Link Classification and Residual Time Estimation Through Adaptive Modeling for VANETs

Nikoletta Sofra and Kin K. Leung
Electrical and Electronic Engineering
Imperial College London
London, UK
Email: {nsofra, kin.leung}@imperial.ac.uk

Abstract— **Vehicular *Ad hoc* Networks design has drawn a lot of research attention. High node mobility, an inherent characteristic of VANETs, results in short lived links and rapid topology changes; this makes traditional ad-hoc protocols inefficient for vehicle to vehicle communication. In this paper we propose a cross-layer approach, where physical layer information can be used by upper layers to monitor the well-being of the various links used, and to estimate their residual time. The algorithm proposed comprises of forming a time series based on physical layer measurements. Utilizing adaptive non-linear parameter estimation methods, this time series can be used to estimate the current state of the link and its residual time. The environment in which VANETs usually operate together with the effect of relative node movement introduce considerable noise. For this reason, a data-driven signal processing technique, Empirical Mode Decomposition is used for denoising. The proposed algorithms are tested against real data and simulations.**

Keywords- Vehicular Ad Hoc Networks, Link Quality, Link Residual Time, Mobility Prediction, Received Signal Strength, Empirical Mode Decomposition, Adaptive Modeling

I. INTRODUCTION

Enabling vehicle to vehicle and vehicle to roadside communication by employing the *Ad hoc* network paradigm is a recognizably challenging and equally rewarding scenario. The field of Vehicular Ad Hoc Networks is one that has been studied intensely among the wireless community ever since it emerged.

VANETs form a subset of Mobile *Ad hoc* Networks, but are distinguishable from MANETs in many ways, as they exhibit rapid topology changes, frequent fragmentation, relatively small network diameter and limited redundancy [1]. In this work, we aim to address the problems arising from the first feature, namely rapid topology changes.

By definition, the routing procedure is strongly linked to the topology of the underlying network, but it has been traditionally assumed that connectivity will be seamlessly dealt with by underlying layers. However, in the case of such high instability, upper layer protocols need to be able to adapt to these changes, and they can only be designed to do so via cross layer interactions. Classification of the physical links in terms of their stability and tendency for improvement or deterioration, as well as estimation of the residual time for which the link will remain efficiently active can give crucial information that can support routing and scheduling decisions. The unfavorable conditions under which VANETs are operating (outdoor environment, multiple scatterers, moving obstacles, Doppler effect) pose a challenge when attempting to derive this kind of information from physical layer measurements, as any data acquired are highly influenced by various sources of noise. Moreover, the received power metric will be numerically available only when packets are sent, and there will be a relatively small number of available samples, especially in the first iterations of an online algorithm. These issues have been already underlined in [2].

The use of link quality feedback to the routing protocol in terms of signal strength has been considered in many instances, such as [3-6]. Various methods to predict or study the link, or even route lifetime based on location and node movement information, or general mobility characteristics have been proposed, some examples being [7-9]. The contribution of this works lies in proposing a method that estimates the link quality in terms of stability and residual lifetime, based solely on physical layer measurements, and evaluating its efficiency by explicitly taking into account the unfavorable conditions under which such measurements are collected in the VANET scenario. This work is an extension to the work presented in [2], using adaptive, more robust parameter estimation methods, proposing link classification criteria, and being tested against real measurements.

II. ADAPTIVE MODEL FITTING

A. Outline

The method proposed utilizes knowledge available at the physical layer for estimating metrics that are meaningful for higher layers. This is done on a link basis, by collecting measurements of the signal strength of any packet received on

that link, and logging the time at which this was received. This includes both *hello* and data packets and does not incur extra overhead on the network. Packet by packet, a time series is formed; the method works online, iteratively, producing new estimates each time new data is received. As measurements at the physical layer are highly influenced by both instantaneous channel conditions and the movement between the nodes, a denoising process needs to be applied to the time series prior to any attempt to acquire estimates. A widely accepted model for outdoor signal propagation is adopted, and a non-linear data fitting method is used to fit the denoised time series to this model. Attention must be drawn to the fact that the time series evolves differently over distinguished phases (improving, deteriorating or stable channel condition) depending on the nature of the relative nodes' movement and the environment. Each phase can be best described by different parameters of the same model, which is why an adaptive algorithm, to recognize phase changes is developed.

