Quick construction of data-driven models of the short-term behavior of wireless links

Ankur Kamthe, Miguel A. Carreira-Perpinan and Alberto E. Cerpa
Electrical Engineering and Computer Science
University of California - Merced

***Abstract*—High-quality wireless link models can enable better simulations and reduce the development time for new algorithms and protocols. However, the models underlying current simulators are either based on too simple assumptions, so they are unrealistic, or are based on sophisticated machine learning techniques that require extensive training data from the target link, so they are more realistic but impractical. We consider the practical scenario where data collection time is limited (e.g. a few minutes) and cannot afford to deploy a testbed infrastructure with cabling, power and storage. We propose techniques that can construct an accurate machine learning model of the short-term behavior of a target wireless link given only limited training data for the latter, by adapting a reference model that was trained with abundant data. The parameters of the target model are a constrained transformation of the parameters of the reference model, thus the actual number of free parameters is much smaller, and can be reliably estimated with much less data. While estimating the target model from scratch requires 1 to 5 hours of target link data, we show our adaptation technique only requires under 3 minutes of data, for all packet reception rate regimes. We also show that we can construct adapted models for target links in different environments, packet sizes, interference conditions and radio technology (802.15.4 or 802.11b).**

I. Introduction

Wireless sensor network (WSN) links are variable due to multi-path effects, interference and transceiver hardware issues. Dynamic link changes can adversely affect the progress of in-network distributed processes, resulting in algorithm failure due to race conditions or inability to converge to a solution. Hence, for testing algorithms and protocols it is imperative to evaluate behavior under realistic conditions.

Performance evaluation is done using testbed deployments or with wireless simulators. Deploying a testbed in the target environment, involves cable installation for power and serial communication (for debugging), as the on-board batteries and memory are insufficient for long term data collection. This is a time-consuming and resource-intensive process, incurring significant human and capital expense.

Another way is wireless simulators that rely on link models to recreate real-world conditions. The chief advantages are repeatability and ease of debugging in a controlled setting. For realistic simulations, construction of a model for each new link in the target environment requires researchers to setup a physical testbed to collect *long* input traces to estimate the many model parameters. However, as explained earlier, this kind of data collection is time, labor and resource intensive.

Our goal is to use model adaptation to enable rapid construction of high quality link models using a few minutes of data using battery-powered sensor nodes and storing the traces in the on-board memory. This is more feasible compared to a full-fledged testbed deployment.

We assume that when a user collects data for short periods of time (a few minutes), the data is more likely to correspond to a regime where the link, although bursty, does not vary widely over time. In prior studies [9], such link regimes have been modeled using a Mixture of Multivariate Bernoulli (MMB) distributions, which captures the short-term correlations in packet reception behavior. In this paper, we present a novel model adaptation technique to estimate the MMB distribution parameters of a target link regime. Our proposed method transforms the parameters of a pre-existing reference MMB model, which captures some of the intermediate link burstiness properties, using the few minutes of data from the target link, collected by the user with minimal effort (e.g. just a pair of nodes, no deployment). We only need to estimate the parameters of the transformation, which are few compared to the parameters in the original MMB model. Therefore, the data required to estimate the parameters of the transformation (i.e., to adapt the model) is far smaller than to estimate the many parameters via retraining (the Expectation-Maximization (EM) [3] algorithm used to compute MMB parameters [5]).

The contributions of this paper are as follows: (1) We are the first to introduce adaptation techniques for data-driven models to the networking literature. (2) Demonstrate that the proposed adaptation technique requires only 3 minutes of data from the target link to estimate MMB models. This is significantly less than the 1-5 hours of data required for retraining. This changes the logistics of data collection, making it easy to impart realism in simulation thru use of adapted link models. (3) Show that the adapted model is quite close to the fully retrained model in terms of log-likelihood.

