Towards a Framework to Support Novice Users in Understanding and Monitoring of Home Area Networks

Yuqian Song, Rob Brennan, David Lewis, Owen Conlan
FAME & KDEG
School of Computer Science and Statistics
Trinity College Dublin
Dublin, Ireland
{yuqians, rob.brennan, dave.lewis, owen.conlan}@scss.tcd.ie

Abstract—Advances in Home Area Network (HAN) technologies have afforded end users increased convenience in performing everyday activities. However, even seemingly trivial issues can cause great annoyance for the novice user who lacks domain expertise of often complex HANs that underpin these advances. A key challenge lies in assisting novice users in understanding and monitoring of the obscure and complex HAN. This research proposes a framework to leverage domain expert knowledge to enable real time semantic up-lift in supporting novice end-users to understand and monitor the home area network. This presents a significant opportunity to increase user satisfaction and reduce associated support costs. This semantic approach has been designed and implemented in an early prototype of our Home Area Network Monitoring System (HANMS). This paper presents a detailed description of the current state of the research, an initial evaluation, and future work.

Keywords—information up-lift; network monitoring; HAN; domain expert knowledge; non-expert user

I. INTRODUCTION

Home Area Network (HAN) technologies are evolving rapidly and becoming more and more pervasive and complex. Recent data shows 43 million US families and over 99% of UK households are connected to a broadband enabled exchange [11][12]. These HANs are typically small in scale with heterogeneous devices interconnected to enable file sharing, on-line gaming, live video streaming, VoIP communication and other applications working together to increase the convenience and satisfaction of HAN users. Such HANs require significant network skills and knowledge to understand how well the home network is used with a rich variety of devices and services. Inexperienced HAN users are considered as novice users who are continuously confused and frustrated with even simple HAN performance and maintenance tasks [10]. For example, an important VoIP call on an iPad suddenly suffers degradation of wireless connection quality, perhaps caused by some other WiFi activities initiating on the same channel – causing interference, or a signal weakens when the device antenna is obstructed, or another user may be downloading a large file on a PC wired connected to the same home network which causes network congestion. Such events may be unremarkable but will impact on the end-user’s perception of the quality of the service supported by the network connection. Additional cost for both the network provider and HAN user will be incurred to detect and diagnose such issues. A significant impediment to recognition, diagnosis and correlation such events for non-expert users is the problem of knowing what is happening, why it is happening and how to solve it in their HANs. However, most existing tools have limited capacity to provide relevant information except through rudimentary but hard-to-understand logging analysis [13].

This research addresses the challenge of how to bridge the semantic gap between the hard-learned knowledge of domain experts and the cognitive competencies of novice users in a HAN context. To address this challenge, this research proposes an autonomic, flexible and harmonious framework to use domain expert knowledge to enable the real time semantic up-lift of meaningful information from raw monitoring data to support novice end-users to understand and monitor home area networks. By autonomic, it is meant that this framework should support automatic HAN monitoring and anomaly diagnosis without any user input; by flexible, it means our framework could not only consume data in a variety of forms and levels of detail, but also adopt diverse measurement and annotation methods to the raw data according to the domain expert knowledge; by harmonious, it means different parts of our framework could be harmonized to share the meaningful information and driven by unified domain knowledge.

This framework aims to achieve the following objectives:
• Properly model and aggregate captured insights and knowledge of the domain expert with open knowledge on the web for the HAN monitoring;
• Up-lift meaningful information from raw data (e.g. wireless network log data in a HAN scenario) through leveraging expert knowledge model in a real-time dynamic process;
• Adopt domain knowledge driven reasoning to facilitate the diagnosis and analysis of anomalies;
• Propose a structure to harmonize the information up-lifting, anomaly diagnosis and visual presentation process with domain expert knowledge;
• Present an organized, correlated and annotated view of the up-lifted information to enable non-expert users to understand potentially obscure details of the home area network.
This work-in-progress paper is organized as follows: First, related research already addressing some of these objectives will be briefly reviewed in section II. Next, the Framework Design section introduces a framework for the proposed approach. Then the Implementation and Evaluation section presents an early prototype of the approach applied to HAN management, followed by preliminary evaluation results. Future work and conclusions are described in the final section.

