Multi-Step-Ahead Prediction of IP Packet Delay Variation Based on a GARCH Model

Takashi Oshiba and Kazuaki Nakajima

Cloud System Research Laboratories, NEC Corporation, Japan

{oshiba@cp, nakajima@ah}.jp.nec.com

***Abstract*—We propose JitSeer, a long horizon prediction method for the delay variation of IP packets; i.e., jitter. In designing JitSeer, we directed our attention to similarities between delay variation, which is the variation of one-way delay, and volatility in finance, which is the variation of financial asset return. From a time series view, the similarities are that both vary over time, and during some periods they swing quickly while in other periods they stay calm. Based on these similarities, the delay variation is modelled with a GARCH model, which is a time series model originally developed in econometrics to capture the dynamics of volatility. Then the model predicts the future multi-step-ahead delay variation. Experimental results obtained from operational mobile networks show that the prediction accuracy of JitSeer is better than that of conventional methods; its prediction accuracy for the next 10 seconds is 74% over 3G and 88% over LTE.**

I. Introduction

Mobile networks such as 3G and LTE cellular networks have become a dominant form of Internet access due to the recent rise of smartphones [1]. In mobile networks, the delay variation of IP packets [2], which is called jitter [3], fluctuates drastically [4] since the wireless access network experiences rapidly changing link rates due to radio interference and occasional instantaneous outages. Most current network applications which cope with the delay variation attempt to ensure quality of service based on feedback control [3][5] using past measurements of delay variation. Unfortunately, the feedback control is applied only after the fluctuation of delay variation, when the user experience has already been degraded. Thus, ensuring quality of service remains a challenging problem.

If long horizon (e.g., the next 10 seconds) prediction of future delay variation can be realized, network applications will have a long preparation time (i.e., up to the next 10 seconds) before the anticipated fluctuation of delay variation, and can take action based on the prediction results. Thus, long horizon prediction will help ensure quality of service.

In this paper, we propose JitSeer, a multi-step-ahead (i.e., long horizon) delay variation prediction method. In designing JitSeer, we directed our attention to similarities between delay variation, which is the variation of one-way delay, and *volatility* [6] in finance, which is the variation of financial asset return (i.e., relative price changes). In JitSeer, the delay variation is modelled with a *generalized autoregressive conditional heteroskedasticity (GARCH) model* [7], which is a time series model originally developed in *econometrics* [8] to capture the dynamics of volatility. Then the model predicts the future multi-step-ahead delay variation.

MMSP'13, Sept. 30-Oct. 2, 2013, Pula (Sardinia), Italy.
978-1-4799-0125-8/13/$31.00 ©2013 IEEE.

Although there are a few conventional methods for delay variation prediction [9][10], these methods have been evaluated only through simulations and their prediction accuracy over real networks has not yet been validated.

The main contributions of this paper are:

- We shed light on *the similarities between delay variation and volatility*, and propose a multi-step-ahead prediction method for delay variation based on a *GARCH model*. To the best of our knowledge, our method is the first attempt to apply a GARCH model to predict delay variation.
- We extend a standard GARCH model by introducing a *correction term* in order to improve its prediction accuracy.

II. Applications Using Prediction of Delay Variation

In this section, we enumerate the wide variety of potential applications using predicted results of delay variation:

(1) *Jitter buffer control in VoIP*: Feedback-based jitter buffer control [11] can be improved through delay variation prediction. In current feedback-based VoIP applications over mobile networks, the jitter buffer length fluctuates rapidly and widely due to a large delay variation, and this degrades speech quality. If long horizon (e.g., the next 10 seconds) prediction can reliably indicate an increase in delay variation, the jitter buffer length can be extended gradually. Thus, jitter buffer starvation can be avoided. Note that the jitter buffer extension will add to latency. However, since the next 10-second interval will generally contain plenty of speech and silent segments, a waveform reconstruction mechanism [12] at a receiver which cooperates with the jitter buffer control should enable a more natural speech pitch. Thus, the latency impact can be mitigated.

(2) *Routing algorithms based on delay variation*: In current delay-variation-based routing algorithms [13], a route is selected based on delay variation feedback. A route with less delay variation can instead be proactively selected using the predicted delay variation.

(3) *Omen detection of packet loss*: Increased delay variation can be an omen of packet loss [4][14]. Thus, the loss-based congestion avoidance algorithm of TCP [15] as well as the adaptive rate control of video and/or audio based on loss events [5] can use the predicted delay variation to take anticipatory action.

(4) *Forward error correction (FEC)*: In existing adaptive FEC [16], the degree of redundancy is decided based on feedback

regarding the packet loss rate. Thus, omen detection of packet loss through the predicted delay variation can improve the degree-of-redundancy decision algorithm.