II. Background and Related Work

In this paper, our focus is adaptation techniques for packet loss models because they have been shown to accurately model burstiness of wireless links [9]. Prior studies [13], [7], [11], [4], [9] compute statistical link models using the long packet traces as input to estimate the numerous parameters of the respective models. This necessitated data collection in a testbed environment to have link models that can closely match communication characteristics of the real deployment.

In WSNs, a protocol or application may need to send a packet at any time. This requires the simulator to know the link conditions at all times, necessitating high frequency data collection for the link model. This results in large amounts of data being generated when collecting traces over long time periods. In turn, this forces testbed deployment, a resource-intensive and time-consuming process. However, the data collection should require least effort from the user, and at the same time the modeling approach should create realistic models by using little training data.

Mixture of Multivariate Bernoullis are widely used to model high-dimensional binary data [5], [9]. The multi-level Markov model [9] models the short-term correlations of wireless links using MMBs. However, it requires long traces ($\geq$1 hour) to estimate the numerous model parameters. In other domains such as speech recognition, the large training data problem has been solved by applying parameter-tying based model adaptation techniques [18]. [8] proposed a very restrictive model adaptation technique for MMBs, wherein all parameter values in a component increase/decrease non-linearly. This kind of adaptation is unsuitable for short term wireless link modeling because the correlations exist at smaller time-scales. In RoofNet [1], it was observed that bursty links showed correlated packet reception at a timescale of $\geq$1 second. In [17], analysis of wireless link RSSI and PRR variations showed that RSSI variations lead to changes in PRR and result in bursty behavior. In bursty links, these short-term correlations were observed to last typically for 500ms. In [16], bursty wireless links have been shown to display self-similarity over a time scale of 640ms. In the design of STLE [2], the threshold for identifying an intermediate link was set to 3 packets (inter-packet interval =250ms). This design choice had the underlying assumption (shown to be empirically correct) that intermediate links showed stable short-term behavior over a period of >750ms. Therefore, a good model adaptation approach should be able to model the complex correlations at these sub-second time-scales observed in prior studies.

III. Proposed Adaptation Approach

A. Terminology

Reference Link is the intermediate wireless link for which an MMB model has been created using extensive packet traces.
Target Link is the new wireless link for which we want to estimate a parametrized model.
Adaptation Data is the available data from a target link to estimate a model for it. This can be used (i) to retrain a model from scratch, initialized either randomly or to the same values as the reference model parameters or (ii) to adapt a reference model (the algorithm proposed in this paper).
Reference Model is a model estimated ahead of time using a large training set for a reference link.
Retrained Model is the model estimated from the adaptation data using the M&M approach.
Model adaptation is estimation of a new model by transforming an existing reference model to match the distribution of a different, and usually smaller adaptation dataset for a new (target) link.
Adapted Model is the model created using the proposed model adaptation approach.

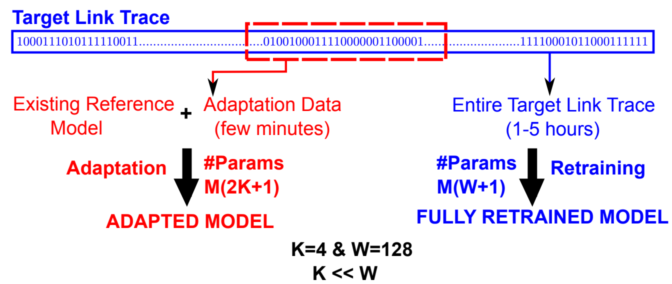


Fig. 1. Model Adaptation Overview: The target link trace is a binary sequence of 1's (successful packet reception) and 0's (lost or corrupted packet). Adapted model reuses an existing, well-trained reference model and adaptation data from target link. The proposed adaptation approach can perform similar to retraining with less training data as it has fewer parameters since $K \gg W$.

Figure 1 shows the proposed model adaptation overview.