B. Large Scale Fading

In the context described herein, we are interested in capturing the characteristics of large scale fading, namely the variations in the value of the received power that are caused mainly by physical movements of the nodes. Small scale variations, caused by scattering and the Doppler effect, as well as deep fades caused by moving obstacles or shadowing will be therefore treated as noise. Having derived the parameters of fading on the large scale, we can derive estimates of the link residual time, defined as the time period for which the *mean temporal value* of the received power will remain above a specified threshold, so that the link can be efficiently used for communication with a high probability. We can also evaluate the mean trend in the change of value of the received power in order to classify the links according to their stability.

For this reason we employ the well-known log-distance path loss model [10] for the received power:

$$P_r = P_t - (PL_0 + 10n\log_{10}(d))) + \varepsilon_i \qquad (1)$$

where PL_0 depends on the frequency, antenna gains and other system loss factors, P_t is the transmitting power, n is the path-loss exponent depending on the specific propagation environment (e.g. existence of obstructions), d is the physical separation distance and ε_i is the noise caused by small scale fading, shadowing and other random factors. Assuming that relative nodes' speed, u, and the path-loss exponent do not change significantly for a specific period of time, defined as a *phase* so that we can assume $d = ut + d_0$, where d_0 is the physical distance that seperates the nodes at the beginning of the phase, (1) yields h:

$$h : P_r = a_P + b_p \log_{10}(c_P t + 1)) + \varepsilon_i \qquad (2)$$

where α_P, b_P and c_P depend on either the path loss exponent factor or the relative speed (or both), so are phase-specific. Factors such as transmitting power, P_t, and system attributes affecting PL_0 are considered as constant for the duration of the link existence.

C. Denoising

As has already been stressed, it is of vital importance that a proper denoising process is applied to the collected samples. We employ a novel signal processing technique, *Empirical Mode Decomposition* (EMD) [11], suitable for non-stationary and non-periodically distributed time series. EMD is fully data driven, and it gradually dissects the time series into a case-dependent number of *intrinsic mode functions* (IMFs), going from finer to coarser temporal scale and thus from higher to lower frequencies. The signal described by the time series can be reconstructed by summation of all the IMFs. The interested reader can refer to [2, 11-14] for detailed descriptions. EMD's use as a filter has been well established [12-13]. Noise and channel fading will naturally occupy the higher frequencies, while changes on the large scale will reside on the lower ones. Since we are interested in large scale fading for this application, we can diminish higher frequencies, by only keeping the later extracted IMFs. Based on a criterion very similar to the one developed in [12], which does not assume any knowledge of the noise level, we may choose which IMFs to keep based on the change on the energy of each IMF. The IMF of the smallest energy, be IMF_j, is considered the threshold between noise-dominated IMFs and signal-component dominated ones. We therefore discard all the IMFs derived before IMF_j.

D. Adaptive parameter estimation

Let

$$S^{P_i} = \{s_{t_1}^{P_i}, s_{t_2}^{P_i}, \dots \qquad s_{t_n}^{P_i}\} \qquad (3)$$

be the set of denoised observations during phase P_i. If M is a non-linear data fitting method, then S^{Pi} can be used to produce estimates of α_P, b_P and c_P :

$$(\hat{a}_P, \hat{b}_P, \hat{c}_P) = M(h, S^{Pi}). \qquad (4)$$

In order to adaptively detect changes between phases, we keep track of the mean squared 2-norm of the residuals. As every time new data is added to S^P, M is repeated on the new set, acquiring new estimates of the parameters, we can calculate this metric for every t_i, and form a set accordingly:

$$R = \{\bar{r}_{t_1}^2, \bar{r}_{t_2}^2, \dots \qquad \bar{r}_{t_k}^2\}. \qquad (5)$$