III. Related Work

A. Prediction of Network-Related Metrics

Although much prior work has been done on predicting network-related metrics, the work has mainly focused on throughput [17]–[19] and end-to-end delay [19]–[24], and the prediction of delay variation has not received enough attention. Note that, in the TCP's retransmission timeout calculation [25], the *future* round-trip time (RTT) is not *predicted,* and the *current* smoothed RTT is merely *estimated*. Also note that RTT cannot be converted to one-way delay variation due to the asymmetric nature of the one-way delay between uplinks and downlinks in cellular networks [26].

B. Delay Variation Prediction

There are a few existing methods for delay variation prediction. In [9], an autoregressive (AR) model is used to predict delay variation. Since [9] is the most closely related method to JitSeer, we compare these two methods in Section VI. In [10], a neural network method is used to predict delay variation. Since this method requires a huge amount of learning datasets and a huge number of iteration times to train the neural network, the method is extremely difficult to run in our environment. Thus, we cannot compare the method of [10] with JitSeer. Note that these existing methods have been evaluated only through simulations [9][10], so their prediction accuracy over real networks has not yet been validated.

C. Prediction with a GARCH Model

In recent years, several GARCH-based prediction methods [27]–[30] have been proposed in the area of IP networks. However, these methods do not predict delay variation.

IV. Delay Variation and Volatility

In this section, we explore the relation between delay variation and volatility.

A. Definition of Delay Variation

In [3], the definition of delay variation is standardized, and it is compatible with [2]. In Fig. 1, a sender transmits IP packets to a receiver. Let S_t denote the sending time for a packet t and R_t denote the receiving time for packet t. For a pair of packets $t-1$ and t, the difference D_t in packet spacing at the receiver compared to that at the sender for the pair of packets is then defined as

$$D_t = (R_t - S_t) - (R_{t-1} - S_{t-1}) = (R_t - R_{t-1}) - (S_t - S_{t-1}) . \quad (1)$$

Clock synchronization is not required to obtain D_t. If the value of D_t is greater than zero, the transmission duration of packet t is longer than packet $t-1$. The definition of delay variation J_t, also called interarrival jitter, is standardized as the smoothed absolute value of a time series of D_t [3]. More specifically, J_t is computed with an exponentially weighted moving average (EWMA) where the smoothing factor is 1/16:

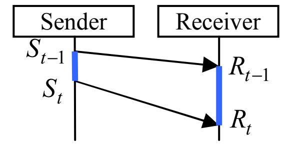

Fig. 1. Difference of packet spacing between the receiver and the sender

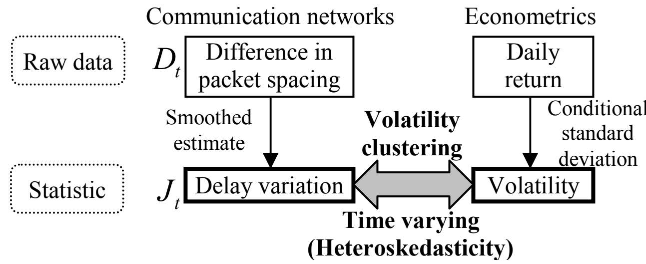

Fig. 2. Similarities between delay variation and volatility

$$J_t = \frac{1}{16} | D_{t-1} | + \frac{15}{16} J_{t-1} . \quad (2)$$

B. Similarities Between Delay Variation and Volatility

We find similarities between the packet spacing difference D_t and financial asset returns. Fig. 2 summarizes the analogy. A time series of D_t for one of our IP packet traces measured over an operational 3G network is shown in Fig. 3a. In the packet trace, UDP/IP packets were transmitted every 20 ms ($S_t - S_{t-1}$ = 20 ms), so the minimum value of D_t is limited to −20 ms (when two packets arrived back to back; i.e., $R_t - R_{t-1}$ = 0 ms). Although the time series of D_t normally moves within the range of ± 20 ms, jumps over 100 ms occur frequently. Fig. 3b shows a time series of daily returns of the S&P 500 index (a representative U.S. stock market index) over the period 1992–2006, as an example of asset returns. In Fig. 3b, a value of 0.03 corresponds to a one-day increase in the index value of 3%.

The first similarity we found is that in both time series there are periods where values swing quickly and other periods of relative calm. In Fig. 3b, 1997–2002 is a swing period, and 1992–1996 and 2003–2006 are calm periods. In econometrics, this phenomenon is called *volatility clustering* [6]. Interestingly, in Fig. 3a, the packet index range 400–1500 (roughly) is a swing period, and the 1–400 and 1500–2000 ranges are calm periods. Thus, we can visually observe volatility clustering in the time series of difference D_t.