B. Reference Model: MMB distribution

Given a data vector $\mathbf{x} \in \{0,1\}^W$ with W binary variables, the density of an MMB distribution is

$$p(\mathbf{x}) = \sum_{m=1}^{M} \pi_m p(\mathbf{x}|m) \quad p(\mathbf{x}|m) = \prod_{w=1}^{W} p_{mw}^{x_w}(1-p_{mw})^{1-x_w}$$

where there are M components and the parameters are the mixing proportions π_m (which are positive and sum to one) and the prototypes $\mathbf{p}_m \in [0,1]^W$. Thus, variables within a component are independent, but not across components. With enough components, an MMB can represent complex high-dimensional distributions. Given a training set, an MMB is usually trained with an Expectation-Maximization (EM) algorithm [3], [5]. The EM algorithm needs initial values for the parameters and can converge to local optima. In the context of adaptation, we will call *retraining* the process of estimating an MMB using this EM algorithm given the adaptation data, and initializing the parameters to those of the reference MMB.

C. Methodology

To capture the complex correlations at sub-second time scales, we propose tying (K) consecutive groups of reference MMB parameters within each reference MMB component. Each group is of length W/K and uses a different sigmoid transformation. Having multiple transformations acting on smaller groups of reference MMB parameters allows for independent variations between these groups. The added flexibility with more parameters facilitates better modeling of short-term correlations. The adapted model parameters ($\widetilde{p}_{mw}$) are:

$$\widetilde{p}_{mw} = \sigma(p_{mw}; a_{mi}, b_{mi}) = \frac{1}{1+e^{-(a_{mi}p_{mw}+b_{mi})}}$$
$$w = (i-1)\frac{W}{K}+1, \ldots, i\frac{W}{K}, \ i = 1, \ldots, K.$$

For each component, we have a vector $\mathbf{a}_m = \{a_{m1}, \ldots, a_{mK}\}$ and $\mathbf{b}_m = \{b_{m1}, \ldots, b_{mK}\}$ of sigmoid transformation parameters. Our approach establishes a continuum between retraining (all parameters are free, $K = W$, like in the M&M model) and the adaptation approach from [8] (all prototype dimensions are transformed with the same sigmoid, just 2 free parameters,

$K = 1$). The mixture proportions, one per component, are free during adaptation subject to the constraint that they sum to 1. Our approach with multiple sigmoids has $M(2K+1)$ free parameters (mixing proportions and sigmoid parameters). This is still less than the $M(W+1)$ parameters that are to be estimated using retraining because $K \ll W$. We estimate these parameters using a generalized EM algorithm.

D. Approach Illustration

The flexibility of our approach is easily seen with the help of an illustration. Consider MMB distributions with $W = 4$ and $M = 1$. Suppose the reference model parameters are $\mathbf{p}_R = [.1\ .9\ .1\ .9]$. and the adaptation data for the target link comes from a target distribution $\mathbf{p}_T = [.9\ .1\ .1\ .9]$. We see that the first two parameters must be inverted while the last two should remain the same. By adapting the MMB using one sigmoid transformation per component [8], we get $\widetilde{\mathbf{p}}_{K=1} = [.5\ .5\ .5\ .5]$. By applying multiple sigmoids, one to the first two parameters ($a_{m1} = -6, b_{m1} = 3$) and a second to the last two parameters ($a_{m2} = 6, b_{m2} = -3$) within $\mathbf{p}_R$, we get a new adapted model $\widetilde{\mathbf{p}}_{K=2} = [0.92\ 0.08\ 0.08\ 0.92]$ We can clearly see that $\widetilde{\mathbf{p}}_{K=2}$ is a better estimate of $\mathbf{p}_T$ compared to $\widetilde{\mathbf{p}}_{K=1}$, showing the greater flexibility when using multiple sigmoid transformations to the reference MMB parameters. Such local transformation approaches have demonstrated better performance compared to global transformations [8] in other domains [15].