II. RELATED WORK

In 2010, the number of world-wide broadband connections has been estimated at over 364 million [14], while the global network monitoring and management market is over 1.9 billion [11]. In this rapidly evolving area, a series of significant challenges lies in the necessity and complexity of understanding, monitoring and managing modern home networks for most households [15][16]. A recent usability study [17] shows that even though there are several tools already embedded in operating systems and routers, most non-expert HAN users still desire a simple, direct and explicit way to monitor their whole home network and to diagnose network problems.

The current HAN monitoring tools that users mostly rely on are built into the local device and the router [17]. Some existing commercial tools like NetworkMagic [19] from Cisco and PTRG network monitor [20] has shown a local-based approach to visually present domestic network topologically with configuration and performance monitoring of local machine. Some other locally-based applications [21][22] aimed at providing extensive measurements at the host level to infer application network requirements, and identifies network related problems through time-series analysis with aggregated monitoring information from other devices in the same HAN. The Homework project [23] intends to create the next generation of domestic infrastructure that combines empirical understanding of use with a fundamental re-invention of the protocols, models and architectures of the domestic setting.

This project has proposed a router-based information plane architecture to enable the flexible real-time measurement and analysis of composed events to drive high-level management and visualization for non-expert HAN users.

Ontology-based network monitoring has recently evolved from a theoretical proposal to a more mature technology, which is considered as a novel and effective strategy to overcome HAN complexity by involving ontology-based modeling and reasoning; e.g., Hoag [3] presents an approach to apply semantic reasoning techniques for network management and resource allocation to avoid overbuilding and improve quality. FOCALE [4] is a semantically rich architecture for orchestrating the behavior of heterogeneous and distributed computing resources. Vergara [24] described and summarized several ontology-driven network management and monitoring projects, detailing the most important facets of how semantic technologies were applied and explaining the advantages and drawbacks. In his paper, he found semantic technologies are explicit, formal, and shareable, which means the ontology-based modeling and reasoning could be composed with other semantic techniques to express the formal network monitoring and management logic and improve current approaches. A novel project [25] proposed an ontology-based formal definition of the different management behavior specifications integrated with the management information definitions, in which SWRL rules are defined directly over the ontology elements and allow for logical reasoning. Another ontology-based approach could dynamically evoke the internal and external ontology models, which remedies the shortcomings of some policy-based approaches. Moreover, the measurements provided by different network monitoring tools and platforms could be modeled and integrated with a syntactic ontology-based solution [26]. However, the continuous evolution and heterogeneity of HAN technologies coupled with varying user requirements still poses a significant challenge for further research to leverage the correlated knowledge to support non-expert users in HAN monitoring.

Another HAN monitoring challenge lies in how to extract meaningful information from the vast number of dynamic network log datasets from heterogeneous HAN devices and services. Driven by Semantic Web technologies, Semantic Annotation can be applied to transform and enrich the data and information with formal semantic meanings. Handschuh and Staab [1] introduce a broad range of technologies and methods for the explicit construction of semantic annotations, including approaches for data extraction, collaboration and translation into Semantic Web metadata. SARA [2] can be used to gather heterogeneous data from different resources and endow the raw data with semantic attributes through rules specified by domain experts. These semantic attributes are intended to support non-expert users exploring an information domain across heterogeneous data sources. Adopting such an approach to enhance and assist the network monitoring process is challenging however due to the highly dynamic and distributed nature of network monitoring data.

Visualization techniques can effectively present complex data by capitalizing on human perception capabilities. Firstly, how to visually present correct information for small scale networks for non-expert users is becoming a recognised challenge in the HCI community and network visualization is considered to be the primary interface for HAN monitoring and management [18]. A novel interactive visualization system [5] was proposed as a router-based approach to monitor the data collected from a home router and control bandwidth usage for family users. Another interactive network management system, Eden [6], delivers a simple conceptual model to help users understand key aspects of the network by eliminating the barrier and cost of the network technical minutia. Our research is also based on [7] to express a semantically enriched visualization approach for non-expert users to monitor networks and services.