The second similarity is the time varying standard deviation. Fig. 3b shows volatility as well as daily returns (the computation procedure for volatility is given in Section V-A). In econometrics, volatility is defined as the *conditional standard deviation* of financial asset returns [6]. This definition is supported by the plot in the figure, which shows that the volatility was strongly correlated with the variation of asset returns. Fig. 3b also shows that the volatility (i.e., conditional standard deviation) varies over time, a property called *heteroskedasticity* (i.e., variance heterogeneity) in statistics [6]. Since the delay variation J_t in Fig. 3a is the smoothed estimate of *difference* of packet spacing D_t, we consider J_t to be a standard-deviation-related statistic. Interestingly, since the delay variation J_t varies over time in

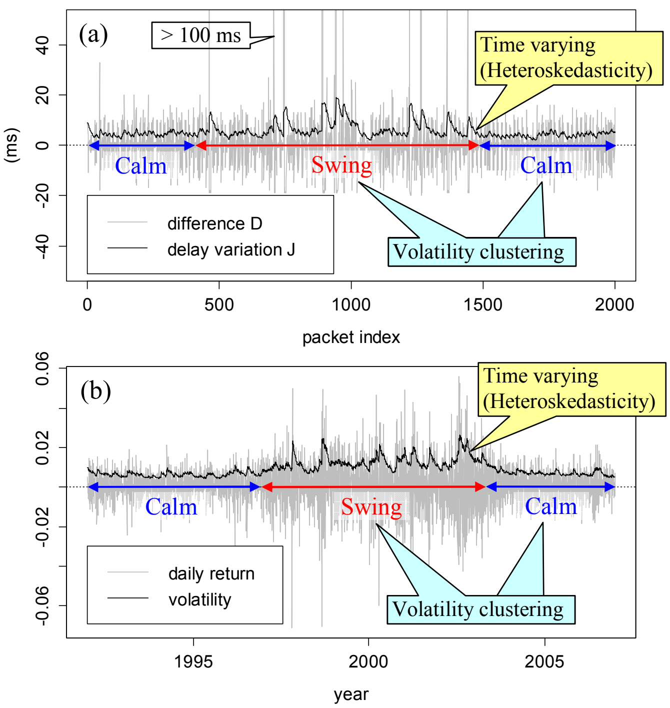

Fig. 3. Time series of (a) difference of packet spacing (top) and (b) daily returns (bottom). Both exhibit *volatility clustering* and *heteroskedasticity*.

Fig. 3a, we also observe heteroskedasticity in the delay variation J_t.

V. Proposal of JitSeer

In this section, we show how delay variation is modelled and predicted with JitSeer.

A. JitSeer Based on a GARCH Model

Observing the similarities of volatility clustering and heteroskedasticity inspired us to design JitSeer, a multi-step-ahead delay variation prediction method. The key idea of JitSeer is to model delay variation J_t as the volatility (i.e., conditional standard deviation) of the difference in packet spacing D_t using a generalized autoregressive conditional heteroskedasticity (GARCH) model. We believe the GARCH model is suitable for modeling delay variation J_t because the GARCH model was originally developed to capture the dynamics of volatility which is characterized by volatility clustering and heteroskedasticity. The input to JitSeer is the time series of difference D_t and the output is a multi-step-ahead prediction of delay variation J_t.

In econometrics, a GARCH model whose order is p and q, referred to as a GARCH(p,q) model [7], is given by

$$D_t = E_{t-1}[D_t] + \varepsilon_t \,, \tag{3}$$

$$\varepsilon_t = \sigma_t z_t, \quad \sigma_t > 0, \quad z_t \sim \text{i.i.d.}, \quad E[z_t]=0, \quad Var[z_t]=1 \,, \tag{4}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \,. \tag{5}$$

In Eq. (3), D_t is daily return, $E_{t-1}[\cdot]$ represents conditional expectations at time $t-1$, and ε_t is a random variable. Thus, D_t is divided into $E_{t-1}[D_t]$ which can be calculated at time

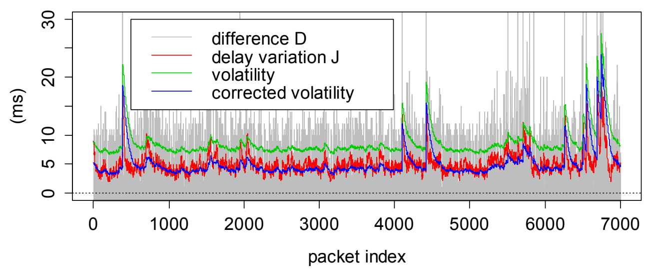

Fig. 4. Example of the differences in packet spacing, delay variation, volatility and corrected volatility

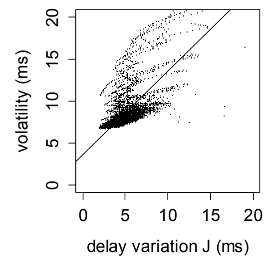

Fig. 5. Scatter plot of delay variation and volatility. While the thick chunk looks flatter than the regression line, the computed slope is close to 1.

