Evaluation of short-term traffic forecasting algorithms in wireless networks

Maria Papadopouli[a,c] Elias Raftopoulos[c] Haipeng Shen[b]
Emails: maria@cs.unc.edu, eraftop@csd.uoc.gr, haipeng@email.unc.edu.

Abstract

Our goal is to characterize the traffic load in an IEEE802.11 infrastructure. This can be beneficial in many domains, including coverage planning, resource reservation, network monitoring for anomaly detection, and producing more accurate simulation models. We conducted an extensive measurement study of wireless users on a major university campus using the IEEE802.11 wireless infrastructure. We proposed several traffic models that take into account the periodicity, recent traffic history, and flow-related information and based on them designed traffic forecasting algorithms. Finally, we evaluated these algorithms and the impact of several parameters on the prediction accuracy.

I. Introduction

Popular applications and services from the wired networks shift in the wireless arena and new applications are increasingly being deployed. The proportion of wireless steraming audio/video traffi c increased by 405% between 2001 and 2003/2004, P2P from 5.2% in 2001 to 19.3% in 2003/4, fi lesystems from 5.3% to 21.5%, and streaming from 0.9% to 4.6%. All these applications not only generate higher trafffi c load but also impose additional quality of service requirements. Currently APs do not perform any type of forecasting or admission control and clients frequently experience failures and disconnections when there is high demand in the wireless infrastructure. Empirical studies and performance analysis indicate dramatically low performance of real-time constrained applications over wireless LANs (such as [2] on the VoIP).

While in several cases over-provisioning in wired networks is acceptable, it can become problematic in the wireless domain. Furthermore, wireless clients have more vulnerabilities than their wired counterparts. Their energy limitations, dynamic characteristics, and mobility of the wireless clients impose additional constraints and the bandwidth utilization at an AP can impact their performance substantially. For designing adequate quality of service provision, capacity planning, load balancing, and network monitoring, it is critical to understand the traffi c characteristics. While there is a rich literature characterizing traffi c in wired networks ([12], [11], [17], [5]), there are only a few studies available that examined wireless traffi c load and even fewer studies on short-term wireless traffi c forecasting. Each AP predicts its traffi c load for the next time interval (e.g., next hour or fi ve minute interval) and use their traffi c load forecasts during admission control to not only better manage their traffi c demand but also advice clients to associate with the appropriate APs to better utilize their local resources. Such predictions can be used to reduce the energy spendings at the client side, improve the capacity utilization of wireless LANs, and better load balance the traffi c.

In this paper, we study a large wireless infrastructure[1] and model the traffi c load at APs. Based on these traffi c load models, we design forecasting algorithms to predict the traffi c load at APs in different time-scales. We then apply them on real traffi c traces acquired from the most heavily utilized APs and evaluate their performance. In [14], we showed that hourly predictions of the traffi c load at an AP have very large prediction error due to the high variability. In this work, we focus on fi ner time scales and extend that work by (a) presenting and evaluating a number of new forecasting algorithms, such as the adaptive moving-average (adaptive MA) and flow based algorithms, (b) integrating different types of information in the prediction algorithm (e.g., snmp- and tcp-based), (c) dramatically improving the forecasting accuracy. We found that the time

a. Department of Computer Science, University of North Carolina at Chapel Hill, US. This work was partially supported by the IBM Corporation under an IBM Faculty Award 2004/2005 grant.

c. Institute of Computer Science, Foundation for Research and Technology-Hellas (FORTH), Greece.

b. Department of Statistics & Operations Research, University of North Carolina at Chapel Hill, US.

granularity and the size of the recent traffic history have dominant impact on the prediction accuracy. That is, the finer the time granularity (5minute vs. hour) and more recent the historical traffic data is, the larger their impact on the prediction error. To the best of our knowledge, these research efforts are the first traffic forecasting studies on large IEEE802.11 infrastructures.

Section II describes briefly the wireless infrastructure at UNC, and data acquisition process. Section III presents several main traffic models and forecasting techniques and Section IV discusses their performance evaluation. In Section V, we discuss previous related research and Section VI summarizes our main results and future work plans.

II. Background and Data Acquisition

The IEEE802.11 infrastructure at the University of North Carolina at Chapel Hill provides coverage for 729-acre campus and a number of off-campus administrative offices. The university has 26,000 students, 3,000 faculty members, and 9,000 staff members. Undergraduate students (16,000) are required to own laptops, which are generally able to communicate using the campus wireless network. A total of 488 APs were part of the campus network at the start of our study. We acquired two types of data, namely, data collected using SNMP and packet header collected from the link between the university and the rest of the Internet.

