreases 22% of the required service false ratio. We simulate a wireless cloud system with 7 wireless AN resources and the basic video server as application server used in this simulation. Uniformly set of [10, 50] users are distributed in this wireless cloud system. We set the duration of video is uniformly [5, 60]; network transmission bandwidth ranges from 500kbps to 2 Mbps. The application servers are set in 3 difference location and all the video data are stored in these 3 application servers.

![Figure 2 Successful request ratio vs. user requested service](image)

Figure 2 illustrates successful ratio of the user application service against the user submitted service requirements, the result shows the different successful ratio between apply agent-based scheme with resource management strategy to select and allocate the AN resource or normal resource selection based on of user’s application constraint. A novel agent-based scheme for supporting service with proposed an optimization resource selection strategy for delivery service. The simulation result shows improvement of resource utilization in wireless cloud environment. Future works are planned to prove the efficiency of energy-awareness resource selection and management algorithm under mainstream wireless cloud environment.