$t-1$ and ε_t which is unpredictable at time $t-1$. In Eq. (4), σ_t is volatility and z_t is a strong white noise process. In Eq. (5), the squared volatility σ_t^2 at time t is comprised of the squared random variable ε_{t-i}^2 all realized by time $t-1$, the squared volatility σ_{t-j}^2 all realized by time $t-1$, and the parameters of the GARCH model (ω, α_i and β_j). To capture the dynamics of volatility clustering and heteroskedasticity correctly, the parameters are estimated with quasi-maximum likelihood estimation (QMLE). We do not have space here to describe the QMLE procedure, but it is described elsewhere [31]. After this procedure, the squared volatility σ_t^2 can be computed with Eq. (5).

In JitSeer, D_t in Eq. (3) is not the daily return but the difference in packet spacing as defined in Eq. (1). We use a GARCH(1,1) model in JitSeer since [32] presents evidence that it is difficult to find a volatility model, including variants of GARCH and higher order models, that outperforms the simple GARCH(1,1). Thus, Eq. (5) is simplified to

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \,. \tag{6}$$

Our choice of GARCH(1,1) drastically reduces the computational load since the order selection procedure, which is a complex task, becomes unnecessary. Moreover, the QMLE for parameter estimation is simplified since only ω, α and β need to be estimated. This is important since a heavy load should be avoided in multimedia applications.

B. Extension of the GARCH Model with a Correction Term

Fig. 4 shows an example of volatility σ_t, computed through QMLE and Eq. (6), along with one of our packet traces (a different one from that in Fig. 3a). Delay variation J_t computed with Eq. (2) is also depicted. As expected, σ_t and J_t are correlated. This can be attributed to both EWMA and the GARCH model working as a kind of low-pass filter. However, there is a gap between σ_t and J_t, and the width of this gap seems to be constant. To verify this, a scatter plot of

σ_t and J_t is shown in Fig. 5, along with a line showing the linear regression. In this example, the slope of the line is 1.034 and the intercept is 3.236 ms. Since the slope is close to 1, it is reasonable to assume the distances between σ_t and J_t are almost constant throughout the time series.

Since the future delay variation is predicted based on past volatility σ_t in JitSeer, we consider it advantageous to bring σ_t close to J_t. To this end, we introduce a *correction term* $\hat{C}$ to the standard GARCH(1,1) model:

$$\hat{C} = \underset{C}{\operatorname{argmin}} \sum_t (\sigma_t - J_t - C)^2 \,. \tag{7}$$

The correction term $\hat{C}$ corresponds to the intercept of the linear regression line and can be computed by the ordinary least squares (OLS) method. In the above case, $\hat{C}$ is 3.236 ms.

With Eq. (6) and (7), volatility σ_t is corrected to

$$\sigma'_t = \sqrt{\omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2} - \hat{C} \,. \tag{8}$$

The corrected volatility σ'_t is shown in Fig. 4 and successfully overlaps the delay variation J_t.

C. Multi-Step-Ahead Prediction of Delay Variation

Assume the model parameters are estimated over the time period $t = 1,2,\cdots,T$ and the current time period is T. According to [31], in a standard GARCH(1,1) model, the future h-step-ahead ($h = 1,2,\cdots$) delay variation prediction σ_{T+h} can be computed as

$$\sigma_{T+h} = \sqrt{\omega \sum_{i=0}^{h-1} (\alpha+\beta)^i + (\alpha+\beta)^{h-1}(\alpha\varepsilon_T^2 + \beta\sigma_T^2)} \,, \tag{9}$$

where ε_T^2 and σ_T^2 are the fitted value computed from the GARCH(1,1) model. In JitSeer, the corrected prediction σ'_{T+h} is

$$\hat{J}_{T+h} = \sigma'_{T+h} = \sqrt{\omega \sum_{i=0}^{h-1} (\alpha+\beta)^i + (\alpha+\beta)^{h-1}(\alpha\varepsilon_T^2 + \beta\sigma_T^2)} - \hat{C} \,. \tag{10}$$

Fig. 6 shows σ_{T+h}, σ'_{T+h} and the actual future delay variation J_{T+h}, i.e., the correct answer to $\hat{J}_{T+h} = \sigma'_{T+h}$ ($h = 1,2,\cdots$). The horizontal line in the figure represents the long run level to which the corrected prediction σ'_{T+h} will eventually settle down. The long run level is the *unconditional standard deviation* $\bar{\sigma}$ [31] minus $\hat{C}$:

$$\bar{\sigma} - \hat{C} = \sqrt{\omega/(1-\alpha-\beta)} - \hat{C} \,. \tag{11}$$

Thanks to the GARCH feature that captures the volatility and the correction term $\hat{C}$, the long horizon prediction of JitSeer successfully hits the actual future delay variation.

VI. Experimental Evaluation

We evaluated JitSeer and conventional predictors with packet traces measured over operational mobile networks.

A. Conventional Predictors for Comparison

The following predictors are compared with JitSeer. Fig. 6 shows the prediction of the following predictors.

