Rate Measurement Over Short Time Scales in Stationary Cellular Receivers

Habtegebreil Haile, Per Hurtig, Karl-Johan Grinnemo

Department of Mathematics and Computer Science

Karlstad University

Karlstad, Sweden

firstname.surname@kau.se

Abstract—An increasing number of cellular congestion control algorithms (CCAs) are becoming reliant on measurements of the delivery rate observed at the receiver. Accordingly, early detection of changes in the receiver's rate would improve the performance of such algorithms. In addition to CCAs, faster detection of rate can also benefit available throughput estimation tools that rely on rate measurements. The upper layers of a cellular receiver could achieve faster rate detection through rate measurements over short time intervals. However, for cellular receivers, upper-layer rate measurements over short time scales produce unreliable results due to the effect of underlying lower layer mechanisms such as scheduling and retransmissions. In this paper, we introduce a Kalman filter based rate estimation approach that reduces the variability observed in short time scale receiver rate measurements and allows faster rate change detection. We also integrate an adaptive mechanism to facilitate online estimations in a network with an unknown or changing characteristic.

Index Terms—Cellular, Rate estimation, Kalman filter

I. INTRODUCTION

The basic form of rate calculation involves adding up the amount of data received and dividing it by the time interval it took to receive the given amount of data. However, different rate estimation approaches diverge from each other based on where the rate is estimated and the time interval for which the rate is estimated. Some algorithms also apply some processing and aggregation mechanisms over one or more rate samples to infer a stable rate that is immune to some of the temporary variations experienced in networks [5] [9]. In cellular networks, the available rate to a specific user is, however, rarely constant and varies depending on various factors. The rate variability has made a timely estimate of the receiver's rate an important input for properly regulating how the packets are sent in recent congestion control algorithms.

Rate estimates computed over longer intervals might not give accurate estimates of the current state of a cellular network. Thus, it can be argued that in cellular networks, estimations done over short intervals are preferable to quickly detect the available rate and adjust sending rates accordingly. However, upper-layer reception rate calculations done over a short time interval can be vulnerable to various temporary lower layer mechanisms that could result in bursts or starvations. Examples of such lower layer mechanisms are eNodeB scheduling and in-order delivery of Radio Link Control (RLC) data. Therefore, a reception rate obtained using short time intervals can experience large variations even when the system finds itself in a relatively stable condition.

In this paper, we propose an approach to reduce the impact of lower layer mechanisms on the rate measured by upper protocol layers over short intervals. The approach applies a Kalman filter to the short interval measurements and enables faster detection of changes in the delivery rate. The Kalman filter and its nonlinear variants have been used in a number of cases related to cellular and wireless communication. Some of the previous uses of the Kalman filter include mobility tracking [16] [12] and load estimation [4]. To the best of our knowledge, this is the first approach to use the Kalman filter for making fast and accurate rate estimations from short time-scale measurements in cellular receivers.

The rest of the paper is organised as follows: Section II reviews recent rate estimation approaches. Section III presents the rate estimation issues considered by this paper. Section IV describes the proposed estimation approach. Results of the proposed estimation are presented in Section VI, and we conclude in Section VII.

II. BACKGROUND AND RELATED WORK

Estimating a receiver's rate has become an important component of a number of recent transport layer cellular congestion control algorithms (CCAs). As such, these types of CCAs can be viewed as entities performing micro-monitoring to enact changes that enable them to get better service from the underlying network. The estimation approaches differ, among others, in where they are placed and the time scale used for rate computation. Rate estimation algorithms can be implemented on the sender's side or on the receiver's side of the communication. Sender side implementations use packet acknowledgements to make estimations about the delivery of packets at the receiver [6] [7]. However, to obtain more reliable sender side receiver rate estimations, provisions for dealing with issues such as data-link asymmetry and acknowledgement (ACK) aggregation are needed [6]. On the other hand, receiver side rate estimations allow a more accurate rate computation, since they have immediate access to the actual count of data being delivered instead of indirectly relying on the response acknowledgements [15] [17].

