Crowdsourcing-based Mobile Network Tomography for xG Wireless Systems

Ergin Dinc* **Mustafa Ozger*** **Ahmet F. Ates†** **Ibrahim Delibalta†** **Ozgur B. Akan***

*Next-generation Wireless Communications Laboratory (NWCL)
Department of Electrical and Electronics Engineering
Koc University, 34450, Istanbul, Turkey
{edinc, mozger, akan}@ku.edu.tr

†Turk Telekom Group
Istanbul, Turkey
{ahmetfeyzi.ates, ibrahim.delibalta}@turktelekom.com.tr

***Abstract*—Network size and number of mobile users are ever-increasing with the advancements in cellular network technologies. Hence, this situation makes the network monitoring highly complex. Although there are numerous network tomography approaches, service providers need real-time network monitoring tools to provide better network utilization. In this paper, we propose a crowdsourcing-based real-time network tomography framework. In the proposed framework, channel condition and user data usage are monitored via an application at the mobile terminals, and then the mobile terminals transmit their data to the server. In this way, the network and user behavior can be continuously monitored, and real-time actions can be implemented to improve the network performance. By using the proposed framework, we propose an optimization framework for the amount and reporting frequency of the transmitted data to avoid battery drain at the mobile terminal and network congestion. At the end, we provide simulation results for the proposed optimization framework.**

***Index Terms*—Crowdsourcing, Network Tomography, Wireless Systems, Network Management.**

I. Introduction

The evolution of cellular network technologies has increased the number of mobile users. There are over 4 million base stations worldwide to serve the mobile users [1], and many more will be deployed to satisfy the needs of the increased mobile users [2]. Furthermore, the increase has also changed the mobile devices. Mobile users are now capable of accessing the Internet, sending and receiving different types of data such as multimedia in addition to voice calls and text messages, which are the early services of mobile operators. These services differ according to the mobile devices and the mobile telecommunication in xG (2/5G) wireless systems.

The increase in the demand of data due to the large number of mobile phone users and the different services offered by the service providers cause problems related to network management. These problems may be degradation in quality of service (QoS) due to excessive load on a base station or the dissatisfaction of some of the users due to the heterogeneity of the telecommunication technologies of the users. This situation necessitates a real-time network monitoring framework, which monitors the network and take measures to increase QoS in the network [3].

Mobile operators use different approaches to analyze and optimize their networks. This includes testing the network, receiving data about the network by test equipment and identifying the network problems via big data analytics. There are different technologies to optimize the mobile cellular networks. ActixOne [5] is a multi-vendor and multi-technology optimization platform. It helps service providers for planning, managing and optimizing their mobile access networks to satisfy the certain level of QoS for their customers. It is used in 2G, 3G, LTE, and VoLTE technologies. Elastic-SON [6] is another solution for cellular networks for reducing operational costs of the network without decrease in delivery and service. It operates over different wireless technologies seamlessly. SmartAir [7] is a software suite to analyze the network behavior. It is deployed in network operation center, and it monitors how each cell performs in terms of quality of user experience. It analyzes and diagnose the problems in the network to boost the performance. Cerion optimiser [8] is a software application that plans the network, improves the capacity and provides best network configurations. The available network monitoring tools provide optimization in the network performance and efficient operation. However, they are commercialized solutions, and the access to their methods is restricted. In addition to commercially available tools, crowdsourcing-based methods [10] are proposed for network monitoring, and this type of methods has high potential to monitor the network real-time with smartphone-based systems.

TABLE I
Comparison for network tomography approaches.

Platform	2G Support	3G Support	4G Support	Fast Response	Crowd-sourcing
ActixOne [5]	✓	✓	✓	✓	X
Elastic-son [6]	✓	✓	✓	✓	X
SmartAir [7]	✓	✓	✓	✓	X
Cerion [8]	X	X	✓	✓	X
Our Approach	✓	✓	✓	✓	✓

In this paper, we propose a crowdsourcing-based real-time network tomography framework. It monitors the network by running an application at the mobile terminal and collecting data about the channel characteristics and data usage. The