1) Naïve predictor: The prediction n_{T+h} ($h = 1,2,\cdots$) for any period equals the last observed delay variation J_T:

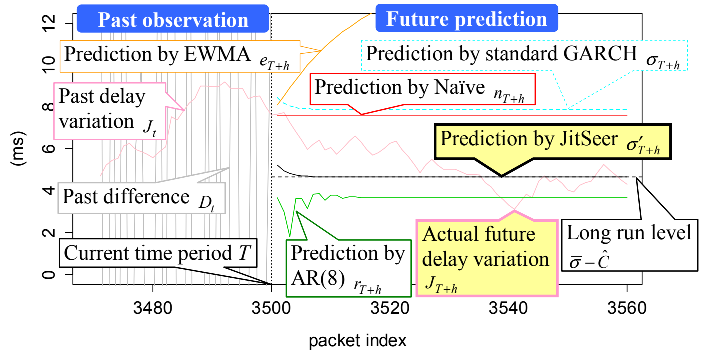

Fig. 6. Example of multi-step-ahead predictions

$$\hat{J}_{T+h} = n_{T+h} = J_T \,. \tag{12}$$

2) *EWMA predictor*: Another straightforward method is to predict the future delay variation based on the last observation of difference D_T using EWMA directly [33]. A multi-step-ahead prediction is simply computed with

$$\hat{J}_{T+1} = e_{T+1} = \frac{1}{16}|D_T| + \frac{15}{16}e_T \tag{13}$$

in a recursive manner. However, the predicted results are far from the actual future delay variation, as shown in Fig. 6, since the prediction e_{T+h} converges to $|D_T|$ which is often much higher than J_T. Thus, we exclude this method from comparison with JitSeer. (Note that if $|D_T|$ and e_T are replaced with J_T in Eq. 13, the method becomes equivalent to the naïve predictor.)

3) AR(8) predictor: In [9], an 8th-order AR model is used to predict the future delay variation as

$$\hat{J}_{T+1} = r_{T+1} = a_0 + a_1 r_T + \cdots + a_8 r_{T-7} \tag{14}$$

where $a_0, a_1, \cdots a_8$ are the parameters of the model. The AR(8) model is fitted to the observed delay variations J_t. A multi-step-ahead prediction is computed with Eq. (14) recursively.

B. Experimental Setup

In our experiments we used five UDP packet traces. Traces 1–2 were measured by us using Android smartphones (Qualcomm 1.5 GHz dual-core CPU, 1 GB RAM) over 3G and LTE networks of Japan's primary mobile operator. The durations of these traces were 440 s and 176 s, respectively. UDP packets (packet size: 44 kB) were transmitted every 20 ms to emulate adaptive multi-rate (AMR) VoIP traffic as is suitable for mobile networks [34]. We also used traces 3–5 obtained from [35] where packets were transmitted over a 3G network of Italy's primary mobile operator [36]. The durations were 117 s, 111 s and 112 s, respectively. UDP packets (packet size: 64 kB) were transmitted every 10 ms. Fig. 7 shows the difference D_t and the delay variation J_t of each trace.

Each method ran on a Windows PC (Intel Core 2 Duo 1.4 GHz CPU, 3 GB RAM), read a trace and predicted offline. In JitSeer and AR(8), the past 3,000 packets (60 s in traces 1–2 and 30 s in traces 3–5) were used for parameter estimation (including $\hat{C}$), and parameter re-estimation was done every 50 packets (1 s in traces 1–2 and 0.5 s in traces 3–5) to catch up