Should a change of phase occur, the model derived through M will no longer be adequate to describe the time series described by S, as this is designed to handle data for a single phase only. Therefore, after a change of phase, the mean squared norms acquired will be significantly higher. For this reason, after each iteration, the statistics of R are observed. If

$$\bar{r}_{t_1}^2 > \overline{\bar{r}^2} + 3 \cdot \sigma(\bar{r}^2), \qquad (6)$$

where $\overline{\bar{r}^2}$ is the arithmetic mean of R and σ its standard deviation, and if the time period R is covering is bigger than T_P, a small enough interval to consider the environment and relative speed would not change dramatically, then a change of phase is considered to have been detected. The exact point of

change is determined among the members of **S** and all previous s_{ti}s are discarded.

E. Metrics derived

Having acquired the set of parameters in (2), we can classify the links according to their stability, and whether channel condition is improving or deteriorating, based on the estimate of $\hat{c}_P$.

Moreover, in the case that the link is deteriorating, we can estimate the time before the mean received power will drop below a specified threshold, γ, for efficient communication, thus an estimate of the link's residual time:

$$L\hat{R}T(t) = \frac{10^{\frac{\gamma - \hat{a}_P}{\hat{b}_P}} - 1}{c_p} - t, \qquad (7)$$

as well as an estimate of the received power at some time point in the future, t_f:

$$\hat{E}_r(t_f) = \hat{a}_P + \hat{b}_P \log_{10}(\hat{c}_P t_f + 1). \qquad (8)$$

III. PERFORMANCE RESULTS

A. Technical Details

We assume that physical layer information is passed on to the algorithm every T ms, in the form of the mean received power over all packets received on that link (if there exist any) during the according period. If T is sufficiently small, in the order of some hundreds milliseconds accuracy is not affected, as a node moving with normal vehicle speed will not have moved considerably during this period. On the contrary, an initial smoothing of the heavily noised time series incurs, while computational demand will be lower, as the algorithm needs only to be updated once, utilizing all the information acquired within the sampling period.

The implementation of Empirical Mode Decomposition used is as described in [12, 14]. The non linear parameter estimation method used is a subspace trust region method, based on the interior-reflective Newton method [15], and is implemented in the Optimization toolbox of Matlab [16].

As the link residual time is computed consecutive times during the phases when received power is deteriorating, we introduce convergence criteria on whether to accept the estimated value. Naturally, the more samples available, the better estimates are acquired, and there is a tradeoff between the mean accuracy of our estimations and the time before the first estimates are accepted.

B. Simulation Evaluation

We use simulation derived measurements to gain more insights on the effect different parameters will have on the efficiency of the method. Measurements are produced by considering a pair of nodes that move relative to each other. The relative speed between nodes is chosen explicitly, but is contaminated with additive white Gaussian noise (SNR=10). The threshold for efficient communication is -105 dBm, but we

assume that packets over -110 dbm can be detected by the receiver. In order to model the distance dependence of the received power, we attenuate the signal by assuming that the received power is declining proportional to da where a is the attenuation exponent factor and d is the distance between the sender and receiver. For small scale fading, we employ the Rayleigh fading channel model, and specifically, the Jakes channel model for the Doppler spectrum, using the frequency of IEEE 802.11g, (2.4 GHz). This particular scenario does not consider the existence of shadowing; however the trace used in IIIC gives initial evidence that the algorithm can respond well to the shadowing phenomenon.
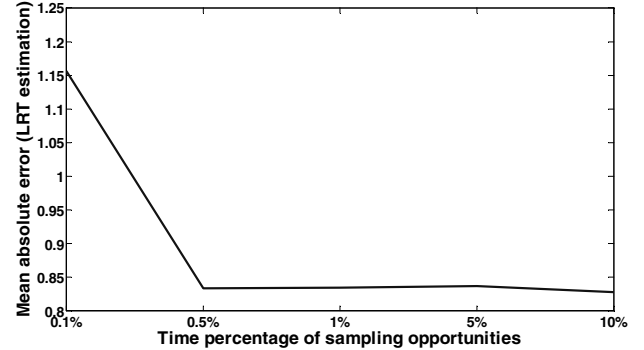