E. Generalized EM (GEM) algorithm for adaptation with multiple sigmoids per component

The new objective function is the log-likelihood (LL) of the adaptation data is as follows:
$L(\Theta) = \sum_{n=1}^{N} \log \sum_{m=1}^{M} \widetilde{\pi}_m p(\mathbf{x}_n; \mathbf{a}_m, \mathbf{b}_m) - \lambda(R(\mathbf{a}) + R(\mathbf{b}))$
where $\Theta = \{\widetilde{\pi}_m, \mathbf{a}_m, \mathbf{b}_m\}_{m=1}^{M}$ are the $M(2K+1)-1$ free parameters of the constrained MMB model. $p(\mathbf{x}_n; \mathbf{a}_m, \mathbf{b}_m)$ is a multivariate Bernoulli with $\widetilde{\mathbf{p}}_m = \sigma(\mathbf{p}_m; \mathbf{a}_m, \mathbf{b}_m)$ ($\mathbf{a}_m = a_{m1}, \ldots, a_{mK}$ and $\mathbf{b}_m = b_{m1}, \ldots, b_{mK}$). $R(\mathbf{a})$ is the variance of all the a's $\in \{\mathbf{a}_1, \ldots, \mathbf{a}_M\}$ and similarly for all b's. This is a commonly used regularization term [6]. Here the term involving λ penalizes LL to prevent $\mathbf{a}_m$ and $\mathbf{b}_m$ within each component from approaching very high (extreme) values.

We estimated the mixture proportions and sigmoid parameters using an EM algorithm. EM is an iterative algorithm that alternates between performing an expectation (E) step and a maximization (M) step, till the algorithm converges to a local minima. In the E–step, the posterior probabilities are computed from the current parameter estimates, and, in the M–step, the current parameter estimates are updated using the posterior probability computed in the E–step. In EM, instead of directly maximizing the objective function, an auxiliary function Q is defined. Iteratively maximizing the auxiliary function by estimating values for the model parameters increases the value of the objective function. In a typical EM algorithm, the M-step where we update our parameter value is closed form. However, the nonlinear sigmoid transformation makes the M step not solvable in closed form for $\{\mathbf{a}_m, \mathbf{b}_m\}$. Hence, we provide a generalized EM algorithm [12]. We make use of BFGS [14], a quasi-Newton numerical optimization algorithm with super-linear convergence, to estimate new values for $\{\mathbf{a}_m, \mathbf{b}_m\}$. The E step is analogous to that of the EM algorithm for MMBs. In short, our algorithm for estimating $\{\mathbf{a}_m, \mathbf{b}_m\}$ is comprised of an outer loop of E- and M-steps, and within each M-step there is an inner loop of BFGS iterations.

E step: Computes $r_{mn}^{\tau} = p(m|\mathbf{x}_n; \widetilde{\pi}_m^{\tau}, \mathbf{a}_m^{\tau}, \mathbf{b}_m^{\tau})$, the posterior probability of component m given data point $\mathbf{x}_n$ and the parameter estimates in iteration τ:

$$r_{mn}^{\tau} = \frac{\widetilde{\pi}_m \prod_{w=1}^{W} (\widetilde{p}_{mw}^{\tau})^{x_{nw}} (1 - \widetilde{p}_{mw}^{\tau})^{1-x_{nw}}}{\sum_{m'=1}^{M} \widetilde{\pi}_{m'} \prod_{w=1}^{W} (\widetilde{p}_{m'w}^{\tau})^{x_{nw}} (1 - \widetilde{p}_{m'w}^{\tau})^{1-x_{nw}}}$$

M step: Maximize an auxiliary function, Q defined as follows:

$$Q(\Theta^{\tau+1}; \Theta^{\tau}) = \sum_{n=1}^{N} \sum_{z_n=1}^{M} r_{mn}^{\tau} \log \left(p(z_n; \widetilde{\pi}_{z_n}) p(\mathbf{x}_n | z_n; \mathbf{a}_{z_n}^{\tau+1}, \mathbf{b}_{z_n}^{\tau+1}) \right) - \lambda(R(\mathbf{a}) + R(\mathbf{b}))$$