The main source of data was collected using the Simple Network Management Protoc ol (SNMP), the most widely available monitoring service in wireless platforms. Any AP in the market supports monitoring using SNMP, so it is important to understand how much operators and researchers can learn from SNMP data. For the SNMP-based data, every AP on campus was polled every five minutes. We developed a custom data collection system, being careful to avoid the pitfalls described in [7]. First, the system was implemented using a non-blocking SNMP library for polling each AP precisely every five minutes in an independent manner. This eliminates any extra delays due to the slow processing of SNMP polls by some of the slower APs. The system ran in a multiprocessor system and the CPU utilization in each of the three processors we employed never exceeded 70%. Second, our characterization of the workload of the APs is derived only from those clients associated with the AP at polling time (and not from roaming ones associated with a different AP). More detailed information about the testbed and wireless infrastructure can be found at [14]. Based on the SNMP trace for each AP, we produce a time series of its traffic load at hourly and five-minute intervals. This traffic is the total amount of bytes received and sent from all clients that were associated with the AP at that time interval. In the rest of the paper, depending on the mathematical expression, we will use two notations for these time series. Specifically, the traffic of the AP i during the h-th hour of day d, that corresponds to t-th time interval, is $X_i(t)$. The forecasted value is indicated as $\hat{X}_i(t)$ When we do not refer to a specific AP, the subscript of X or $\hat{X}$ is ommitted. The SNMP allows us to query for both aggregate and client-based data. The latter corresponds to the traffic that clients associated with the given AP had accessed during the polling interval.

The packet header traces consist of a total of 175GB and collected between 19:06 PM on Wednesday April 13rd ,2005, and 05:18 AM on Thursday April 21st, 2005 resulting in a continuous trace of 178.2 hours. Packet headers were acquired using a high-precision monitoring card (Endance's DAG 4.3 GE) attached to the receiving end of a fiber split. The card was installed in a high-end FreeBSD server. Neither the server nor the card's driver reported any failures or packet drops during the monitoring. We focus our analysis on the **hotspot APs**, namely the APs with the highest traffic load demand. We describe the criteria for their selection in detail in [14].

III. Traffic forecasting methodology

We distinguish three main categories of forecasting algorithms based on their traffic models, namely, the periodic-based, the AR-based, and the hybrid algorithms. The first category exploits the periodicities (e.g., diurnal and weekly patterns) in the traffic by incorporating historical means of traffic. The second type considers only a window of recent traffic history (e.g., the traffic during the last three hours). The window size could be fixed throughout the forecasting process or dynamically adjusted. In the latter case, the algorithm monitors the traffic dynamically, detects its prominent changes (i.e., level shifts) decides about the window size (amount of traffic history to be used) and applies the prediction algorithm using this window of traffic. Examples of such algorithms are the moving average-based algorithms described in the following paragraphs. When the traffic load exhibits both strong periodicities and temporal dependencies, hybrid models can be better choices. The classification can be extended depending on whether the algorithm uses the *predicted* or *actual* values of the recent history. More specifically, a forecasting algorithm is *one-step ahead,* when it uses the *actual* values of traffic for the traffic during the recent history window and *multi-step ahead* when it employs the predicted ones. To enhance the performance of the forecasting algorithm, we may use additional information about the type of traffic, application, and user profile. Depending on the different types of data used, a forecasting algorithm can be further classified into single-source or multi-source algorithm. For example, a forecasting algorithm using SNMP-based data and tcp-packet headers is a multi-source algorithm. Our general methodology consists of the following steps: (A) Time-series extraction, data cleaning, and treatment of missing values; (B) Power spectrum and partial autocorrelation analysis; (C) Traffic load modeling and forecasting using the traffic load models.

A. Time-series extraction and treatment of missing values

While our monitoring system requested traffic load information from each access point precisely every five minutes, missing values are relatively frequent in our dataset. They are due to several reasons: (1) an access point may be down for maintenance, or in the middle of an accidental reboot; (2) an access point may be too busy to reply to an SNMP query; (3) the network path between our monitor and the access point may be temporarily broken; and (4) query packets and response packets may be lost (they are transported using UDP). While these pathologies are expected to be infrequent, our dataset is large enough to contain numerous instances of each of them. Thanks to the cumulative nature of SNMP counters, we were able to reconstruct missing values quite accurately. The basic technique for extracting an equally-spaced time-series $X = \{x_1, x_2, \ldots, x_n\}$ from SNMP data is to subtract the cumulative counters from two consecutive *polling* operations. In [14], we describe in detail the algorithms for the detection and treatment of missing values and reboots.