Figure 1. The effect of packet sending rate, as a percentage of time, on the accuracy of Link Residual Time estimation. The accuracy is expressed in terms of the mean absolute error, measured in sec.
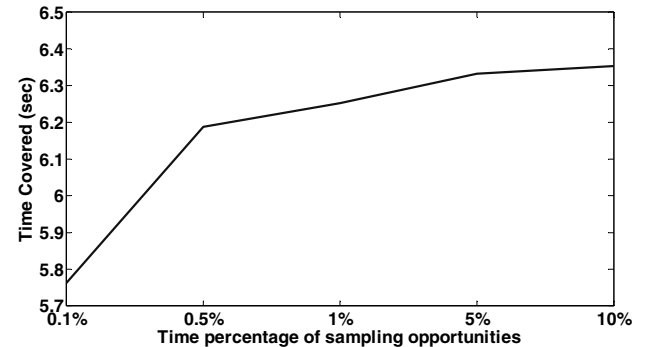

Figure 2. The effect of packet sending rate, as a percentage of time, on convergence

Obviously, the method depends on the number of packets that are being received, namely the percentage of time that we can actually sample the channel. For this reason we analyze the following scenario with simulation: the times that the packets are sent are drawn from a uniform distribution in such a way, so that the percentage of time that the link is active is fixed on average, and every packet that falls below the nominal threshold is discarded and not used by the method. This means that under certain circumstances, the rate of sent packets will not equal the rate of sampling opportunities for the channel, but we find the latter to be more realistic. Should the estimation of the residual time be negative, the method will consider the link as broken; this is possible as the threshold for efficient communication is different than the receipt threshold.
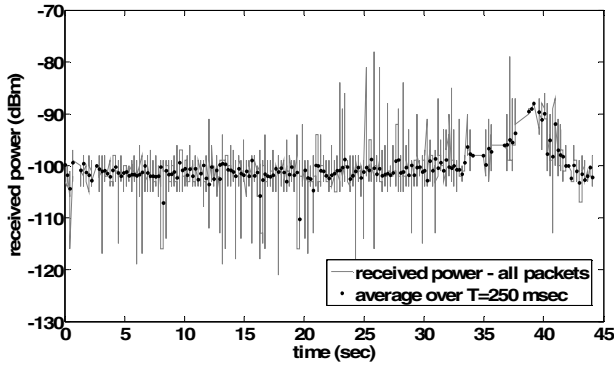
Figure 3. Received power of every packet and received power averaging over packets received within 250 ms, as given by the trace used.
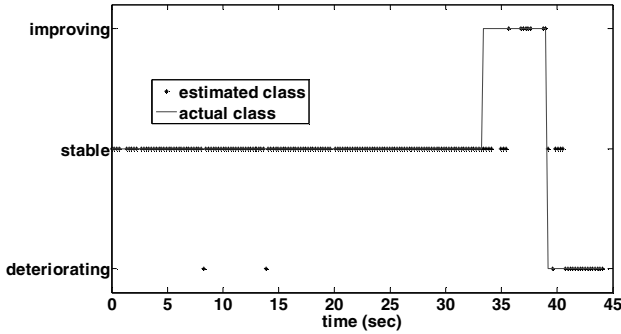

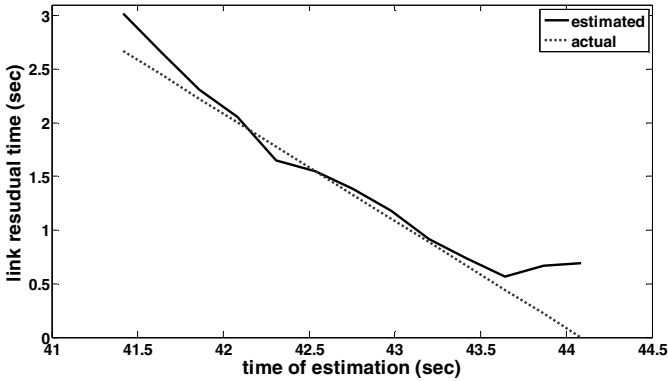Figure 4. Classification of link condition with respect to time for the trace.