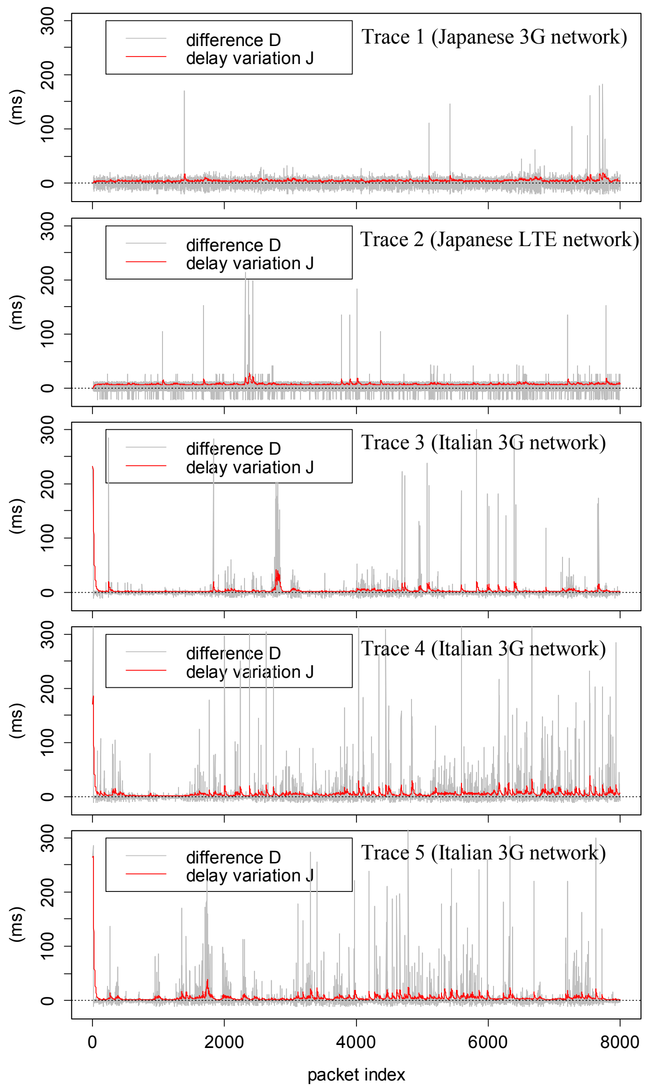

Fig. 7. Difference of packet spacing and the delay variation of traces 1–5. Only packet index 1–8,000 (160 s in traces 1–2 and 80 s in traces 3–5) is shown.

to changes in the network conditions. Each method computed 500-step-ahead ($h = 500$) predictions (10 s in traces 1–2 and 5 s in traces 3–5) for every packet. Thus, the rolling prediction was executed from packet index 3,001 to the end of the trace. After the rolling prediction throughout the trace, we computed the prediction accuracy in percentage A_T based on mean absolute percentage error (MAPE), which is the most commonly used metric for percentage error [37]:

$$A_T = 100 \times (1 - \text{MAPE}) = 100 \times \left(1 - \frac{1}{h} \sum_{t=T+1}^{T+h} \left| \frac{\hat{J}_t - J_t}{J_t} \right| \right) \qquad (15)$$

where $\hat{J}_t$ is the prediction of each method ($t = T+1, \cdots, T+h$).

C. Experimental Results

The mean of A_T from the 50-step-ahead to the 500-step-ahead predictions throughout each trace are shown in Figs. 8 and 9. Overall, our JitSeer outperformed the other methods. The prediction accuracy of JitSeer ranged from 76% to 90% for 50-step-ahead predictions and from 66% to 88% for

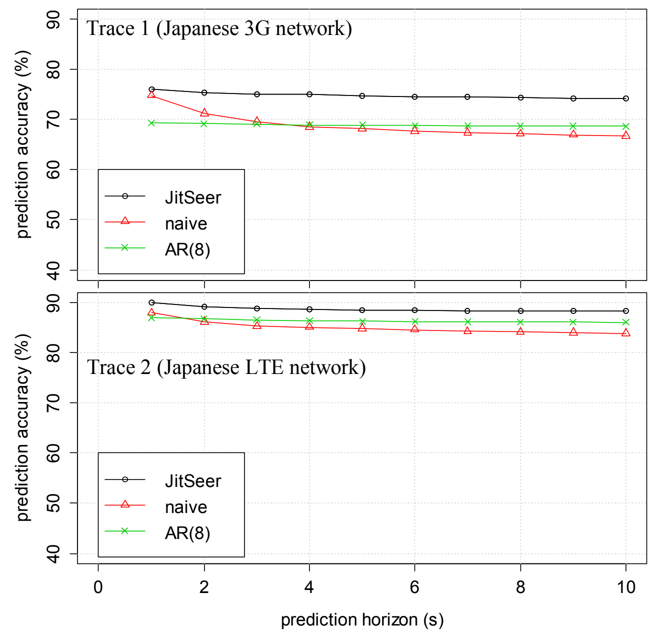

Fig. 8. Prediction accuracy for traces 1–2 (Japanese mobile networks)

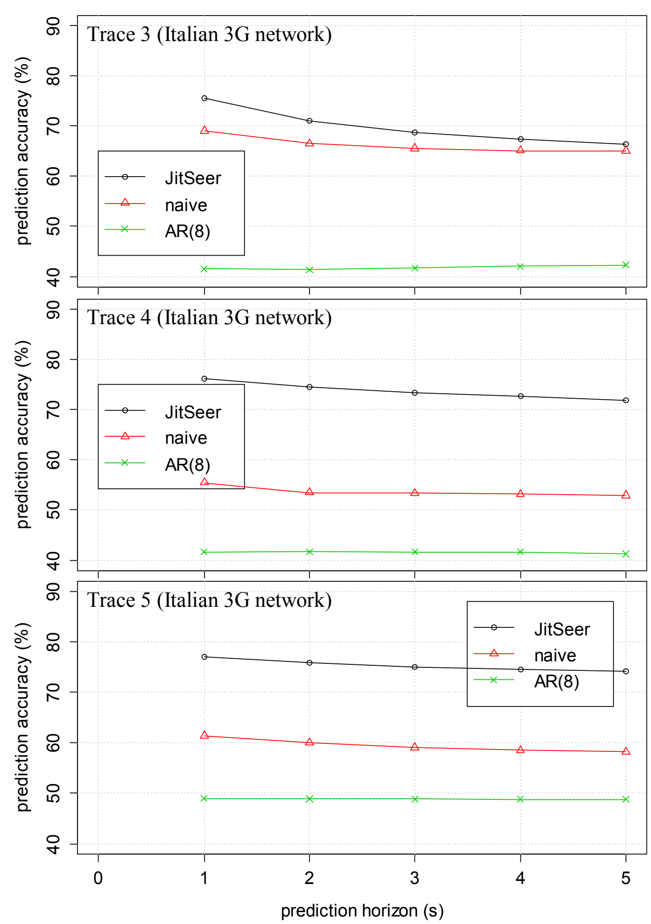

Fig. 9. Prediction accuracy for traces 3–5 (Italian 3G networks)