III. FRAMEWORK DESIGN

In order to meet the research objectives and challenges above, this section describes the HANMS framework (Figure 1). It is designed to leverage domain expert knowledge and linked data on the web to enable the real time semantic up-lifting of meaningful monitoring information from raw monitoring data to support novice end-users in understanding and monitoring their HAN. This HAN event processing framework is designed in a layered structure associated with both a domain expert
knowledge model and relevant linked data on the web. It is structured as: Information Uplifting Layer, Semantic Processing Layer, and Visual Representation Layer.

Figure 1: HANMS Framework Design

A. Domain Expert Knowledge Model

The insights and knowledge of domain experts are captured and modeled in the domain knowledge model. As partly exposed in Figure 2, captured domain expert insights are decomposed into schemas and rules, which are associated with domain ontologies and modeled in a domain expert knowledge model. In this model, there are four main classes: Condition, Event, Reason, and Solution. The Condition class indicates a condition which could be used to trigger an event. These condition rules in the knowledge model, which can be cascaded, are currently encoded in simple SWRL [8] rules. These can be loaded, parsed and evaluated by the semantic reasoner. For instance, if AntennaNoise_Bad and SignalStrength_Good, then Interference. Based on condition rules, one condition could contain one or several semantic attributes, which are pieces of semantic encodings captured from domain experts. For example, the condition to trigger the event Interference is composed by two semantic attributes: AntennaNoise_Bad and SignalStrength_Good. The Semantic Attribute Schema is XML-based schema with expert defined parameters and mappings to bridge the semantic meaning in raw log data stream, like “Antenna Noise is too high”, with the semantic attribute AntennaNoise_Bad, where restrictions and thresholds can be adjusted by domain experts via the visual widget. The Event class is used to describe the network problem or network status. Events could also be conditions to trigger other events. The relationship between Condition and Event is described as SWRL rules, which can be automatically reasoned over by the ontology reasoner. Event instances are also associated with the HAN component: device, service, gateway or the HAN itself.

The Reason class indicates the expert-defined reasons which may cause an event of a given type. Reason also has conditions which could be considered as a SWRL rule for the ontology reasoner to analyze whether the event is actually caused by this reason. For example, OutOfRange could be the reason for events of type SignalLost.

Figure 2: Domain Expert Knowledge Model

The Solution class describes expert-defined solutions which could be applied against one particular reason for the event. The event, reason, and solution classes can then be associated with the linked data entity which is retrieved and mapped with relevant linked data on the on-line data repository.

B. Linked Data

Linked data provides an approach to interlink and publish data, information or knowledge on the web so that they are machine processable. The DBpedia [27] project extracts structured data from Wikipedia and re-publishes it on the web with a public API to enable queries and interlinking with other datasets and applications. This provides a general purpose knowledge base with a very large number of instances (3.64 million things in version 3.7). Many HAN devices are included in this data-set and it can be used as a source of further context for them. In addition more abstract HAN concepts such as “router”, “wireless networking” and so on are also described, enabling the presentation of introductory material on relevant concepts to home users in the context of a particular network event.

However this knowledge is not strongly structured and suffers from inconsistencies so significant effort is required in performing mappings to our local knowledge base. For example, it may be necessary to process DBpedia’s SKOS [28] category tree in addition to the class inheritance hierarchy to correctly classify an instance and instance values must have some checking to guard against erroneous values present due to human input error in Wikipedia itself. It is also necessary to carefully handle incompleteness since some information though present in Wikipedia is not amenable to the automatic extraction techniques of DBpedia and is thus either not present or very weakly classified and hence hard to extract with standard queries.

In our framework, we harvest the structured linked data from DBpedia and map them to corresponding local domain ontology concepts to gather relevant information about connected HAN components. This information is used to identify the type of detected device/service and provide rich and easy-to-understand introductory text for technical concepts in the HAN environment. This enables us to partially automate the generation of Linked Data Entity instances in our local knowledge base.