B. Power Spectrum and partial autocorrelation analysis

There are hotspot APs whose traffic load exhibits strong diurnal periodicity. In the hourly time series of the traffic load for several hotspots, the most dominant period is the 24-hour one, with smaller ones corresponding to 12 hours (day/night), and 168 hours (weekly period). 10 out of the 19 hotspots have a clear spike at 24 hours/cycle and do not have a high frequency variation. Also, some APs have weekly patterns at around 168 hours/cycle. There are other APs with no clear periodic pattern, for which there is little prediction power among the historical data. Further smoothing does not appear to be helpful, at least with our current relatively short traces.

C. Modeling for Traffic Forecasting

1) Periodic-based forecasting: We propose several models that facilitate the diurnal and weekly periodicity of the traffic load. Let us first consider hourly traffic time series. We define the **historical mean hour** (PH) traffic of an AP as the mean of the traffic during that hour for each day in the history of that AP (N_{days} days). Similarly, the **historical mean hour-of-day** (PHoD) traffic is the mean of the traffic at such hour of day in the history of that AP. We taylor two simple models based on the historical mean hour and mean hour-of-day. Such periodic models can be extended for finer time scales (e.g., five-minute time series).

2) Recent history based algorithms: We apply some simple linear predictors based on localized regression. They are less demanding than more complex linear predictors, such as ARIMA where selecting the order and coefficiants requires a large amount of previous historical data. The linear models used are moving average and exponentially moving average.

A **simple moving average** (MA) is the unweighted mean of the previous w data points in the time series. $\hat{X}(t+1) = \frac{1}{w}\sum_{k=t-w+1}^{t} X(k)$

A **weighted moving average** is a weighted mean of the previous w data points in the time series. A weighted moving average is more responsive to recent movements than a simple moving average.

An **exponentially weighted moving average** (EMA) is an exponentially weighted mean of previous data points. The parameter α of an EMA can be expressed as a proportional percentage. For example, in a 10% EMA, each time period is assigned a weight that is 90% of the weight assigned to the next (more recent) time period. $\hat{X}(t+1) = \alpha X(t) + (1-\alpha)\hat{X}(t)$

A higher α cannot smooth out the measurement noise whereas a lower value is slow in adapting to changes in the time series. Note that for α equal to 1, we have a simple AR(1) model, very sensitive to the level changes in the traffic. For α equal to 0, the algorithm corresponds to a simple moving average less sensitive to the level changes. The prediction algorithm uses the aforementioned models to compute the predicted traffic for the next time interval. We will assume that these prediction algorithms consider a *fixed window size* for the recent traffic history. Section III-C.3 presents the adaptive moving-average based algorithm that dynamically decides about the window size.

3) Adaptive Moving-Average based forecasting: Motivated by the need to better capture the burstiness of the traffic and adapt to its sudden changes during the forecasting process, we propose a novel forecasting approach. The **adaptive moving-average** algorithm dynamically detects the level shifts (i.e., prominent changes of traffic) in the traffic and establishes a new window size of the recent traffic. It then applies a moving-average based algorithm using this window of recent traffic.

As the traffic is received and sent at an AP during each interval (e.g., k-th interval), the traffic time series $X(k)$ is being formed. It employs a sliding window that moves across the traffic time series $X(k)$ and detects the level shifts. A recent history window has a minimum and maximum size (w_{min} and w_{max}, respectively), currently set at 3 and 12 time intevals.

Let us assume that the current window is L_m starting and ending at the beginning of the k_m-th and k_{m+1}-th interval, respectively ($k_{m+1} - k_m \leq w_{max}$ and $k_m + w_{min} \leq k_{m+1}$)). The traffic accessed during the window L_m is the set of all values in the $X(k_m), \ldots, X(k_{m+1})$. Each traffic value $X(k_m)$ corresponds to the aggregate traffic accessed from the access point during the k_m-th time interval. Depending on the time-scale the k_m-th interval is one hour or 5-minute interval. The algorithm "scans" *dynamically* the time series on the fly (while the AP operates), starting from the end of the current interval k_j ($k_j \geq k_{m+1}$) and tries to detect level shifts in the traffic (accessed during those time intervals). A level shift flag is triggered when one of the following conditions becomes true.

