Journal of Network and Computer Applications 34 (2011) 1928–1941

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Cross-layer signalling and middleware: A survey for inelastic soft real-time applications in MANETs

Sarogini Grace Pease [a,*], Lin Guan [a], Iain Phillips [a], Alan Grigg [b]

[a] Department of Computer Science, Loughborough University, Loughborough Leicestershire LE11 3TU, UK
[b] Systems Engineering and Innovation Centre, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK

A R T I C L E I N F O

Article history:
Received 13 January 2011
Received in revised form
29 June 2011
Accepted 2 July 2011
Available online 26 July 2011

Keywords:
Cross-layer Design
Real-time traffic
MANETs

A B S T R A C T

This paper provides a review of the different cross-layer design and protocol tuning approaches that may be used to meet a growing need to support inelastic soft real-time streams in MANETs. These streams are characterised by critical timing and throughput requirements and low packet loss tolerance levels. Many cross-layer approaches exist either for provision of QoS to soft real-time streams in static wireless networks or to improve the performance of real and non-real-time transmissions in MANETs. The common ground and lessons learned from these approaches, with a view to the potential provision of much needed support to real-time applications in MANETs, is therefore discussed.

Contents

1. Introduction . . . 1928
2. Background . . . 1929
   2.1. Real-time applications for MANETs . . . 1929
   2.2. Cross-layer optimisation defined . . . 1930
3. A taxonomy of cross-layer signalling methods . . . 1931
   3.1. Network-wide signalling . . . 1931
   3.2. Local node signalling . . . 1932
4. QoS control with protocol tuning . . . 1934
   4.1. Network-adaptive tuning . . . 1935
   4.2. QoS-adaptive tuning . . . 1937
   4.3. Hybrid network and QoS-adaptive tuning . . . 1938
5. Conclusions . . . 1939
   Acknowledgement . . . 1940
   References . . . 1940

1. Introduction

Mobile ad hoc networks (MANETs) are emerging in all sectors as the vision for future communications. This vision has at its basis the belief that a mobile device, whatever its location and speed, should have the ability to connect to the rest of the world. For example, a cellphone user may require access to a video stream while at a distance from a cellular mast. In the military sector, the mobile device may take the form of an aircraft transmitting mission critical video data to a ground unit, on friendly vehicles in the vicinity. The usefulness of such connectivity is not limited to communications and there is growing interest in the transmission of command and control data over ad hoc links, for example in the operations of remote industrial or medical safety-critical devices.

The provision of such services to users is dependant on ability to guarantee a high level of performance or QoS. A MANET has several performance-limiting factors, stemming from the mobility of the infrastructure devices (or nodes) and the nature of the transmission medium. A MANET is a self-configuring wireless network where mobile devices connect to each other, when in range, creating a dynamic and somewhat unpredictable topology over which packets can be forwarded. Such networks may stand-alone or be connected

* Corresponding author. Tel.: +44 7944976196.
E-mail addresses: S.G.Pease@lboro.ac.uk (S.G. Pease), L.Guan@lboro.ac.uk (L. Guan), I.W.Phillips@lboro.ac.uk (I. Phillips), A.Grigg@lboro.ac.uk (A. Grigg).

1084-8045/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2011.07.005

to other wired and static wireless nodes or networks, although these are generally not considered part of the MANET. As a result of node mobility, connections are intermittently set up and torn down and there is also the potential for no end-to-end path to exist at a point in time. A single node will therefore discretely act as transmitter, receiver and router. In the latter case it is referred to as an intermediate node, on the path between packet transmission and receipt.

Problems faced in providing QoS guarantees in wireless networks are extended to ad hoc networks. Radio frequency transmissions, propagated from multiple transmitters into a real world environment, are subject to interference, multipath fading, Doppler effects and shadowing. In this environment, a channel, or the path of a packet from one node to another, varies in quality along its length and over time. With no spatial separation of channels low frequency radio transmissions can also interact with each other. All these conditions are characterised by varying signal to interference-plus-noise ratios (SINR) that decrease with inter-nodal distances. SINR determines channel and application performance. What results from low SINR is limited and varying available bandwidth, a high frequency of bit errors and packet loss and increased packet latency and jitter.