In addition to the placement of the rate estimation algorithm, the time scale used to compute the rate estimation also varies
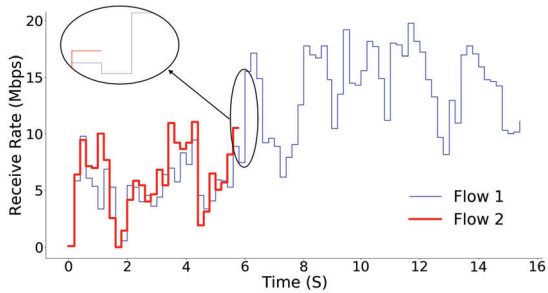
Fig. 1: Rates calculated over 200 ms

Fig. 2: Rates calculated over 20 ms

among the algorithms. Decisions on the time scale used for measuring the rate at a receiver is typically justified through theoretical [7] [17] or empirical reasoning [5] [13]. The time scale for calculating the receiver rate can be dynamically changing or statically set.

There are a number of algorithms where the rate calculation interval is related to the round trip time (RTT). The delivery rate estimation employed by the Bottleneck Bandwidth and Round-trip propagation time (BBR) algorithm [5] [6] takes samples of the delivery rate on an interval that lasts between the the ACK reception time just before the transmission of a data packet and the reception of its acknowledgement. Therefore, the granularity of the rate computation varies depending on the round trip time (RTT). Client Driven Bandwidth Estimation (CDBE) [17] is another algorithm with dynamic rate computation intervals. The algorithm aggregates multiple short interval rate estimations over a longer time window. The longer interval is set to a single RTT, and each long interval comprises five short intervals. Performance-oriented Congestion Control (PCC) [7] extracts the rate from ACKs received over a monitoring interval (MI), which is at least a value between 1.7 and 2.2 times the RTT.

PropRate [9] calculates an estimate of the receiver's rate at the sender. However, unlike BBR and PCC, the rate computation interval is not dependent on the RTT. The estimation is done by taking an exponentially weighted moving average (EWMA) of instantaneous rates over a time period. The time period is set to the duration of 50 consecutive packet bursts with a maximum limit at 500 ms. On the other hand, Sprout [15] sets a static tick period of 20 ms to calculate the rate to be used for updating a probability distribution of rates.

It should be noted that this work is not an alternative estimation mechanism to the above approaches. In fact, it is more of a complementary approach that could improve the rate samples collected by the above algorithms. This can particularly be useful for scenarios where the time scale over which algorithms collect a rate sample is dependent on the RTT, and the connection happens to be between nodes in close proximity to each other. In such cases, this approach could help in avoiding adverse effects from aggregations that use maximum filter or statically defined weight parameters. Some algorithms might choose to use longer intervals to avoid
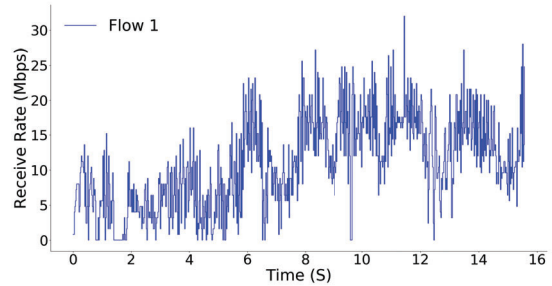
the high variations experienced in short time scales. For such cases, this approach enables the use of short intervals and allows a faster detection of a change in the packet delivery rate resulting from a change in cellular network conditions. Another application area of fast rate measurement techniques is in network measurement and monitoring tools. Fast estimation techniques can reduce the time required to saturate the network to obtain a reliable measurement. Real time monitoring of throughput to observe the immediate effect of changes in the network condition is another practical application that could integrate fast rate estimation techniques.

III. PROBLEMS WITH UPPER LAYER RATE ESTIMATION

The use of longer time scales for computing the receive rate of a cellular node will make the detection of the current capacity take longer. This is because a number of the packets used in the calculation of the rate were received in a period preceding the change. Figure 1 and 2 show the rates computed for two different time scales from the same trace collected using the setup described in Section V. Figure 1 shows the upper-layer receive rate of a saturating UDP flow that lasts about 15 seconds. A second UDP flow was launched at the same time as the first one and ended 5 seconds later. The rate is calculated at every 200 ms interval. It can be seen that there is a delay between the time when *flow2* ends and *flow1* detects the new higher rate.