As described in Figure 3, our framework gathers information of active devices and services, for example using UPnP device detection, classify those devices as specific HAN
component or device types in our local knowledge base and then retrieve the related concepts from the linked data on the web to get extra information, such as device/service type, manufacturer, model, introduction text, etc. This could also be useful for the information uplifting process itself but that is to be studied in future work. The information harvest from the web is also potentially of great benefit for novice users to understand the details of detected problems, expert-supplied reasons and solutions. However ensuring information quality and user overload are critical concerns when presenting this additional information. Given the characteristics of sources like DBpedia it means that run-time dynamic queries of these sources are unlikely to be productive in the short term and instead they provide a way to quickly seed explanatory information that must be pre-checked by domain experts for its validity and utility.

The information up-lifting process in the domain knowledge model are checked one by one in an event diagnosis loop, in which the information is iteratively annotated with events from low-level to high-level. For example, a particular semantic attribute could be considered as a low-level annotation, and if there is another event whose condition is based on this initial annotation we can refer to higher-level events, and so events are annotated level by level. All annotated events are kept in an event pool. In the up-lifting process, the event pool constantly checks the semantic annotation loop until there are no more new events (and no rules to fire) and at that time, the up-lifting of the data in this time interval is finished. The events in the event pool are then maintained for use in other layers.

C. Information Up-lifting Layer

The information up-lifting layer supports diverse annotation patterns and processes. In a given time interval, wireless log data of all monitored devices and services in the HAN is collected and aggregated. A pattern is applied to detect the change point (CP) of the real-time data stream. Our pattern is based on the algorithms in [21]. With the detected change points, we could divide the data stream into different intervals and discrete points. As shown in Figure 4, the linear fitting method is applied to the interval between two CPs. According to the slope and mean of the line segment, we annotate data between two CPs with semantic attributes like “sharp_increase”, “throughput_low”, and so on.

According to the expert-defined semantic attribute schema, the intervals and discrete points are annotated with the corresponding semantic attributes, e.g. “increase_CP”, “lost_signal”, “new_service” and so on. There are several types of annotation process for this semantic annotation pattern(P): The first process is the real-time change point annotation, which could be considered as a sequence of semantic meaning points(S) annotated for a real-time data stream, i.e. \( P = \{S_1,...,S_m\} \), where \( S = (s, t) \) is a pair with the semantic meaning (s) at timestamp t. The second process is a status annotation, which annotates status events like AntennaNoise_BAD to data intervals, i.e. \( P = \{S_1,...,S_m\} \), where \( S = (b, t, t) \) is a triple with the data status in a period \((t_1, t_2)\). The third process is a high-level semantic meaning annotation. The high-level semantic attribute with the corresponding semantic meaning (s) is determined according to other semantic attributes \((s_1,...,s_n)\), i.e. \( P = \{S_1,...,S_p\} \), where \( S = (s, f_1, f_2) \) is a triple with the behavior (b) happened in a period \((t_1, t_2)\), and e.g. “Play” is a behavior for an IPTV service. When the IPTV service is playing, the semantic attribute playing is dynamically annotated into the log data flow.

Through these semantic annotation patterns and processes, semantically meaningful information is extracted from raw data. According to annotated semantic attributes, all related events in the domain knowledge model are checked one by one in an event diagnosis loop, in which the information is iteratively annotated with events from low-level to high-level. For example, a particular semantic attribute could be considered as a low-level annotation, and if there is another event whose condition is based on this initial annotation we can refer to higher-level events, and so events are annotated level by level. All annotated events are kept in an event pool. In the up-lifting process, the event pool constantly checks the semantic annotation loop until there are no more new events (and no rules to fire) and at that time, the up-lifting of the data in this time interval is finished. The events in the event pool are then maintained for use in other layers.