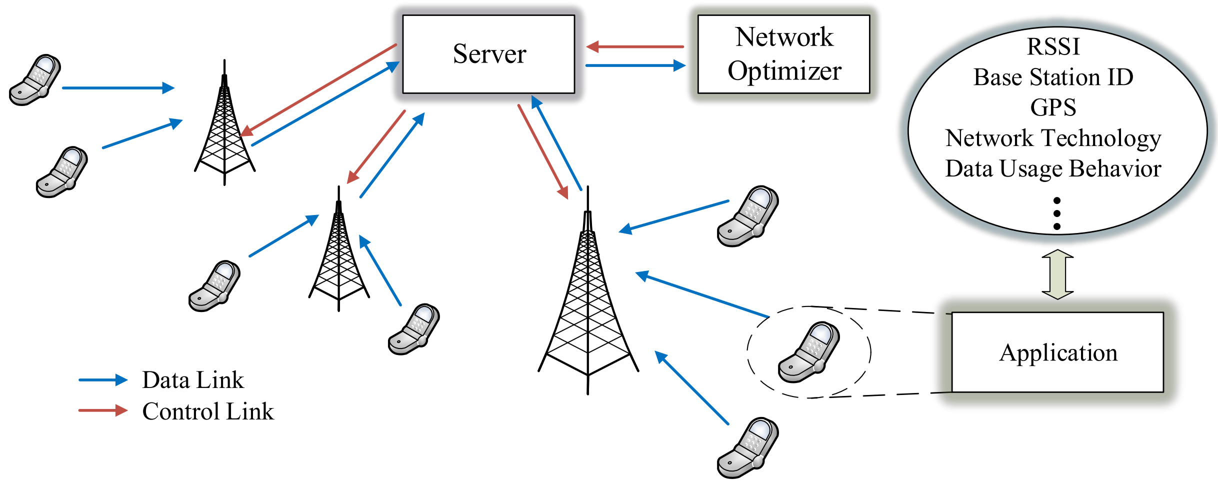

Fig. 1. System architecture of the proposed framework.

collected data by the users is sent to the server via base stations to detect the problems of the network in a timely manner and to take necessary measures accordingly. Storing the data provided by the mobile users can also be used to estimate the future failures and problems of the network. The service provider can solve the future problems beforehand. Thus, the proposed method have high potential to improve the network management compared to the other commercially available network tomography methods as in Table I. The mobile operator also uses the data to optimize the network at the server. To this end, we determine the wireless channel information extent we can use in mobile phones running Android Operating System (OS). Each user will report this information to the base stations. We optimize reporting frequency of the mobile terminals according to the estimation of the channel within certain level of accuracy by transmitting less packets. This will decrease the amount of data to be sent to the base station such that the mobile users do not suffer from the battery drain. Furthermore, we find the optimal number of reporting users to estimate the channel conditions within different confidence levels.

The remainder of this paper is organized as follows. In Section II, the system architecture is presented. Section III explains our network tomography framework. Section IV presents our simulation environment and simulation results. Finally, the concluding remarks are given in Section V.

II. System Architecture

The proposed system architecture can be seen in Fig. 1. We provide the detailed explanation of the system architecture in the following subsections.

A. Mobile Terminal

At the mobile terminal, users collect data related to the channel conditions and data usage behavior. Then, mobile terminals send their collected data to the base stations. However, transmission frequency and amount of transmitted data are critical to avoid battery drain that is highly undesirable for the users. For these reasons, the amount of transmitted data and reporting frequency will be optimized at the mobile terminal and server, respectively. On Android OS, the types and sizes of information about the channel and user are listed in Table II [9].

TABLE II
Information about the channel and user [9].

Network Information	Data Type	Data size (byte)
Received Signal Strength Indicator (RSSI)	*int*	*4*
Serving and neighboring cell Identities	*int*	*4*
GPS Location (lattitude, longirude)	*(int, int)*	*8*
Network technology (2G, 3G, 4G or Wi-Fi)	*int*	*4*
User cellular data usage (upload, download)	*(int, int)*	*8*
User Wi-Fi data usage (upload, downlad)	*(int, int)*	*8*
Timestamp	*long*	*8*

B. Server

Base station transmits the collected data from mobile terminals to their server as seen in Fig. 1. The server optimizes the network by the help of this data to take necessary actions. In addition, the collected data will be stored on the database to estimate future network traffic and failures. Some of the applications of the proposed framework can be listed as:

- **RSSI Monitoring:** Based on the RSSI readings by the mobile users. The RSSI distribution of network area can be determined. Based on this distribution, service providers can detect locations with low network coverage. In addition, new base stations or mobile base stations can be placed accordingly to maximize the throughput.
- **Interference Management:** Mobile users can monitor the control signals from both serving and neighboring base stations. The power of neighboring base station can be utilized to determine the interference levels, and these

analysis can be useful for decreasing the interference by frequency partitioning or power management.