1) $X(k_j) > X(k_l)$, $\forall k_l \in L_m$
2) $X(k_j) < X(k_l)$, $\forall k_l \in L_m$

If none of the conditions are true, the current window L_m is expanded to include the current intervals up to k_j-th and the algorithm continues with the next interval k_{j+1}. When the current window size exceeds its maximum size, the algorithm keeps as its current window only the w_{max} most recent values. When one of the aforementioned conditions becomes true, a new interval L_j that corresponds to the subset $\{X(k_j), \ldots, X(k_{j+3})\}$ is formed. A level shift at k_j is *detected* if the confidence intervals of the traffic accessed during L_i and L_j are *non overlapping*. This is the *level shift criteria*. In the case of a new level shift, the current window becomes the L_j and the algorithm continues to the next interval k_{j+1}.

4) Hybrid algorithms based on periodicities and recent history: To incorporate both the recent traffic and periodicities, we build a hybrid model that uses the moving average and historical means. Specifically, for each AP (e.g., AP i), we introduce the **hybrid model** (P-MA), a *weighted average* of the **periodic-based** models of the historical mean hour and hour-of-day and the **simple moving average** (MA) defined as

$$(P-MA) \quad \hat{X}_i(h,d) = a \times (1/w) \sum_{k=t-w}^{t-1} X(k) + b \times \mu_i(h,d) + c \times \mu_i(h).$$

We experiment with different window sizes and weights to evaluate the impact of the recent history and periodicity on forecasting. Note that the P-MA with weights (a,b,c) equal to (1,0,0) and history window w is a simple MA model of window w and has the form of an autoregressive process of order w, $AR(w)$. In that case, the prediction takes into account only the recent traffic history instead of the periodicity.

The weights of the P-MA can be established using multiple linear regression. The weights will reveal which of the predictors, namely, the historical mean hour, historical mean hour-of-day, recent-history have the greatest effect. The linear model takes the form $y = Xb + e$, where y is a vector of observations, X is a matrix of independent variables (regressors/predictors) and e is a vector of random disturbances. P-MA with weights established via multiple linear regression is denoted as **P-MA-RG**.

5) Multi-sourced algorithms with flow-related information: The majority of wireless traffic is TCP-based. Furthermore, we found a high correlation (above 90%) in the log-log scale between the traffic load and number of flows in both the hourly and five-minute time intervals. We designed two novel algorithms that use flow-based information to enhance the predictions.

Unlike the previous algorithms that use only SNMP data, the new algorithms will integrate SNMP and flow-based information. Specifically, in this paper, we investigated the impact of number of flows and type of application on forecasting. For that, we correlated the SNMP-client based and TCP-packet headers information. Since the time granularity of the SNMP data is five minutes, we created five-minute time series of the number of active flows at each interval for each AP using the TCP-packet headers.

We established the relation between number of active flows and traffic load per interval for each AP by applying linear regression on the two time-series generated using a *training data set*. Therefore , for each AP, the number of flow based model was a function of the form: $log(X(k)) = alog(N_{flows}(k-1)) + b$, where $X(k)$ and $N_{flows}(k)$ are the total traffic and number of flows during the k-th interval of that AP, respectively. To forecast the traffic during the $k-$th interval, $\hat{X}(k)$, the **number-of-flows-based** prediction looksup the number of flows at the previous interval and predicts as traffic demand the

$$\hat{X}(k) = e^b (N_{flows}(k-1))^a. \tag{1}$$

The type-of-flow-based forecasting algorithm predicts the traffic at the next time interval based on the estimation of the traffic "contribution" of its active flows at that interval. To compute the contribution of a flow, we use the median traffic size and duration of flows of the same port number (application type). There is a one-to-one mapping between the application type and port number of the destination of a tcp flow. For example, let us assume that $\bar{l}_p$ and $\bar{d}_p$ are the median traffic size and duration of flows of port number p and d the time-scale. To forecast the traffic load of an interval k, we find the set of active flows at the end of the $(k-1)$-th interval. For each active flow f (of application type p) that started at time t_s we estimate its expected traffic contribution at the k-th interval as the ratio of the median bandwidth requirement of this type of flows by the estimated duration of that flow,

$$\frac{\bar{l}_p}{\bar{d}_p} min((t_s + \bar{d}_p - d * (k-1), d).$$

The **type-of-flow-based** forecasting algorithm will predict as the traffic load of the next time interval, the sum of the contributions of all active flows in that interval. Note that the type-of-flow-based algorithm does not consider the arrival of new flows at the next time interval.