500-step-ahead predictions. While the prediction accuracy of JitSeer declined for longer prediction horizons, the rate of decline was gradual. In traces 1–2, the prediction accuracy for the next 10 s was 74% over 3G and 88% over LTE. The next

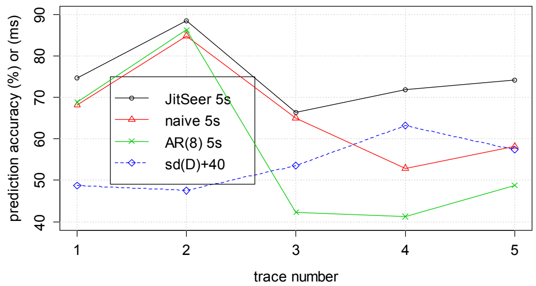

Fig. 10. Standard deviation vs. prediction accuracy in each trace. We raised the level of each standard deviation (+40 ms) to put it into the 40–90 range.

best performing method was the naïve predictor, whose prediction accuracy ranged from 53% to 84% for 500-step-ahead prediction. The AR(8) predictor was less accurate overall. Its prediction accuracy ranged from 42% to 68% for 500-step-ahead predictions in 3G networks.

D. Discussion

Our analysis of the relation between the degree of fluctuation of D_t and the prediction accuracy revealed that D_t in traces 3–5 was more unstable than in traces 1–2 (Fig. 7), and the overall prediction accuracy for traces 3–5 was lower than that for traces 1–2. We therefore compared the standard deviation of D_t throughout each trace and the prediction accuracy for the next 5 s in each trace. Fig. 10 shows negative correlation between the standard deviation and the prediction accuracy of the naïve and AR(8) predictors. However, since the prediction accuracy of JitSeer remained higher for traces 4–5 despite the large standard deviation, we believe JitSeer is more robust than other predictors to the fluctuation of D_t.

VII. Conclusion and Future Work

JitSeer is a long horizon method based on a GARCH model to predict the delay variation of IP packets. Experimental results over operational mobile networks show that the prediction accuracy of JitSeer exceeds that of conventional predictors; its prediction accuracy for the next 10 seconds was 74% over 3G and 88% over LTE.

In future work, we will conduct larger scale experiments spanning longer durations up to several months.

References

[1] F. Qian et al., "How to reduce smartphone traffic volume by 30%?," *Passive and Active Measurement (PAM)*, LNCS, pp. 42–52, 2013.

[2] C. Demichelis et al., "IP packet delay variation metric for IP performance metrics (IPPM)," IETF RFC 3393, 2002.

[3] H. Schulzrinne et al., "RTP: a transport protocol for real-time applications," IETF RFC 3550, 2003.

[4] E. H-K. Wu et al., "JTCP: jitter-based TCP for heterogeneous wireless networks," *IEEE JSAC*, Vol. 22, No. 4, pp. 757–766, 2004.

[5] P. Frojdh et al., "Adaptive streaming within the 3GPP packet-switched streaming service," *IEEE Network*, Vol. 20, Issue 2, pp. 34–40, 2006.

[6] M. P. Visser, "Volatility proxies and GARCH models," *Doctoral Dissertation*, University of Amsterdam, Amsterdam, Netherlands, 2009.

[7] T. Bollerslev, "Generalized autoregressive conditional hetero-skedasticity," *Journal of Econometrics*, Vol.31, pp. 307–327, 1986.

[8] J. G. D. Gooijer et al., "25 years of time series forecasting," *Int'l Journal of Forecasting*, Vol. 22, Issue 3, pp. 443–473, 2006.

[9] A. L. Corte, A. Lombardo, S. Palazzo and G. Schembra "A feedback approach for jitter and skew enforcement in multimedia retrieval services," *IEEE GLOBECOM*, Vol. 2, pp. 790–794, 1995.

[10] H. Sun et al., "End-to-end delay prediction by neural network based on chaos theory," *IEEE Int'l Conf. on Wireless Communications Networking and Mobile Computing (WiCOM)*, pp. 1–5, 2010.

[11] C-C. Wu et al., "An empirical evaluation of VoIP playout buffer dimensioning in Skype, Google talk, and MSN Messenger," *ACM NOSSDAV*, pp. 97–102, 2009.