The first term is the expected complete-data log-likelihood, over π_m and $\{\mathbf{a}_{mi}, \mathbf{b}_{mi}\}_{i=1}^{K}$ for each component. $1 \le z_n \le M$ is the (unknown) index of the mixture component that generated data point $\mathbf{x}_n$. The second term involving λ is for regularization. The mixing proportions are updated as below:

$$\widetilde{\pi}_m^{\tau+1} = \frac{1}{N} \sum_{n=1}^{N} r_{mn}^{\tau}$$

The gradient of Q w.r.t. a_{mi}, b_{mi} $(i \in \{1, \ldots, K\})$ is:

$$\frac{\partial Q}{\partial a_{mi}^{\tau+1}} = \sum_{n=1}^{N} r_{mn}^{\tau} \sum_{w} p_{mw}(x_{nw} - \widetilde{p}_{mw}) + \frac{2\lambda(a_{mi}^{\tau} - E(\mathbf{a}_m))}{K}$$

$$\frac{\partial Q}{\partial b_{mi}^{\tau+1}} = \sum_{n=1}^{N} r_{mn}^{\tau} \sum_{w} (x_{nw} - \widetilde{p}_{mw}) + \frac{2\lambda(b_{mi}^{\tau} - E(\mathbf{b}_m))}{K}$$

where w goes from $(i-1)W/K$ to iW/K. We solve for $a_{mi}^{\tau+1}, b_{mi}^{\tau+1}$ using BFGS.

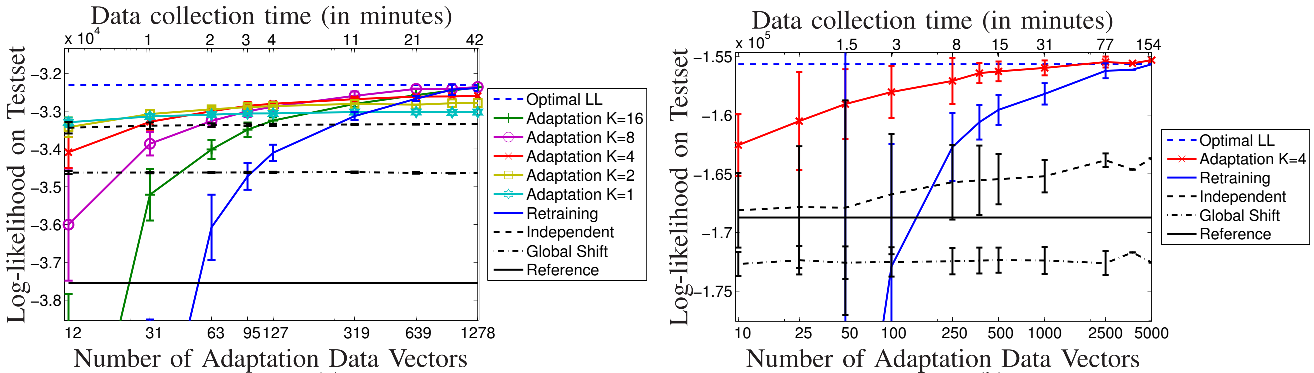
IV. Results

A. Data Collection

We collected comprehensive packet reception traces using 802.15.4 compliant CC2420 radios in different environments, transmission power levels and interference conditions. In addition, we used an 802.11 dataset submitted to the CRAWDAD repository [10] to test our approach under different packet sizes and transmission power that varies significantly from 802.15.4 radios. In all cases, we have single transmitter and multiple receivers logging the packet reception traces.