- **Data usage behavior:** Our proposed application can help the service providers to categorize their users and Wi-Fi off-loading locations. For example, most of the users utilize Wi-Fi connection at home and work to provide longer battery life. In this case, these users do not need cellular data usage in these locations. Therefore, service providers can detect these regions and use this information while allocating their resources.
- **Real-time resource allocation:** Based on the user locations, service provider can continuously estimate the data traffic and take actions to provide better QoS. In case of concerts and social gatherings, cellular network may suffer from outage due to abrupt traffic increase. However, real-time monitoring through the proposed framework can also help service providers to estimate these traffic conditions and take the necessary actions before the service outage occurs.

The number of application can be further increased. With the deployment of this framework, adaptive control of the network can be maintained with big data and machine learning techniques.

III. Proposed Optimization Framework

As discussed in Section II, the proposed framework can be utilized in a number of different applications. In this paper, we utilize this framework to estimate the signal-to-noise ratio (SINR) distributions of base stations based on the data collected from users. However, transmission of this data can cause battery drain at the mobile terminals or create network congestion. Therefore, we propose an optimization framework to reduce the amount of transmitted data at the mobile terminal and server side as described in the following subsections.

A. Mobile Terminal

In the optimization framework, SINR readings at the mobile users are transmitted to the base station and base stations forward this information to the server. In this framework, the most accurate estimation can be performed by sending all the collected information to the server. However, transmitting large amount of data will cause battery drain at the mobile terminal which is highly undesirable. To this end, we propose an optimization framework such that only some portion of the collected data will be sent to the base stations to reduce the power consumption.

There is a trade-off between the amount of transmitted data and the accuracy of the estimations since it is not possible the transmit all of the collected data. For this reason, we propose a two-sample Kolmogorov-Smirnov test [12] based approach to reduce the information packets while satisfying desired accuracy. In statistics, the two-sample Kolmogorov-Smirnov test is utilized to determine whether the two data sets have the same probability distribution or not. Assume that $F_1(x)$ and $F_2(x)$ represent cumulative distribution functions (CDFs) of two arbitrary distributions as shown in Fig. 2. For

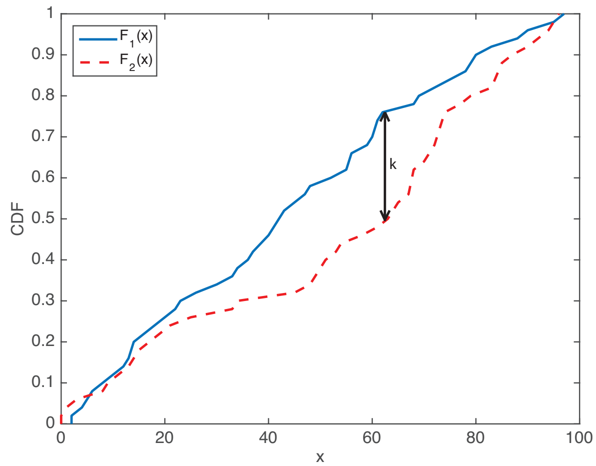

Fig. 2. An example presentation of Kolmogorov-Smirnov statistic for $F_1(x)$ and $F_2(x)$.

these two distributions, Kolmogorov-Smirnov statistic is the maximum amount of difference between CDFs that can be shown as

$$k = \sup_x |F_1(x) - F_2(x)|. \tag{1}$$

We use Kolmogorov-Smirnov statistic to find a sampling period for the collected data such that the difference between the original data and sampled data satisfies the desired threshold value. In this way, the amount of transmitted information can be reduced while providing required estimation accuracy at the serve side. Algorithm 1 shows the optimization framework at the mobile terminal. The collected data is represented as $\mathbf{D}$ and sampled data is represented as $\mathbf{S}$. The algorithm tries to find the sampling interval such that the Kolmogorov-Smirnov statistic is equal or slightly higher than the threshold value (k_{th}^M), where N_{max}^S is the maximum sampling interval. In this way, this algorithm can provide reliable estimate of the channel conditions and reduce the amount of transmitted data.

