Available online at www.sciencedirect.com
ScienceDirect
Ad Hoc Networks 6 (2008) 524–538
www.elsevier.com/locate/adhoc

Adaptive audio streaming in mobile ad hoc networks using neural networks

Daniel W. McClary [a], Violet R. Syrotiuk [a,*], Vincent Lecuire [b]

[a] *Department of Computer Science and Engineering, Arizona State University, Brickyard Suite 523, 699 South Mill Avenue, Tempe, AZ 85281, United States*

[b] *Centre de Recherche en Automatique de Nancy (CRAN UMR 7039), Nancy-Université, CNRS Faculté des Sciences et Techniques, BP 239, F-54506 Vandoeuvre-lés-Nancy Cedex, France*

Received 24 July 2006; received in revised form 10 January 2007; accepted 19 April 2007
Available online 22 May 2007

Abstract

We design a transport protocol that uses *artificial neural networks* (ANNs) to adapt the audio transmission rate to changing conditions in a *mobile ad hoc network*. The response variables of throughput, end-to-end delay, and jitter are examined. For each, statistically significant factors and interactions are identified and used in the ANN design. The efficacy of different ANN topologies are evaluated for their predictive accuracy. The Audio Rate Cognition (ARC) protocol incorporates the ANN topology that appears to be the most effective into the end-points of a (multi-hop) flow, using it to adapt its transmission rate. Compared to competing protocols for media streaming, ARC achieves a significant reduction in packet loss and increased goodput while satisfying the requirements of end-to-end delay and jitter. While the average throughput of ARC is less than that of TFRC, its average goodput is much higher. As a result, ARC transmits higher quality audio, minimizing *root mean square* and Itakura–Saito spectral distances, as well as several parametric distance measures. In particular, ARC minimizes *linear predictive coding* cepstral (sic) distance, which closely correlates to subjective audio measures.

Keywords: Adaptation; Transport protocols; Neural networks; Mobile ad hoc networks

1. Introduction

Streaming audio and video over computer networks poses numerous challenges, yet offers great potential. The promised benefits of *voice over IP* (VoIP), entertainment, telemedicine, and other applications drive research in robust media streaming. In *mobile ad hoc networks* (MANETs), the challenges are more pronounced because the quality constraints must be satisfied under rapidly changing network conditions.

Our focus is on an adaptive transport protocol for media streaming. Perhaps the most common approach to adaptation is to have the destination of the flow explicitly return feedback to the source. Consider the *TCP-Friendly Rate Control* (TFRC) protocol [1], designed for applications that vary their transmission rate in response to congestion.

* Corresponding author. Tel.: +1 480 965 7034; fax: +1 480 965 2751.
E-mail address: syrotiuk@asu.edu (V.R. Syrotiuk).

1570-8705/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2007.04.005

In the congestion control mechanism of TFRC, the destination measures the loss event rate and feeds it back to the source. The source also uses this feedback to measure the *round-trip time* (RTT). The loss event rate and RTT are used as input to TFRC's throughput equation, to calculate the acceptable transmission rate. The source then adjusts its transmission rate to match the calculated rate.

The throughput equation used by TFRC is a simplification of that used in TCP Reno [2]. It is a function of the factors of transmission rate, packet size, RTT, and the loss event rate. TFRC was designed for flows operating in an Internet environment and, as such, has shortcomings in its adaptation in mobile wireless networks [3].

Our first goal is to identify those factors and interactions that contribute most to the response variables of throughput, end-to-end delay, and jitter in MANETs. The factors considered are similar to those used in TFRC, with the addition of node speed. The mathematical and statistical techniques of *Design of experiments* (DoE) are used for screening the factors [4].

The regression equations obtained allow the characterization of average throughput, end-to-end delay, and jitter. Experimentation shows that while the terms in the equation defining the response surface do not change with network conditions, the coefficients of the terms do change. This motivates the use of machine learning to approximate the function with changing network conditions. We select a simple machine learning approach that is sufficient – an *artificial neural network* (ANN) – since it can approximate any continuous function [5].

We consider three topologies for the ANN. The first ANN topology has an input for each main effect, each interaction term, and intercept. It has no hidden layers and a single output neuron, thus it computes a linear function of its input [6, Chapter 4]. This topology corresponds most closely to the linear model derived from regression analysis in the factor study.

In the second topology, only the main effects are provided as input. One hidden layer with a single neuron is used, hence the ANN is capable of computing linear and nonlinear functions of the input [6, Chapter 4]. The topology of the third ANN is identical to the second except that the input is augmented to include the interaction terms and intercept. This topology computes a nonlinear function of the main effects and factor interactions.

The Audio Rate Cognition (ARC) transport protocol includes an audio streaming rate control mechanism driven by ANNs, with one ANN for each response variable. Using ANNs at the source and the destination of a multi-hop audio stream, we adjust streaming rates among discrete values. In simulation, we first compare the predictive accuracy of the three ANN topologies and find the third topology to be the most accurate.

Using simulation, we compare the ARC transport protocol using the third ANN topology to three well-known protocols: the *user datagram protocol* (UDP), the *real-time transport protocol* (RTP) without a control protocol, and a discrete variant of TFRC. In mobile ad hoc networks with several sources, ARC wastes significantly less bandwidth by reducing loss between 55% and 95% while satisfying requirements on end-to-end delay and jitter. As a result, ARC transmits higher quality audio, minimizing *root mean square* and Itakura–Saito spectral distances, as well as several parametric distance measures. In particular, ARC minimizes *linear predictive coding* cepstral (sic) distance, which closely correlates to subjective audio measures.

In summary, we make three contributions:

1. A screening experiment is used to quantify the contribution of factors and their interactions on the response variables of throughput, end-to-end delay, and jitter.
2. Machine learning is used to approximate these response variables as a function of the factors and interactions and use them to adapt the transmission rate. We select a simple approach that is sufficient, *artificial neural networks* (ANNs). Three ANN topologies are evaluated for predictive accuracy.
3. The ARC protocol equipped with the ANN topology that appears to be the most accurate is compared in simulation to competing protocols. ARC shows significant reduction in packet loss and increase in goodput, resulting in high quality audio transmissions.

The rest of the paper is organized as follows. Section 2 describes the factor study conducted. In Section 3 we describe the ARC protocol, focussing on the design of three ANNs topologies motivated by the factor study. An evaluation of the topologies is performed in Section 4. We provide the results of performance comparisons among ARC, UDP, RTP, and a discrete variant of TFRC in Section 5.