Figure 2 shows that upper-layer rate measurements performed over very short time intervals (20 ms) are extremely variable. This variability is caused by the effect of lower layer mechanisms such as eNodeB scheduling and in-order delivery. Therefore, it is difficult to obtain a clear understanding of the steady rate at the receiver. This variability can get worse if measurements are done over an ever shorter time interval. The CDF plots in Figure 3 show the different levels of variability experienced when different measurement time scales are applied on the arrival trace used in Figures 1 and 2. It can be seen that the proportion of extremely high or low rates increases as the time scale of measurement decreases.

IV. SYSTEM DESCRIPTION

A. System and measurement models

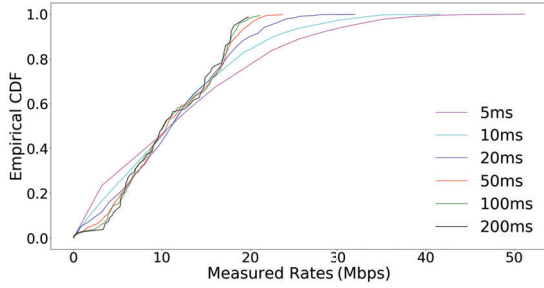In this section we describe the standard Kalman filter based approach used to minimise the high level of variations

Fig. 3: ECDF of rates for different sampling time scales

$$K = p_k^- C (C p_k^- C^T + R)^{-1} \qquad (5)$$

$$x_k = A x_k^- + K(y_k - C A x_k^-) \qquad (6)$$

$$p_k = p_k^- - K C p_k^- \qquad (7)$$

In current cellular networks, the achieved receive rate is determined by the signal quality of the channel and the amount of load in the network. In addition, the sending rate at the source will have to be greater than the available rate determined by the channel quality and network load. Thus, the above three quantities are the control of the system. Having full knowledge of the change in the above quantities would allow better tracking of the dynamics of the rate. However, since it is difficult to determine the amount of load in the network from a cellular receiver, the contribution of the load variation can be included as an uncertainty affecting the state transition, i.e., process noise.

In 4G/LTE networks, the change in the receive rate resulting from change in the channel quality is dependent on the type of scheduling employed by the eNodeB. However, it has been shown in [3] that when a network is loaded most eNodeB schedulers result in a sigmoid-like relation between the achieved rate and the channel quality. Since we are only considering a node at a static position, the change in signal quality resulting from propagation loss is relatively low. Additionally, we assume that the scale of the external interference is limited to a linear region of the scheduler dependent sigmoid-like relation. Thus, we opt for a simplified and generalised model that includes the rate variations resulting from channel quality in the uncertainty of the state transition model.

Based on the above conditions Equations (1) and (2) can be simplified as:

$$r_k = r_{k-1} + w_k \qquad (8)$$

observed in rate measurements over short time scales. The Kalman filter [8], is an optimal estimator for linear dynamical systems with state transition is given by Equation (1), and the observation of the state is expressed as in Equation (2). We chose the standard Kalman filter because it is: (1) easy to implement and understand; (2) relatively low cost computations allowing online adjustments of sending behaviour, if used in an application or a CCA; and (3) adaptable to different measurement procedures and network conditions by adaptively changing the parameters of the filter online.

$$x_k = A x_{k-1} + B u_k + w_k \qquad (1)$$

$$y_k = C x_k + v_k \qquad (2)$$

In Equations (1) and (2), x_k and x_{k-1} are N_x dimensional vectors representing the state of the system at times k and $k-1$, respectively, with transition matrix A. The contribution of the control vector, u_k, to the next state of the system is determined by the control-input matrix, B. The process noise, w_k, characterizes the uncertainty in the state model. The measured state, y_k, is the observable quality of the system. The translation between the state of interest and the observable state is given by C. v_k is the measurement noise, which quantifies the uncertainty in the measurement process.