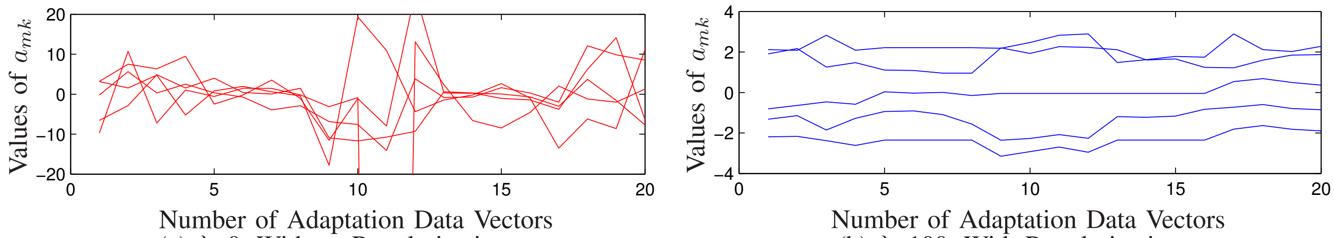
B. Methodology

For adaptation, one needs a good reference model. We treated *one* of the links from the Indoor dataset as the reference link. The reference link data trace is comprised of 1's and 0's, indicating packet reception and loss, respectively. The binary input trace ($\geq$230,400 packets) is split into smaller sequences, each of length $W = 128$, and clustered in two groups to separate sequences based on PRR (*link regimes*). This tends to group sequences associated with similar PRR values. For each group, using all input sequences, we used the EM algorithm for MMBs as described in [5], for estimating reference MMB model parameters, namely, the proportion and Bernoulli W-dimensional vector for each of the M mixture components.



(a) (b)

Fig. 2. Log-likelihood (LL) when adapting a reference MMB from a 802.15.4 reference link (packet = 28 bytes, Tx power = -11dBm) from Indoor testbed using adaptation data from (a) 802.15.4 target link from Motelab (packet = 28 bytes, tx power = 0dBm) and (b) 802.11 target link from a NIIT (1000 byte payload, with interference). The "optimal LL" is the LL of the retrained model with 100% (all available) adaptation data.



(a) λ=0, Without Regularization (b) λ=100, With Regularization

Fig. 3. Variation in values of a_{mk} and b_{mk} ($M = 5$ and $K = 4$,i.e., 20 a_{mk} and 20 b_{mk}) for a representative adapted model constructed with 45 adaptation data vectors. Sequences of a_{mk} and b_{mk} for multiple adaptation datasets are overlapped in the figure. Without regularization ($\lambda = 0$), the values of a_{mk}, b_{mk} change quite drastically. Note that in (a) the y-axes in truncated between -20 and 20. With regularization ($\lambda = 100$), the values vary much less.

All other links in the Indoor, MoteLab, Outdoor and 802.11 datasets will be treated as target links. Similar to the reference link, for each target link, we convert the binary input sequence into vectors of length $W = 128$. We term these binary sequences as *adaptation vectors* or *adaptation data*. The sequences are separated into two clusters, wherein each cluster has sequences with similar PRR values. The adaptation data for each link regime are divided into training set (used for estimating the adapted and retrained model) and validation sets (for evaluating the model LL) with a 70-30 split.

In addition to the adapted and retrained models, we compare with traces generated from the following models:

Independent model is constructed for each state of the training set link with $p = PRR(AdaptationData)$.

Reference Model MMB that is closest to the target link state in terms of PRR.

Global Shift Model is created by modifying the Bernoulli parameters from MMB distribution (all increase or decrease) of each state of the reference MMB to match the PRR of the adaptation data (Section 3.7.1 of the M&M paper). The problem with this approach is the same as with the linear transformation. Once any of the p-values reach 1/0, it prevents less extreme values from adapting.

For both the adapted and retrained models, we vary the amount of data used for estimating the model from 100% (all adaptation data) to 1%. This helps identify how the quality of the models degrade as a result of decreasing amount of training data. In our evaluation, we assume that adaptation data is available for all link regimes states when creating the different MMB models. For each of the link regimes in the target links, we compare testset LL of the adapted, retrained, independent and global shift MMB models in Section IV-C.