6) Normalized ARIMA multi-step ahead time-series forecasting (NAMSA): The NAMSA forecasting approach can be summarized with the following steps: a) Transform the load in a reasonable way to make the data more normally distributed. Here the transformation ($Y(t) = (X(t))^{1/4}$) is subjectively chosen, and it seems to be working well in the current application. We intend to work out a more automatic procedure to decide on the appropriate transformation in the future for better generalization. b) Investigate time-varying patterns of the mean and variability of the transformed load. c) Normalize the transformed load if the mean and variability are indeed time-varying. We normalize the transformed load $Y(t)$ as $e(t) = \frac{Y(t)-\mu_h(t)}{\sigma_h(t)}$, where $\mu_h(t)$ is the mean of $Y(t)$ during those time periods with the hour-of-day being $h(t)$ while $\sigma_h(t)$ is the stadard deviation of $Y(t)$ during those intervals. The function $e(t)$ can be treated as a normalized version of $Y(t)$. d) Develop standard time series models like AR(p) for the normalized series, and employ rigorous model selection procedures like AIC to select the optimal model. e) Perform one-step-ahead or multi-step-ahead forecasting on the normalized series using the fitted model, and then back-transform the forecast to the original scale. We describe the process in more detail in [14].

D. Prediction metrics: absolute and relative prediction errors

To evaluate the performance of the prediction algorithms, we compute the *absolute prediction error* defined as the absolute difference of the predicted from the actual traffic and the *relative prediction error* which is the ratio of this absolute error over the actual traffic. The relative prediction error $r(t)$ at an AP i durinth the t-th time interval is defined as $r(t) = |\hat{X}_i(t) - X_i(t)|/X_i(t)$, where $X_i(t)$ and $\hat{X}_i(t)$ are the actual and predicted traffic of AP i during the t-th time interval. A perfect prediction algorithm has absolute and relative prediction error equal to 0. A weakness of the relative prediction error is that in intervals with zero actual traffic, the relative prediction error is not estimated. This results in conservative evaluations. Furthermore, large relative prediction errors indicate large over or under-estimations of traffic. To better appreciate the performance of the forecasting algorithms, both the relative and absolute prediction errors need to be considered. A good prediction algorithm should have low absolute and relative prediction errors. The **mean (median) relative prediction error** of all hotspots is the average of the mean (median) relative prediction error considering all hotspots. Similarly, we compute the mean and median absolute prediction error. For the evaluation of the forecasting algorithms, we use the aforementioned metrics.

IV. Evaluation of the forecasting algorithms and discussion

Algorithm	**Type**	**Trace**	**Tracing period**	**Forecasting period**	**Scale**
P-H HoD	Periodic Periodic	SNMP-aggregate	3 weeks excl. weekends start: Mon 18/10/04	5-day start: Mon 11/8/04	hour
P-MA,P-MA-RG	Hybrid	SNMP-aggregate	(same as above)	(same as above)	hour
NAMSA	Hybrid	SNMP-aggregate	(same as above)	(same as above)	5 min, 1 hour
MA fixed window size=2,3,4	Recent history	SNMP-aggregate	no training	(same as above)	hour
MA,EMA fixed window size	Recent-history	SNMP-aggregate	no training	1 day start: Wed 9/29/05	5 min
Adaptive EMA (α=0.4)	Recent history	SNMP-aggregate	no training	1 day, Wed 9/29/05	5 min
Type-of-flow	multi-traced	SNMP-client, tcp headers	no training	2 days, start: 4/14/05	5 min
Number-of-flows	multi-traced	SNMP-client, tcp headers	one day, Wed 4/13/05	2 days, start: Thu 4/14/05	5 min

TABLE I

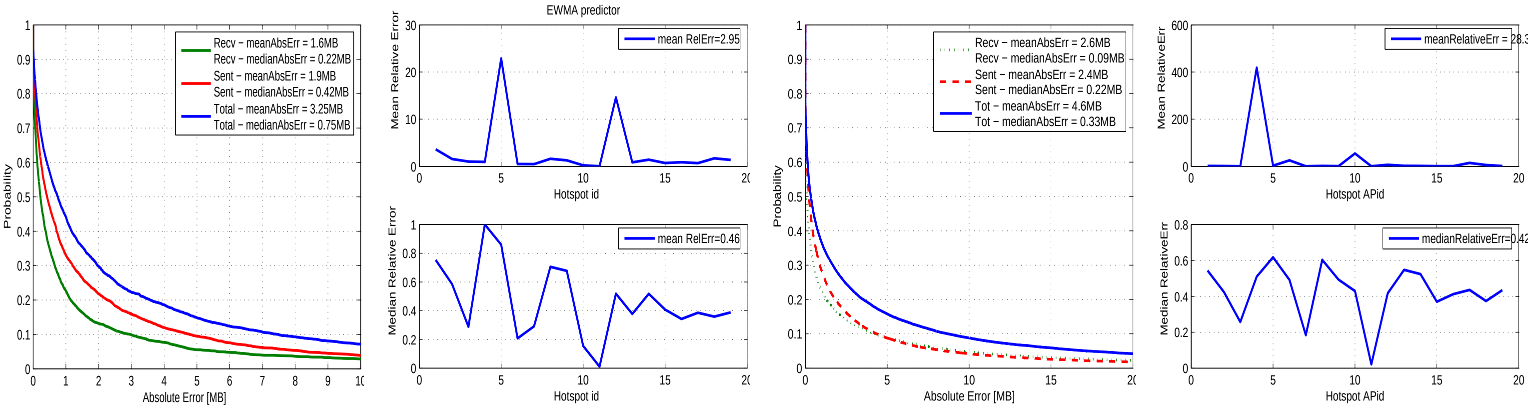
Summary of the forecasting algorithms with the type of traces used, their tracing and forecasting period, and time scale