D. Semantic Processing Layer

The events generated in the information up-lifting layer are highly distributed, structured and in real time. “Distributed” means events from heterogeneous resources and different annotation patterns and processes. “Structured” indicates all these distributed events are maintained in a formal and explicit structure, which exposes the source, time stamp and even the relationship between different events. These events are also highly dynamic to support real time anomaly diagnosis. Driven by the domain knowledge model, the event aggregation engine reviews all the events in the event pool to evaluate the health and status of all devices and services in the HAN and detect existing and potential problems. If some problems lead to the
health degradation of devices and services, we consider that this kind of problem is anomaly. Once the anomaly is detected, a root cause analysis process will be applied on this anomaly. In Figure 5, an IPTV quality degradation problem is referred to be caused by the AntennaNoise_Bad of source device. The aggregation, diagnosis and analysis result is also represented in the visual user interface (Figure 6), which allows novice HAN users to explore what is happening, what will happen, what caused the problem, and the available solutions.

![Figure 5: The reason analysis tree of an anomaly](image)

![Figure 6: The screenshot of the anomaly analysis widget of the HANMS Prototype](image)

E. **Visual Representation Layer**

In the Visual Representation Layer, several user-friendly widgets are designed for the non-expert users to understand and monitor the HAN according to the aggregated and up-lifted information through the communication middleware. It is particularly noteworthy that the information received from the semantic processing layer is independent of any particular visualization widget, so the visualization layer can embed additional expertise-driven logic to select or personalize the most appropriate presentation widget for a given combination of information and user. This separation of domain-specific expertise from visualization-specific expertise improves on the current approach of embedding domain reasoning, and associated domain-level assumptions, in the presentation layer.

IV. **IMPLEMENTATION AND EVALUATION**

A. **Implementation**

An early stage prototype of Home Area Network Monitoring System (HANMS) has been developed and deployed from a local server (CPU:E8500, Memory:4GB). Raw wireless log data was generated from a physical wireless monitoring testbed [9] and collected in device/service data metrics (CSV format), which are loaded into a simulation engine for demo and test purposes. HANMS is fed by the simulation engine to generate several wireless device log data streams every second seeded by the data collected from real wireless devices. The expert defined schemas are stored in an XML database, which can be adjusted via a simple widget. The server-side HANMS logic, which imports schemas and domain ontologies, was driven by Jena [29] and Pellet [30], an ontology API and reasoner. This is deployed on a Tomcat server in the local network and the client-side visual components are implemented in Flex (the anomaly analysis widget in Figure 6).

B. **Evaluation**

The current HANMS prototype has been initially evaluated under three different tasks with two target groups: novice HAN users (19 volunteers) and domain experts (5 experts) [9]. Novice HAN users have no or limited network technology background. Domain experts are HAN engineers with many years’ experience in network configuration, monitoring and management. In this study, participants were required to use HANMS prototype to monitor the HAN and detect/understand three different network performance problems by themselves. They then completed a SUS-based feedback questionnaire on the functionality, effectiveness, efficiency, usability, user-interface, and limitations of HANMS. Most participants agree that HANMS is usable by a non-expert user. More than 70% of participants believe HANMS could detect and analyze common network problems effectively. As a result of the test tasks, 50 out of 57 HANMS observations were correct, which indicates that even in its current prototype state HANMS is suitable for some tasks in different scenarios. Moreover, nearly all participants agree that HANMS is more efficient than traditional network monitoring tools. Some limitations of the current prototype is also identified in this evaluation: only one anomaly can be diagnosed at the same time; the structure in the event pool is not suitable for large amount of real-time events; more annotation labels and textual information are desired for the components of visual widgets. These limitations will be addressed in future development iterations.
V. FURTHER WORK AND CONCLUSION

This paper has introduced a framework to leverage domain expert knowledge to enable the real time up-lift and semantic reasoning of meaningful information from raw HAN log data in order to support novice end-users in understanding and monitoring their HAN. This layered framework consists of a domain expert knowledge model, an information up-lifting process, a knowledge-driven event diagnosis and analysis process and a visualization interface, which are partly implemented in a HANMS prototype. The current prototype is still under active development. The evaluation results show the early stage HANMS prototype has fulfilled some of proposed objectives and has the potential to support rich features. Future releases of the prototype will further demonstrate how semantic web technologies can improve traditional network monitoring applications in different application scenarios.

ACKNOWLEDGMENT

This work is partly funded by Science Foundation Ireland via grant 08/SRC/I1403 — Federated, Autonomnic Management of End-to-End Communications Services (FAME).

REFERENCES