C. Model Comparisons

Figure 2 shows the variation in LL of the retrained and adapted MMB models, for target links from two different testbeds, packet sizes and transmission power, as a function of the amount of adaptation data used for estimating the model parameters. This figure is representative of behavior seen across all our target links when we vary K and the amount of adaptation data. The LL of the adapted models is significantly better than that of the corresponding retrained model when the number of adaptation data vectors is small. As we increase the amount of adaptation data used for training, the LL of the adapted model approaches the optimal LL. The "optimal LL" is the LL of the fully retrained model constructed using 100% (all available) adaptation data. The retrained model converges to the optimal as more adaptation data is used, but it needs much more data to achieve a comparable LL to the adapted model; and its performance is more variable (large error bars). For small amounts of adaptation data, the retrained model parameters converge to a local optima and, hence, it performs worse on the test set than even the reference model. Given the same amount of adaptation data, the adapted model parameters generalize better than the corresponding retrained model.

In Figure 2(a), we adapt the reference model to construct an MMB model for 802.15.4 target link from MoteLab. We observe that adaptation with $K = 1$ stagnates because the transformation is very restrictive since it is tying $W = 128$ parameters. Consequently, it is not flexible enough to model the variations observed in intermediate quality target link data. With $K > 1$, the model has more parameters, and hence greater flexibility to model. One has to be careful about adding many more transformations (i.e. tying fewer parameters) which can cause over-fitting. At the same time the transformations should be capable of changing small groups of reference MMB parameters to model the short term correlations that occur at sub-second time scales. $K = 1$ affects all Bernoulli parameters within an interval of 2seconds ($W/K = 128$). Higher K's reduce this interval by tying smaller groups of parameters. In [17], [16], it has been shown that these short-term correlations change after 500-640ms. This means that we need to tie groups of parameters that affect link quality over similar time intervals. Any $K \geq 4$ satisfies this requirement. However, with too many sigmoids (e.g. $K = 8, 16$), the adaptation approach performs worse on the test set with less adaptation data ($< 10\%$). Typically, adaptation approaches are attractive when they do not require extensive data collection. So, $K > 4$ are not feasible as they require more adaptation data to come close to optimal LL. Therefore, from the LL results and other factors such as data collection and short-term correlation interval requirement [17], [16], we determined the optimal K to be 4. For the MMB models shown in Figure 2(a), with $K = 4$, and ≥ 45 adaptation data vectors, the adapted MMB model does better, in terms of LL, than the one with $K = 1$. For simulating traces and comparing models, we use these values ($K = 4$ and 45 adaptation data vectors) for constructing adapted MMB models for each link regime for all target 802.15.4 links. So, adapted models require 90 adaptation data vectors or 3 minutes of target link data to model it with 2 link regimes. Figure 2(a) also shows the LL of the independent, reference and global shift MMB distribution. It is worse (to varying degree) as compared to adaptation. In Figure 2(b), we adapt the reference model to construct an MMB model for 802.11b target link from NIIT testbed. Using greater than 50 (0.1%) adaptation vectors (and $< 10\%$ data), the LL of the adapted model with $K = 4$ is greater than the corresponding retrained, global shift, reference and independent models.

Due to space constraints, we do not show the full regularization term evaluation with varying λ. Figure 3 summarizes the effect of the regularization term. When the regularization term is absent ($\lambda = 0$), a_{mk}, b_{mk} can take high positive/negative values for adapted MMB models constructed with ≥ 45 adaptation vectors (see Figure 3(a)). This implies that the sigmoids are saturating so as to make all the adapted model p_{mw}'s close to 1/0. This results in over-fitting. In contrast, when using our proposed model adaptation technique with all available (100%) adaptation data, a_{mk}, b_{mk} do not take high positive/negative values. Figure 3(b) shows that by using $\lambda = 100$, we can prevent a_{mk}, b_{mk} from taking high values by imposing a strong penalty on the model LL.