A. *Evaluation of the simple periodic-based and recent-history based algorithms*

In [14], we evaluated the P-MA, P-H, P-HoD and NAMSA algorithms for the hourly time series considering all hotspot APs. We summarize the main results in the following paragraphs. The means in these algorithms were computed based on the history for each AP. The history corresponds to the three weeks of the trace, excluding weekends and starting on Monday, October 18th, 2004. We predicted the traffic for each AP, for all the hours during the weekdays of the following week (Monday, November 8th until Friday, November 12th). We called this period *forecasting period*. For P-MA, we varied the window size to be 2, 3, 4, and 5 hours. We evaluated P-MA for various weights (values of a,b, and c), including also values resulted from applying multiple linear regression for each AP. We observed that the MA with window size of two hours performs better than the PH and PHoD with respect to their mean and median relative prediction errors of all hotspot APs. Specifically, the mean (median) relative prediction errors of P-H, PHoD, MA, and NAMSA are 617.89 (1.23), 170.52(0.80), 185.34 (0.66), and 93.33 (0.75), respectively. These high relative prediction error indicate the high burstiness of the traffic. MA's mean and median absolute error of 17.38MB and 0.66MB, respectively.

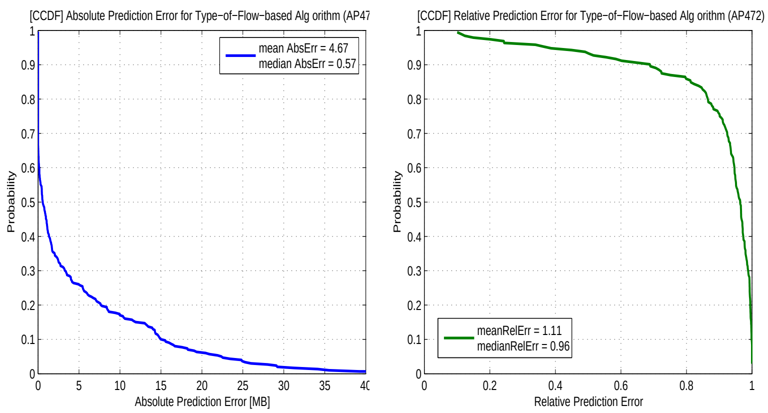
B. *Dramatic Improvement in forecasting performance using finer time scales*

The performance of the MA and EMA when forecasting in finer time scales (in 5minute traces) has been improved dramatically. To evaluate the algorithms, we considered one-day from the original snmp-aggregate data in 5-minute intervals and run variations of the MA and EMA algorithms. We were particularly interested in testing the adaptivity of the adaptive-EMA algorithm, and for this, we select an $\alpha = 0.4$ and 95%-confidence intervals, for the level-shift detection criteria. The adaptive EMA reports a relative prediction error of 2.95 and 0.46 , for its mean and median considering all hotspots and 5-minute interval predictions, repsectively. The mean and median absolute errors are 3.33MB and 0.75MB. This algorithm performs much worse on hourly traffic time-series. The algorithm first forecasts the received and sent traffic at each five-minute interval. Based on these forecasts, it reports as the total traffic for that five-minute interval, the sum of the predicted receive and sent. Notice that its performance is one or two orders of magnitude better than PH's, PHoD's, and P-MA's performance on hourly based time-series. Although several APs have strong diurnal patterns, the performance analysis reveals that the more “recent” the historical data and the finer their time scale, the stronger its impact on forecasting.



(a) (b) (c) (d)

Fig. 1. Adaptive EMA on Figs.(a,b) and NAMSA on Figs. (c,d) with five-minute traces.