Algorithm 1 Pseudocode for Mobile Terminal Optimization

procedure MOBILE TERMINAL
  Require D
  Calculate CDF of $\mathbf{D}$: $F_D(x)$
    for $n = 2, \ldots, N_{max}^S$ **do**
      $\mathbf{S} = \mathbf{D}(1 : n : \mathrm{End})$
      Calculate CDF of $\mathbf{S}$: $F_S(x)$
      $k = \sup_x |F_D(x) - F_S(x)|$
      if $k \geq k_{th}^M$ **then**
        $N_S = n$
        Break
      end if
    end for
end procedure
Return: N_S, $\mathbf{S}$

B. Server

As described in Section III-A, the amount of transmitted data can be reduced by preprocessing at the mobile terminal.

However, transmission of the sampled data from all users may cause congestion in the network by generating excessive traffic. To tackle this problem, we also propose server optimization framework such that only some portion of the users reports their data to the base stations. In this way, both network congestion and power consumption can be further decreased.

In the proposed framework, the server side also utilizes the Kolmogorov-Smirnov test based approach as in Section III-A, and the server optimization framework is presented in Algorithm 2. Assume that the server initially has the data from all users for a base station, and this data is represented as

$$\mathbf{D}_{BS} = \bigcup_{i=1}^{N_U} \mathbf{D}_i, \tag{2}$$

where N_U is the number of users in the cell and $\mathbf{D}_i$ is the data of i^{th} user. In Algorithm 2, the server combines only the data of n number of randomly selected users to find the number of reporting users (N_R) that can provide SINR distribution of the cell with Kolmogorov-Smirnov statistic of k_{th}^{BS}. In this way, only N_R number of randomly selected users will send their data to the base station to lower network usage and power consumption. In addition, the complexity of data processing at the server will be decreased.

Algorithm 2 Pseudocode for Server Optimization

procedure SERVER
Require $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{N_U}$
$\mathbf{D}_{BS} = \bigcup_1^{N_U} \mathbf{D}_i$
for $n = 2, \ldots, N_U$ **do**
    Select n distinct number from $[1, N_U]$: $\mathbf{u}$
    $\mathbf{S}_{BS} = \bigcup_{j=1}^{n} \mathbf{D}_{\mathbf{u}(j)}$
    $k = \sup_x |F_{\mathbf{D}_{BS}}(x) - F_{\mathbf{S}_{BS}}(x)|$
    if $k \leq k_{th}^{BS}$ **then**
        $N_R = n$
        Break
    end if
end for
end procedure
Return: N_R

IV. Simulation Environment and Results

This section includes the simulation environment and results for the proposed framework.

A. Simulation Environment

We perform the simulations on MATLAB. The modeled network has 16km^2 area with 16 base stations placed as a grid as seen in Fig. 3. 1000 mobile users are randomly positioned in the network. We assume that the users can only communicate with the closest base station with the same power level. As the path-loss model, we utilize the empirical path-loss model given in [11]

$$PL = A + 10\gamma \log_{10}(d) + X_s, \tag{3}$$

where $A = 20 \log_{10}\left(\frac{4\pi d_0}{\lambda}\right)$ is the path-loss intercept, d_0 is the reference distance, λ is the wavelength, d is the distance between user and base station, γ is the path-loss exponent and

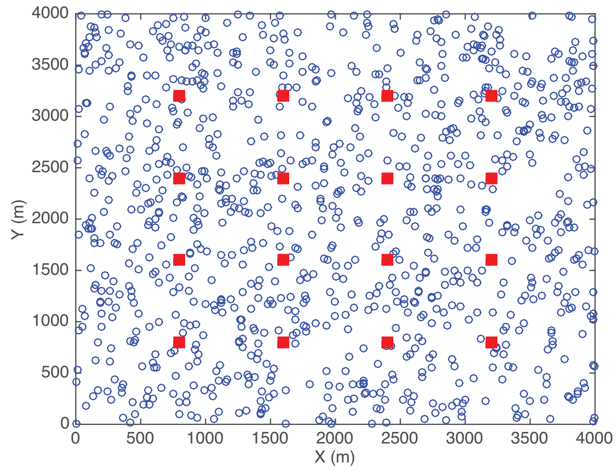

Fig. 3. Simulation environment, where circles represent mobile users and squares represent base stations.

TABLE III
SIMULATION PARAMETERS.

Parameter	Value	Unit
Carrier Frequency	*10.5*	*GHz*
Bandwidth	*10*	*MHz*
Transmitter Power	*0.1*	*W*
Path-loss Exponent	*2*	-
Bandwidth	*2*	*MHz*
σ	*4.3*	*dB*
d_0	*10*	*m*
v_{max}	*30*	*m/s*
T_M	*1*	*ms*

X_s is the shadow fading that has Gaussian distribution with zero mean and σ standard deviation. The noise is modeled as thermal noise. The utilized simulation parameters can be found in Table III. In addition, user mobility is modeled with updating d according to the speed and angle of the user. Both speed and angle of the users are uniformly distributed over $[0, v_{max}]$ and $[0, 2\pi]$, respectively. Mobile terminals measure the channel in every 1 ms (T_M), and reporting time is assumed as 5 s.