Figure 5. Link residual time estimation for the trace. The mean absolute error is 178.163 msec

Figs 1 and 2 show the effect of the mean number of packets sent as a fraction of time on the proposed methods. The relative speed between the nodes is 15m/sec (contaminated with AWGN), and the parameter T is fixed at 250 msec. In fig. 1 we plot the mean absolute error in estimating LRT (Link Residual Time) over every step of the method (after convergence), for one hundred different realizations of the nodes' movement and the channel fading. We can see that the error is dropping as the

rate of sent packets increases, which is to be expected because more information is available to the method. However, with the exception of a drop between 0.1% and 0.5%, this decrease is not that significant, which shows that the method is robust with respect to the packet sending rate. What appears to be more significant is the effect the sending rate has on convergence. The method does not give an estimate of LRT as an output, unless the appropriate convergence criteria have been met. Fig. 2 shows the mean time during which the method is capable of giving estimations of LRT, having a mean absolute error shown in Fig.1, before link breakage occurs. We can see that if the link is used actively (transmission 10% of the time), LRT estimates are being issued, on average, for more than 6 seconds before link breakage with a mean absolute error well below 1 sec.

C. An illustrative example

In order to illustrate the method proposed, we use a trace acquired by a set available online [17]. The connection lasts approximately 45 seconds, and the vehicles move opposite to one another with speeds varying around 100 km/h, separated by 8 lanes of traffic and a concrete median.

Our study has shown that a preprocessing of the data on the physical layer before calculating the mean is beneficial. When estimating the mean of the received power over T msec, as has been explained previously, outliers, which are samples that have a low likelihood of being consistent with the rest of the data, are removed for the calculation of the mean. Outliers are defined as values that are at least three standard deviations away from the mean of this batch. The physical intuition behind this is that outlier preprocessing removes deep, sudden fades which do not extend to more than a few seconds, enabling EMD to correctly identify the fades caused by multiple scattering, nodes' movement and shadowing. Fig. 3 shows the received power with respect to time for all the packets received, as this is given by the traces, as well as the samples that will consecutively be available to the method with T=250 msec. From the figure we may see that there is a stable phase for approximately 33 sec. Fig. 4 shows the temporal classification of the link, based on the amplitude and sign of the estimate for $\hat{c}_P$ for all iterations. The classification is accurate for approximately 92.39% of the times. A lag is to be observed, as when a change of phase occurs, the adaptive method does not respond immediately. Finally, fig. 5 shows the estimated link residual time along with the real residual time. The mean absolute error for these estimations is 0.178 sec.

D. Discussion

In this work we have proposed a method for estimating two metrics; link stability and link residual time. Future average received power, i.e. future link quality can also be easily derived by this analysis. Numerous protocols that are capable of utilizing such knowledge have been developed [4,7…]. The main idea is that such metrics can be used as selection criteria when forming routes in reactive protocols, or saved and advertised as the table metrics in proactive protocols. The different weights assigned to these metrics depend on the strategy of the routing algorithm. For instance, one routing algorithm might favour a link that is likely to be of high quality

short term, but is not stable, in order to use the high rate that it offers, even for a small period of time, while another may choose a stable link of lower quality in order to reduce the overhead of re-establishing a link. It should also be mentioned that the metrics can be used in order to initiate handoffs (i.e. discovering a new route before the current one expires), or resolve scheduling conflicts (i.e. priority should be given to a link that is prone to break).