$$x_k^- = A x_{k-1} + B u_k \qquad (3)$$

$$p_k^- = A p_{k-1} A^T + Q \qquad (4)$$

Given Equations (1) and (2), the Kalman filter estimation works by acquiring a new a priori state estimate (x_k^- in Equation (3)) using the state transition model. Then the a priori estimate is corrected using the measurement value to obtain an a posteriori estimate (x_k in Equation (6)). The uncertainty in the a priori estimate (p_k^-) is calculated in Equation (4) using the covariance of the process noise (Q). The uncertainty is then updated for the a posteriori estimate (p_k) in Equation (7). The Kalman gain (K), which is calculated by applying the measurement noise covariance (R), determines the contribution of the measurement to the a posteriori estimate.

$$m_k = r_k + v_k \qquad (9)$$

Where r_k is the steady rate at time k and m_k is the measured rate at time k.

B. Parameter selection

It is often difficult to have knowledge of the process and measurement noise and the corresponding covariance matrices Q and R. Furthermore, the values of Q and R might not stay steady, changing constantly throughout the measurement duration. In the particular case studied in this paper, the process and noise parameters are unknown and also expected to vary in time.

The value of Q will vary depending on the level of load and signal quality variation in the network. Since the scale of load and signal quality variability could differ between different time instances, the value of Q must be adjusted to reflect the change in variability. In addition, the value of R is dependent
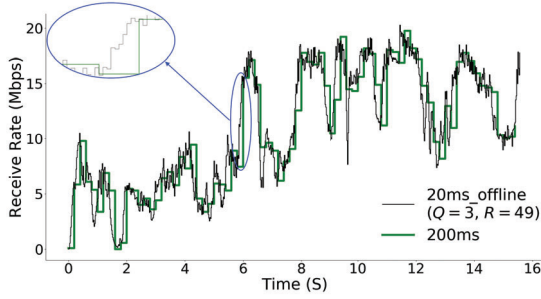
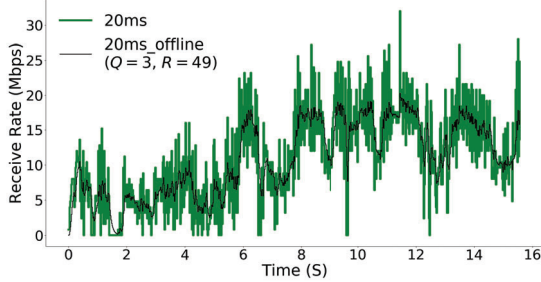Fig. 4: Long time-scale measurements vs. offline filtered output

Fig. 5: Unfiltered short time-scale measurements vs. offline filtered output

on the variability produced by the measurement process. This variability can depend on the type of scheduler, the time scale of rate measurement and the amount of packets available in the queue, among others. For instance, measurement over shorter time scales are more likely to be affected by lower layer mechanisms such as scheduling and retransmissions.

Measurement noise characteristics for Kalman filtering applied to device based measurement systems can usually be acquired from device specification and properties. Process noise is usually more challenging to characterise. Various approaches are applied by different applications to deal with unknown or variable noise parameters. These approaches range from trial-and-error (state space search) to more systematic Bayesian, maximum likelihood, correlation and covariance matching approaches [11].

We apply the noise parameter estimation mechanism presented in [1] to track and update the values of Q and R. The measurement noise covariance matrix R is estimated using the *residual* of the state estimation. The *residual*, ε, is the difference between the actual and estimated measurement values as given in Equation (10). Then, an estimate of R is obtained using Equation (11) .

$$\varepsilon_k = z_k - Cx_k \tag{10}$$

$$R_k = \alpha R_{k-1} + (1 - \alpha)(\varepsilon_k \varepsilon_k^T + Cp_k^- C^T) \tag{11}$$

The process noise covariance Q is obtained using the *innovation* of the system. The *innovation*, d, can be computed using Equation (12) as the difference between the measurement and the projection of the a priori estimate onto the measurement plane. Then, an estimate of Q can be obtained by using the *innovation* as shown in Equation (13). Detailed description of the derivation of the equations for Q and R can be found in [1] and [14].

$$d_k = z_k - Cx_k^- \tag{12}$$

$$Q_k = \alpha Q_{k-1} + (1 - \alpha)(K_k d_k d_k^T K^T) \tag{13}$$

V. Test Setup

The estimation approach is tested on data collected using the MONROE [2] platform. The MONROE platform consists of a collection of mobile devices distributed over several countries for measurement and experimentation on mobile broadband networks. The rate computations presented in this paper use data collected from sending UDP datagrams at a very high rate from a local server to a MONROE cellular receiver node.