(a) (b)

Fig. 2. The performance of the type-of-flow based algorithm on hotspot AP 472.

Figure 1(a) and (b) illustrate the performance of adaptive EMA when it first predicts the receive and sent traffic (from the corresponding recent history measurements) and then forecasts as total traffic the sum of the predicted receive and sent. We found that the improvement in forecasting the total traffic "indirectly" (i.e., through the forecasts of the received and sent traffic) over forecasting directly the five-minute time-series of total traffic is negligible. On the other hand, the adaptive mechanism in MA or EMA does not show any improvement when it is applied on hourly time-series. For example, the adaptive MA has a mean relative prediction error 3213.90 and a median relative prediction error 0.78 which is larger than the prediction errors of simple MA-based models [14].

As Figures 1 (c) and (d) illustrate, the NAMSA on 5-minute traces has slightly better medians but worse means in both the absolute and relative prediction errors. However, the improvement in the performance of NAMSA on 5-minute traces over the one on hourly time intervals is distinct. In both the cases NAMSA was applied on traces that correspond to the same period (mentioned in Section IV-A).

Compared to the aforementioned algorithms that require some statistical analysis of the traffic, the flow-based algorithms presented in Section III-C.5 are simpler and require less processing overhead. We used the SNMP client-based and TCP-flow based data sets with one-day training period to create the 5-minute interval times series of traffic load and number of flows at each AP and then fit the parameters of Eq.1. We evaluated the number-of-flows algorithm on the 5-minute-interval trace of the next two days. The (median, mean) pairs of the relative and absolute prediction errors are (0.89,19.68) and (0.23MB, 4.24MB), respectively.

The type-of-flow based algorithm applied on a 5-minute traces for the hotspot AP 472 has a median and mean absolute prediction error of 4.67MB (0.02MB) and 0.57MB (0.48MB), respectively. The median and mean relative prediction error of

0.96 (1) and 1.11 (30.33). In parenthesis, we indicate the performance of the number-of-flow based algorithm in 5-minute intervals for the same tracing period and AP. Figures 2 illustrate the CCDF of the relative and absolute errors when the type-of-flow algorithm was used on hotspot AP 472.

V. Related work

There are several measurement studies that have examined the workload of 802.11 APs in production environments. In general, most of them have considered a wider range of issues, such as overall usage of a wireless infrastructure, and client mobility patterns, providing only a limited picture of the utilization of APs. Balazinska and Castro [4] used SNMP to characterize a much larger wireless network in three IBM buildings (177 APs). Their study examined the maximum number of simultaneous users per AP (mostly between 5 and 15), total load and throughput distributions. Two interesting observations found in this paper are that offered load and number of users are weakly correlated, and that user transfer rates are dependent on the location of the AP. Balachandran *et al.* [3] performed measurements in a three-day conference setting, also focusing on the offered network load and global AP utilization. They characterized wireless users and their workload and addressed the network capacity planning problem. The overall bursty behaviour and peaks and troughs are similar at all APs, though the absolute peak throughout at each AP varies. They observed that offered load is more sensitive to individual client traffic characteristics rather than just the total number of clients. Kotz *et al.* [10], [7] studied the wireless network at Dartmouth College using syslog, SNMP, and tcpdump traces and reported the distribution of average and maximum daily traffic at APs. A subset of the same data (syslog messages and tcpdump traces from 31 APs in 5 buildings) was revisited by Meng *et al.* [13] for flow modeling purposes. The authors proposed a two-tier (Weibull regression) model for the arrival of flows at APs and a Weibull model for flow residing times, and they also observed high spatial similarity within the same building. Papagiannaki *et al.* [16] modeled the evolution of aggregated IP backbone traffic at large time scales, and developed long-term (up to 6-month ahead) forecasting models that can be used for capacity planning purposes. In particular, they analyzed eight inter-PoP aggregate demand time series from October 2000 to July 2002 with a granularity level of 90 minutes. Recently, Dovrolis *et al.* [6] developed a number of formula-based and history-based algorithms for predicting the throughput of large TCP transfers in wired networks. The formula-based algorithms rely on mathematical models that express the TCP throughput as a function of the characteristics of the underlying network path. We were inspired by their level-shift/moving-average algorithm and decided to apply it on our data.