IV. CONCLUSION

In this paper we have presented a method that utilizes information available in the physical layer for estimating the residual time and the stability of the link between a given pair of nodes using signal processing and non linear parameter estimation methods when sparse samples of the received power are available. The proposed techniques are validated against a trace of actual vehicular communication and simulation results are presented to give more insights.

Future work includes utilization of the results of the method in the design of upper layer protocols and further study of the effect shadowing will have on the accuracy of the proposed algorithm, as well as analysis of different scenarios.

REFERENCES

[1] J.J. Blum, A. Eskandarian, L.J. Hoffman, "Challenges of intervehicle ad hoc networks," *Intelligent Transportation Systems, IEEE Transactions on* , vol.5, no.4, pp. 347-351, Dec. 2004

[2] N. Sofra, K.K. Leung, "Estimation of Link Quality and Residual Time in Vehicular Ad Hoc Networks," *IEEE WCNC, Las Vegas, April 2008.*

[3] T. Hsin-Mu, N. Wisitpongphan, O.K. Tonguz, "Link-quality aware ad hoc on-demand distance vector routing protocol," 1st Int. Symp. on *Wireless Pervasive Computing,* 2006, p.6.

[4] R. Dube, C. D. Rais, W. Kuang-Yeh, and S. K. Tripathi, "Signal stability-based adaptive routing (SSA) for ad hoc mobile networks," *IEEE Personal Commun.* [see also *IEEE Wireless Communications*], vol. 4, no. 1, pp. 36-45, 1997

[5] T. Goff, N. B. Abu-Ghazaleh, D. S. Phatak, and R. Kahvecioglu, 2001. Preemptive routing in Ad Hoc networks. In *Proceedings of the 7th Annual international Conference on Mobile Computing and Networking* (Rome, Italy). MobiCom '01. ACM, New York

[6] R. Chang, S. Leu, "Long-lived path routing with received signal strength for ad hoc networks," *Wireless Pervasive Computing, 2006 1st International Symposium on* , vol., no., Jan. 2006

[7] W. Su, S. Lee and M. Gerla, "Mobility prediction and routing in ad hoc wireless networks," *Int. J. Network Management*, vol. 11, no. 1, pp. 3-30, 2001.

[8] A.B. McDonald, T. Znati, "A path availability model for wireless ad-hoc networks," *Wireless Communications and Networking Conference, 1999. WCNC. 1999 IEEE*

[9] Y. Tseng, Y. Li, Y. Chang, "On route lifetime in multihop mobile ad hoc networks," *Mobile Computing, IEEE Transactions on* , vol.2, no.4, pp. 366-376, Oct.-Dec. 2003

[10] T.S. Rappaport, Wireless Communications: Principles and practice, 2nd Ed., Prentice Hall, 2002

[11] N. E. Huang, et al, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. of Mathematical, Physical and Engineering Sciences* vol. 454, no. 1971, pp. 903-995, 1998.

[12] G.Rilling, P. Flandrin, and P. Gonçalves, "On Empirical Mode Decomposition and its Algorithms," IEEE Signal Processing Letters, vol.11, no 2, p.112, 2004.

[13] A.-O. Boudraa, J.-C.Cexus, "EMD-Based Signal Filtering," *Instrumentation and Measurement, IEEE Transactions on* , vol.56, no.6, pp.2196-2202, Dec. 2007

[14] http://perso.ens-lyon.fr/patrick.flandrin/emd.html

[15] T.F. Coleman and Y. Li, "An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds," SIAM Journal on Optimization, Vol. 6, pp. 418-445, 1996.

[16] http://www.mathworks.com/

[17] R. M. Fujimoto, R. Guensler , M. P. Hunter, H. Wu, M. Palekar, J. Lee and J. Ko, CRAWDAD data set gatech/vehicular (v. 2006-03-15), downloaded from http://crawdad.cs.dartmouth.edu/gatech/vehicular