This demonstrates that the proposed model adaptation procedure can perform similar to a fully retrained model, given a well-trained reference model and little adaptation data.

V. Conclusion

Statistical learning techniques using real data make it possible to create realistic wireless link models. In this paper, our goal was to enable simulation users to compute high-quality short term wireless models without extensive data collection. We have shown that by tilizing a total 3 minutes of data from a target link, we can adapt a reference MMB using multiple sigmoid transformations to model target link MMB distributions. This is an order of magnitude decrease in the amount of data required to train a high quality model to capture the short term correlations of a link. The main advantage is having realistic simulations with minimal user data collection that allows extensive testing of protocols and applications in a controlled environment reducing the debugging effort in the field. Also, our technique can adapt to target links under different environments, packet sizes, transmission power levels and radio data rates.

Acknowledgements

Work funded in part by NSF award IIS-0711186, CNS-0923586 and CNS-1254192, and CITRIS awards SPF-81 and SPF-165.

References

[1] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. Link-level measurements from an 802.11b mesh network. *SIGCOMM Comput. Commun. Rev.*, 34(4):121–132, 2004.

[2] M. H. Alizai, O. Landsiedel, J. Ágila Bitsch Link, S. Götz, and K. Wehrle. Bursty traffic over bursty links. In *SenSys '09*.

[3] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stats.*, 41(1):164–171, 1970.

[4] G. Boggia, P. Camarda, and A. D'Alconzo. Performance of markov models for frame-level errors in IEEE 802.11 wireless LANs. *Int. Journal of Comm. Sys*, 22(6):695–718, 2009.

[5] M. Á. Carreira-Perpiñán and S. Renals. Practical identifiability of finite mixtures of multivariate Bernoulli distributions. *Neural Computation*, 12(1):141–152, Jan. 2000.

[6] M. Farhadloo and M. Á. Carreira-Perpiñán. Regularising an adaptation algorithm for tongue shape models. In *ICASSP*, 2012.

[7] C. Jiao, L. Schwiebert, and B. Xu. On modeling the packet error statistics in bursty channels. *LCN '02*.

[8] A. Kamthe, M. Á. Carreira-Perpiñán, and A. Cerpa. Adaptation of a mixture of multivariate bernoulli distributions. In *IJCAI '11*.

[9] A. Kamthe, M. Á. Carreira-Perpiñán, and A. Cerpa. M&M: Multi-level markov model for wireless link simulations. In *SenSys '09*.

[10] S. A. Khayam and Y. Cho. CRAWDAD trace set niit bit errors 802.11 (v. 2008-07-08), July 2008.

[11] A. Kopke, A. Willig, and H. Karl. Chaotic maps as parsimonious bit error models of wireless channels. In *INFOCOM '03*.

[12] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley, 2008.

[13] G. T. Nguyen, R. H. Katz, B. Noble, and M. Satyanarayanan. A trace-based approach for modeling wireless channel behavior. In *WSC '96*.

[14] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2006.

[15] C. Qin, M. Á. Carreira-Perpiñán, and M. Farhadloo. Adaptation of a tongue shape model by local feature transformations. In *INTERSPEECH*, 2010.

[16] T. Rusak and P. Levis. Burstiness and scaling in the structure of low-power wireless links. *SIGMOBILE Mob. Comput. Commun. Rev.*, 13:60–64, June 2009.

[17] K. Srinivasan,, M. A. Kazandjieva,, S. Agarwal,, and P. Levis,. The β-factor: measuring wireless link burstiness. In *SenSys '08*.

[18] P. C. Woodland. Speaker adaptation for continuous density HMMs: A review. Adaptation Methods for Speech Recognition, ISCA Tutorial and Research Workshop (ITRW), 2001.