[12] N. Aoki, "VoIP packet loss concealment based on two-side pitch waveform replication technique using steganography," *IEEE TENCON*, Vol. 3, pp. 52–55, 2004.

[13] S. Li et al., "An interference-aware routing algorithm for multimedia streaming over wireless sensor networks," *Int'l Journal of Multimedia & Its Applications*, Vol. 2, No. 1, pp. 10–30, 2010.

[14] H. Nishiyama et al., "Wireless loss-tolerant congestion control protocol based on dynamic AIMD theory," *IEEE Wireless Communications*, Vol. 17, Issue 2, pp. 7–14, 2010.

[15] A. Afanasyev, N. Tilley, P. Reiher and L. Kleinrock, "Host-to-host congestion control for TCP," *IEEE Communications Surveys & Tutorials*, Vol. 12, Issue 3, pp. 304–342, 2010.

[16] P. Frossard and O. Verscheure, "Joint source/FEC rate selection for quality-optimal MPEG-2 video delivery," *IEEE Transactions on Image Processing*, Vol. 10, Issue 12, pp. 1815–1825, 2001.

[17] J. Padhye et al., "Modeling TCP throughput: a simple model and its empirical validation," *ACM SIGCOMM*, pp. 303–314, 1998.

[18] M. Mirza et al. "A machine learning approach to TCP throughput prediction," *ACM SIGMETRICS*, pp. 97–108, 2007.

[19] Q. Xu et al., "PROTEUS: network performance forecast for real-time, interactive mobile applications," *ACM MobiSys*, 2013.

[20] A. C. Begen et al., "Predictive modeling of video packet delay in IP networks," *IEEE ICIP*, pp. 1325–1328, 2006.

[21] Q. Li and D. L. Mills, "Jitter-based delay-boundary prediction of wide-area networks," *IEEE/ACM TON*, Vol. 9, No. 5, pp. 578–590, 2001.

[22] A. G. Parlos, "Identification of the Internet end-to-end delay dynamics using multi-step neuro-predictors," *INNS/IEEE Int'l Joint Conf. on Neural Networks*, Vol. 3, pp. 2460–2465, 2002.

[23] M. Yang et al., "Predicting Internet end-to-end delay: an overview," *IEEE Southeastern Symposium on System Theory*, pp. 210–214, 2004.

[24] V. Bui et al., "Long horizon end-to-end delay forecasts: a multi-step-ahead hybrid approach," *IEEE ISCC*, pp. 825–832, 2007.

[25] V. Paxson and M. Allman, "Computing TCP's retransmission timer," IETF RFC 2988, 2000.

[26] M. Laner et al., "A comparison between one-way delays in operating HSPA and LTE networks," *IEEE/IFIP WiOpt*, pp. 286–292, 2012.

[27] D. Niu et al., "Risk management for video-on-demand servers leveraging demand forecast," *ACM Multimedia*, pp. 1229–1232, 2011.

[28] B. Zhou et al., "Traffic predictability based on ARIMA/GARCH model," *IEEE NGI*, pp. 200–207, 2006.

[29] N. C. Anand et al., "GARCH — non-linear time series model for traffic modeling and prediction," *IEEE NOMS*, pp. 694–697, 2008.

[30] S. Kim, "Forecasting internet traffic by using seasonal GARCH models," *Journal of Communications and Networks,* Vol. 13, Issue 6, pp. 621–624, 2011.

[31] E. Zivot, "Practical issues in the analysis of univariate GARCH models." *Handbook of Financial Time Series*, T. G. Andersen et al., Ed. Springer-Verlag Berlin Heidelberg, pp. 113–155, 2009.

[32] P. R. Hansen and A. Lunde, "A forecast comparison of volatility models: does anything beat a GARCH(1,1)?," *Journal of Applied Econometrics,* Vol. 20, Issue 7, pp. 873–889, 2005.

[33] J. Prins, *NIST/SEMATECH e-Handbook of Statistical Methods, 6.4.3.2. Forecasting with Single Exponential Smoothing*, [Online]. Available: http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc432.htm

[34] J. Matta et al., "A source and channel rate adaptation algorithm for AMR in VoIP using the Emodel," *ACM NOSSDAV*, pp. 92–99, 2003.

[35] Traffic Research Group webpage, *Heterogeneous Networks Data Traces*, University of Napoli Federico II, [Online]. Available: http://traffic.comics.unina.it/Traces/dtraces.php

[36] A. Botta, A. Pescape and G. Ventre "Quality of service statistics over heterogeneous networks: analysis and applications," *European Journal of Operational Research,* Elsevier, Vol. 191, Issue 3, pp. 1075–1088, 2008.

[37] R. J. Hyndman, "Another look at measures of forecast accuracy for intermittent demand," *Foresight: the Int'l Journal of Applied Forecasting,* Issue 4, pp. 43–46, 2006.