The sender is a UDP application on a Linux version 4.4 machine that is connected to a 100 Mbps Ethernet connection. The application generates datagrams at a specified constant rate. The sending rate is set much higher than the cellular link capacity, to ensure there is always a queue in the cellular access network. The receiver is an application running inside a container on the MONROE node. A Linux operating system is installed on the node that is stationary to limit signal quality variations resulting from node mobility. The arrival time of each UDP datagram is recorded. The rate is calculated by summing the size of all received datagrams within an interval, and dividing the result by the selected time interval.

VI. Results

A. Offline parameter search

The noise parameters used in Figures 4 and 5 (Q=3, R=49) are obtained by exploring the search space on the same trace used for Figures 1 and 2. The filtered output for the measurements are shown against the long and short time scale measurements given in Figures 4 and 5, respectively.

The results shown in Figure 4 show a Kalman-filtered 20-ms rate estimates along with 200-ms rate measurements. It can be seen that the filtered short time-scale measurements almost always produce rate estimates that are close to the measurements obtained using the longer interval. However, the longer interval measurements are almost always lagging behind the filtered estimates. This shows that the filtered output removes the dragging effect imposed on long interval estimates by the inclusion of packet data that is much earlier than the time of interest.

The close-up image in Figure 4 shows the improvement in detecting the new receive rate when the flow to a second node is ended as shown in Figure 1. It can be seen that the filtered

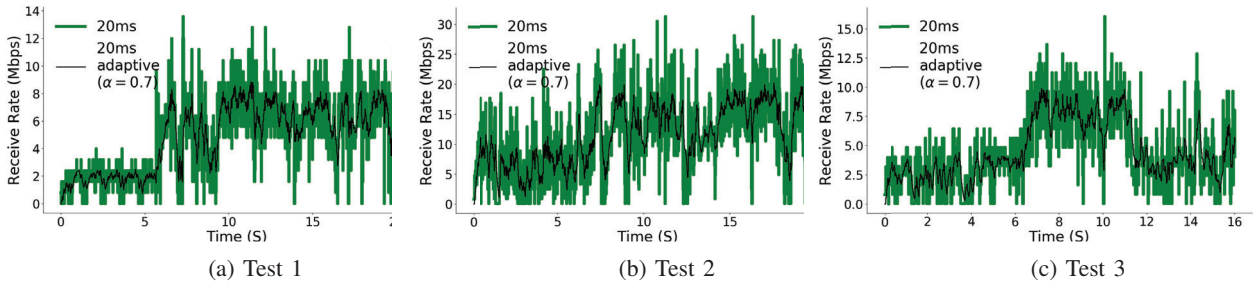(a) Test 1   (b) Test 2   (c) Test 3

Fig. 6: Filtered short interval rates vs. raw short interval rates for different tests
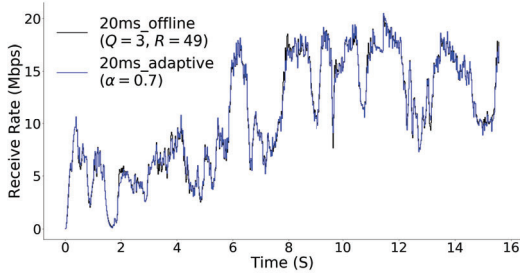

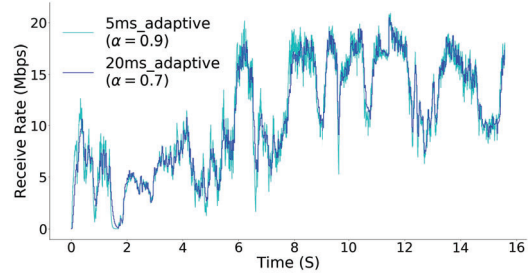Fig. 7: Adaptive estimation compared to basic offline estimation

Fig. 8: Adaptive estimation at different time scales

short interval approach is able to detect the available capacity much earlier than rates computed over the long interval.