The traditional, layered protocol stack of wired networks institutes static relationships between modular protocol layers. These relationships are characterised by the encapsulation of functions in higher layer objects so that they are hidden from their underlying sources. Layered approaches to network QoS control perform poorly in ad hoc networks due to the MANET-specific characteristics that do not appear in static networks (Perkins and Hughes, 2002):

- Node mobility leading to random and possibly rapid topology changes.
- Available throughput that is variable and asymmetric and lower than maximum transmission rates.
- Lack of centralised control.
- Limited processing capacity, memory and energy resources.

This creates a challenge in providing guarantees of bounded jitter and latency to applications with real-time (RT) deadlines. However, when layer boundaries are blurred and layer information (representing the quality and availability of channel resources as well as QoS requirements) can be shared and tuned between layers, QoS guarantees can again be provided. Cross-layer design, moves away from the oblivious layered approach, introducing layer interdependence.

Many cross-layer models have been proposed in the two fields of performance improvement in MANETs and of soft RT (SRT) in wireless networks, with a few straddling both. Proposals concentrate either on the performance of the signal transfer mechanism or on the tuning of specific protocol parameters to improve network performance. The majority of these have been developed in order to meet highly specialised network performance goals such as video quality improvement in spite of changing channel conditions. Caution has therefore been suggested in avoidance of "spaghetti design" wherein the complexity of a cross-layer interaction can reduce its re-usability (Kawadia and Kumar, 2005).

The focus of this paper is to provide a survey of recent cross-layer proposals in both of these fields, identifying the common ground and learning points garnered from both. This is done with a view to identify the signalling and protocol tuning methods that can provide necessary QoS guarantees to delay critical SRT applications in MANETs. Key holistic or middleware proposals for cross-layer information exchange are first investigated, beginning with cross-layer implementations that rely on global, network-wide information such as the contention-aware admission control protocol (CACP) and dual carrier sensing with parallel transmission (DSCPT) awareness. Then proposals that utilise only local information contained within the node are evaluated, including the Mobile Metropolitan Ad hoc Networks (MobileMAN) architecture and the Efficient Cross-Layer Architecture (ECLAIR). A taxonomy of these is created, with a view to their potential to support delay critical SRT.

Aside from architectural approaches to cross-layer design, numerous research proposals have also concentrated on adaptively fine-tuning certain protocol parameters according to QoS and network requirements. The QoS of delay critical SRT applications has a high sensitivity to channel quality changes, hence evaluation of this second group of proposals provides a grounding for reducing the optimisation requirements of a cross-layer model to only those parameters with a strong influence on network performance. Section 5 thus concludes on the signalling mechanism and tunable parameters that should be incorporated in developing an appropriate optimisation model for delay critical SRT in MANETs.

The rest of the paper is structured as follows, Section 2 presents the background of the paper, by identifying the requirements of typical RT applications from a MANET scenario. This is followed by a suggested definition of cross-layer optimisation. A taxonomy of existing signalling mechanisms with representative examples is developed in Section 3, first considering those that use network-wide information in their optimisation and second those that utilise only local nodal information. This is followed, in Section 4, by discussion of the literature on protocol tuning to compensate for failing network conditions or to respond to stringent QoS requirements and of some approaches that propose to do both. The learning points from the models discussed in Section 3 and the fine-tuning approaches from Section 4 and also their usefulness to SRT applications running over MANETs is concluded upon in Section 5.

2. Background

2.1. Real-time applications for MANETs

Timeliness is key to RT flows, for which QoS depends strongly upon deadline achievement and high packet arrival rates. All application processes and transmitted packets can be categorised as RT or non-real-time (NRT). RT processes are time-triggered, based on an internal system schedule or event-triggered by environmental stimuli, and explicitly use global physical completion time constraints to manage their resources. QoS for RT packets is therefore often expressed primarily in terms of deadline achievement. There is no benefit in delivering RT packets early, as this in fact may be detrimental to the system due to the consumption of buffer resources by the storage of early arriving packets. NRT processes may perform computations which satisfy their timing requirements but resource management is not time or constraint driven.

The definition of RT is divided into hard real-time (HRT) and soft real-time (SRT) and the latter has further been subdivided into inelastic and elastic SRT (Li et al., 2011):

- HRT processes have strict end-to-end delay requirements, and late packets are considered unusable. This is because the completion of a related computation after its deadline will impede a systems ability to operate correctly or have a critical impact on the system. Hard deadlines are therefore used in safety-critical systems to guarantee no damage to equipment or personnel. HRT packet deadlines must always be realised in order for minimum QoS guarantees to be met, for example all