VI. Conclusions and future work

We designed a number of forecasting algorithms and evaluated their performance on the hotspots of a larger production wireless network. Short-term forecasting on wireless networks is challenging but the use of finer time-scale can improve dramatically the mean prediction errors. Recently, we found some interesting phenomena on the traffic characteristics of APs (e.g., uploading vs. downloading dichotomy, large diversity in flow sizes, clusters of APs based on their building types, time-varying poisson time arrivals of clients) [14], [15], [8], [9]. We intend to study more systematically the temporal and spatial correlations of APs, classify APs based on various parameters (e.g., traffic characteristics, building type, number of associations, and distinct clients), and explore the impact of the above characteristics on forecasting.

The trace collection is still ongoing. We plan to investigate forecasting in various time-scales. Shorter-term forecasting (e.g., next minute) can assist in designing more energy-efficient clients. Long-term forecasting is essential for capacity planning and understanding the evolution of the wireless traffic and networks. For that, we will study the performance of MRA on a longer trace once it becomes available, and compare that with [16].

This research is a part of a comparative analysis study on wireless access patterns in various environments, such as a medical center, research institute, campus, and a public wireless network. We intend to analyze traces from testbeds in these environments and contrast their traffic models. Understanding and forecasting the traffic of APs can have a dominant impact on the operation of wireless APs and clients and this study sets a direction for exploring further these issues.

References

[1] America's most connected campuses. http://forbes.com/home/lists/2004/10/20/04conncampland.html.

[2] F. Anjum, M. Elaoud, D. Famolari, A. Ghosh, R. Vaidyanathan, A. Dutta, P. Agrawa, T. Kodama, and Y. Katsube. Voice performance in WLan networksàn experimental study. In *Proceedings of the IEEE Conference on Global Communications (GLOBECOM)*, Rio De Janeiro, Brazil, December 2003.

[3] Anand Balachandran, Geoffrey Voelker, Paramvir Bahl, and Venkat Rangan. Characterizing user behavior and network performance in a public wireless lan. In *Proceedings of the ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, 2002.

[4] Magdalena Balazinska and Paul Castro. Characterizing mobility and network usage in a corporate wireless local-area network. In *First International Conference on Mobile Systems, Applications, and Services (iMobiSys)*, May 2003.

[5] Mark Crovella and Azer Bestavros. Self-similarity in world wide web traffic: Evidence and possible causes. In *Proceedings of SIGMETRICS '96*, 1996.

[6] C. Dovrolis, Q. He, and M. Ammar. On the predictability of large transfer tcp throughput. In *SIGCOMM Symposium on Communications Architectures and Protocols*, Philadelpha PA, 2005.

[7] Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campuswide wireless network. In *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, Philadelphia, September 2004.

[8] Felix Hernandez-Campos and Maria Papadopouli. Assessing the real impact of 802.11 wlans: A large-scale comparison of wired and wireless traffic. In *14th IEEE Workshop on Local and Metropolitan Area Networks*, Chania, Crete, Greece, 2005.

[9] Felix Hernandez-Campos and Maria Papadopouli. A comparative measurement study of the workload of wireless access points in campus networks. In *16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, Berlin, Germany, 2005.

[10] David Kotz and Kobby Essien. Analysis of a campus-wide wireless network. Technical Report TR2002-432, Dept. of Computer Science, Dartmouth College, September 2002.

[11] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic. *ACM Computer Communication Review*, 25(1):202–213, 1995.

[12] W. E. Leland, W. Willinger, M. S. Taqqu, and D. V. Wilson. Statistical analysis and stochastic modeling of self-similar datatraffic. In *Proc. 14th Int. Teletraffic Cong., 6-10*, volume 1, pages 319–328, Antibes Juan Les Pins, France, June 1994.

[13] Xiaoqiao Meng, Starsky Wong, Yuan Yuan, and Songwu Lu. Characterizing flows in large wireless data networks. In *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pages 174–186, Philadelpia, 2004.

[14] Maria Papadopouli, Haipeng Shen, Elias Raftopoulos, Manolis Ploumidis, and Felix Hernandez-Campos. Short-term traffic forecasting in a campus-wide wireless network. In *16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, Berlin, Germany, 2005.

[15] Maria Papadopouli, Haipeng Shen, and Manolis Spanakis. Modeling client arrivals at access points in wireless campus-wide networks. In *14th IEEE Workshop on Local and Metropolitan Area Networks*, Chania, Crete, Greece, 2005.

[16] K. Papagiannaki, N. Taft, Z. Zhang, and C. Diot. Long-term forecasting of internet backbone traffic: Observations and initial models. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, San Francisco,CA, April 2003.

[17] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high-variability: Statistical analysis of ethernet lan traffic at the source level. *ACM Computer Communication Review*, 25(4):100–113, October 1995.