In addition, Figure 5 shows that the filtered output is significantly less variable than the original measurements over short time interval. It can be seen that the filtered output is able to maintain the general progression of the short time-scale measurements while removing the extreme variations present in the original measurements.

B. Adaptive Estimation

The adaptive mechanism results in the filtered outputs given in Figure 6 (cf. the unfiltered short time-scale measurements) when applied to measurements done at three different time instances, where each instance is at a different location within the university building. It can be seen that the adaptive mechanism is able to track the receivers rate even with different unfiltered measurement sequences supplied as an input. In addition, Figure 7 shows that the adaptive mechanism is able to achieve similar performance to the offline approach. The value of α, which determines how fast Q and R, is updated is set to 0.7. Higher α updates the noise parameters slowly, and by varying α it is possible to get faster estimates but at the cost of higher variability.

Another advantage of the adaptive mechanism is that it can produce a good estimate from measurements over different time scales with relatively small and predictable changes to the filter parameter α. This is shown in Figure 8 where adaptively filtered measurements over 5 and 20 ms are compared. It can be seen that it is possible to get a good rate estimate even
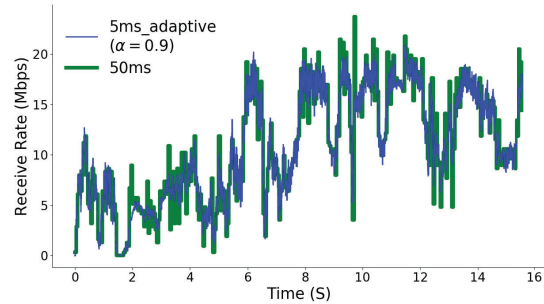

Fig. 9: Filtered estimations vs. unfiltered measurements using an intermediate time scale
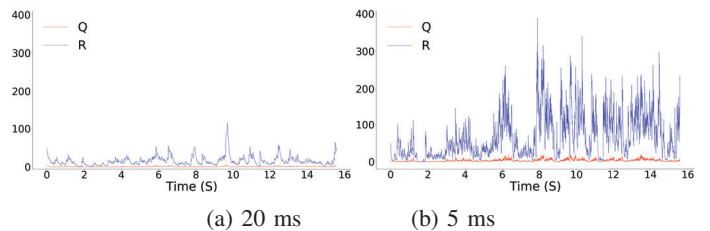

(a) 20 ms   (b) 5 ms

Fig. 10: Evolution of noise parameters for different measurement time scales

for extremely short rate sampling durations by making slight modification to the speed of update. Modifying α for different time scales is easier and more predictable than offline search for acceptable values of Q and R over the entire search space.

Figure 10 shows how Q and R are automatically updated for different measurement time scales. As expected, as the time scale of measurement decreases the measurement noise covariance R increases noticeably.

Applying the adaptive mechanism can also produce less variable results compared to raw measurements done over intermediate length intervals (e.g. 50 ms). As shown in Figure 9, 5 ms estimations experience less spikes and dips than the 50 ms measurements while achieving similar accuracy to the unfiltered measurements.

VII. Conclusion and Future works

In this paper, we presented a rate estimation approach for cellular receivers that applies the Kalman filter on short time-scale measurements to detect rate changes faster than long interval measurements. The approach is also able to reduce the extreme rate variations observed in short time-scale measurements. It is also shown that an adaptive estimation mechanism can be used to get estimates of the unknown filter noise parameters. Rate estimations from measurements over short intervals can facilitate faster reaction to network changes by algorithms that rely on receiver rate estimation to control the manner they send packets out at the sender. The approach could also benefit available bandwidth estimation tools that rely on rate measurements by allowing fast estimations that avoid saturating the network for longer intervals.

Future plans include further improvements on the speed and accuracy of the proposed approach. Such improvements can be achieved by using a system model that leverages cross-layer information, such as channel quality. A system dynamics based on the rate allocation scheme of a selected scheduler is one possible alternative for a model using cross-layer information. Further evaluations of the current online adaptation mechanism for non-stationary scenarios could also be performed to see if modifications are required for comprehensive use. Additionally, other online parameter adaptation mechanisms can also be evaluated to explore for a solution that is a better fit for the specific problem of cellular receiver rate estimation.

Although the approach shows promising results, the evaluations carried out so far are limited to traces collected on stationary receivers connected to a real operator network. We still plan to carry out wide scale measurements over multiple operator networks to further verify the applicability of the proposed approach. However, the network conditions that result in a specific set of rate measurements are not fully observable in a commercial operator's network. Therefore, a more definite evaluation of the approach in an environment that allows full control of the number of nodes in the network and the channel quality of each node will be performed to further verify the applicability of the proposed approach. The evaluations will also include mobility scenarios and other factors that affect signal propagation. Evaluating the performance of algorithms such as Sprout [15], CDBE [17] and CQIC [10] that could benefit from integrating the proposed estimation approach is also among the planned future tasks.

VIII. Acknowledgement

References

[1] S. Akhlaghi, N. Zhou, and Z. Huang. Adaptive adjustment of noise covariance in kalman filter for dynamic state estimation. In *2017 IEEE Power Energy Society General Meeting*, pages 1–5, July 2017.

[2] Ozgü Alay et al. Monroe: Measuring mobile broadband networks in Europe. In *Proceedings of the IRTF & ISOC Workshop on Research and Applications of Internet Measurements (RAIM)*, 2015.

[3] Y. Benchaabene, N. Boujnah, and F. Zarai. Performance comparison of packet scheduling algorithms for voice over IP in LTE cellular network. In *2016 4th International Conference on Control Engineering Information Technology (CEIT)*, pages 1–7, Dec 2016.

[4] G. Bianchi and I. Tinnirello. Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network. In *IEEE INFOCOM 2003 (IEEE Cat. No.03CH37428)*, volume 2, pages 844–852 vol.2, March 2003.

[5] Neal Cardwell et al. BBR: Congestion-Based Congestion Control. *Queue*, 14(5):50:20–50:53, October 2016.

[6] Yuchung Cheng et al. Delivery rate estimation. Internet-Draft draft-cheng-iccrg-delivery-rate-estimation-00, IETF Secretariat, July 2017.

[7] Mo Dong et al. PCC: Re-architecting Congestion Control for Consistent High Performance. In *NSDI*, volume 1, page 2, 2015.

[8] Rudolf Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering (ASME)*, 82D:35–45, 01 1960.

[9] Wai Kay Leong et al. TCP congestion control beyond bandwidth-delay product for mobile cellular networks. In *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*, pages 167–179. ACM, 2017.

[10] Feng Lu et al. CQIC: Revisiting Cross-Layer Congestion Control for Cellular Networks.

[11] R. K. Mehra. Approaches to adaptive filtering. In *1970 IEEE Symposium on Adaptive Processes (9th) Decision and Control*, pages 141–141, Dec 1970.

[12] L. Mihaylova et al. Mobility tracking in cellular networks using particle filtering. *IEEE Transactions on Wireless Communications*, 6(10):3589–3599, October 2007.

[13] Shinik Park et al. ExLL: An Extremely Low-latency Congestion Control for Mobile Cellular Networks.

[14] Jinling Wang. Stochastic modeling for real-time kinematic GPS/GLONASS positioning. *Navigation*, 46:297–305, 12 1999.

[15] Keith Winstein, Anirudh Sivaraman, and Hari Balakrishnan. Stochastic Forecasts Achieve High Throughput and Low Delay over Cellular Networks. In *Proceedings of the 10th USENIX Conference on Networked Systems Design and Implementation*, nsdi'13, pages 459–472, Berkeley, CA, USA, 2013. USENIX Association.

[16] Zainab R. Zaidi and Brian L. Mark. Real-time mobility tracking algorithms for cellular networks based on kalman filtering. *IEEE Transactions on Mobile Computing*, 4(2):195–208, March 2005.

[17] Z. Zhong et al. CDBE: A cooperative way to improve end-to-end congestion control in mobile network. In *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 216–223, Oct 2018.