B. Simulation Results

Fig. 4 presents the Kolmogorov-Smirnov statistics vs. sampling period (N_S) 5000 measurement points. The red line shows the quadratic polynomial fitting to these points. As noticed, the sampling period of ≈ 90 can satisfy the desired Kolmogorov-Smirnov statistic value, so that the CDF of the sampled data may show maximum of k_{th}^{M} difference from the original distribution. With this sampling operation, instead of 5000 measurement point only ≈ 55 data point can be sent. Therefore, both the battery drain in the mobile terminal and network congestion can be avoided.

Fig. 5 shows the Kolmogorov-Smirnov statistics vs. number of reporting users for a cell in the network without mobile terminal optimization. As noticed, the data coming from 6 users can be utilized to estimate the SINR distribution in the

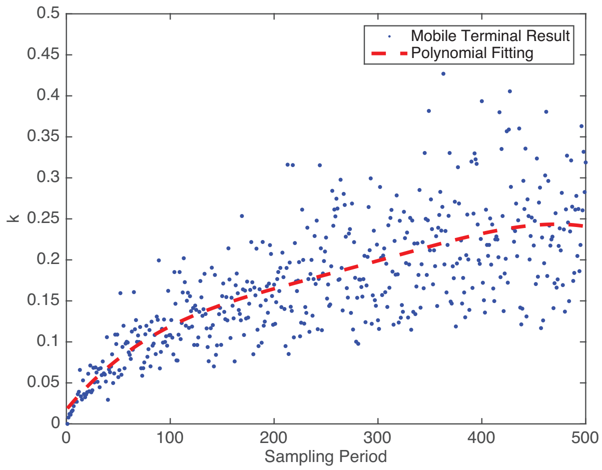

Fig. 4. Kolmogorov-Smirnov statistics vs. Sampling period at mobile terminal.

network while satisfying the Kolmogorov-Smirnov statistic of 0.1. In each reporting time, only 6 users can report their SINR readings instead of all of 42 users.

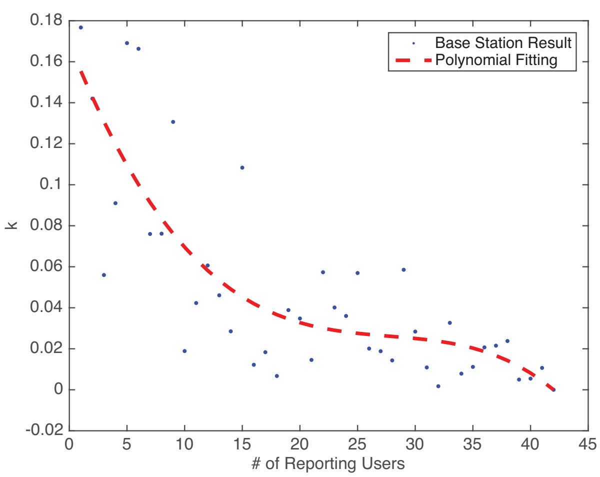

Fig. 5. Kolmogorov-Smirnov statistics vs. the number of reporting users at server side.

In Fig. 6 and Fig. 7, the number of reporting users are analyzed with both mobile terminal and server side optimization frameworks as presented in Algorithm 1 and Algorithm 2 such that only selected users sample their data and transmit to the base station. Since the preprocessing of the data introduces k_{th}^{M} difference, the same Kolmogorov-Smirnov statistic level ($k_{th}^{BS} = 0.1$) can be maintained with higher number of users. However, the required number of users increase 6 to 7 and 8 for $k_{th}^{M} = 0.1$ and $k_{th}^{M} = 0.2$, respectively as in Fig. 6 and Fig. 7. However, the proposed framework is still opportunistic because the sampled data only includes 1.8% and 0.3% of the measurement points for $k_{th}^{M} = 0.1$ and $k_{th}^{M} = 0.2$, respectively as in Fig. 5. Therefore, mobile terminal and server side optimization frameworks can provide significant reduction in the amount of transmitted data, and also reduce the complexity of network monitoring framework.

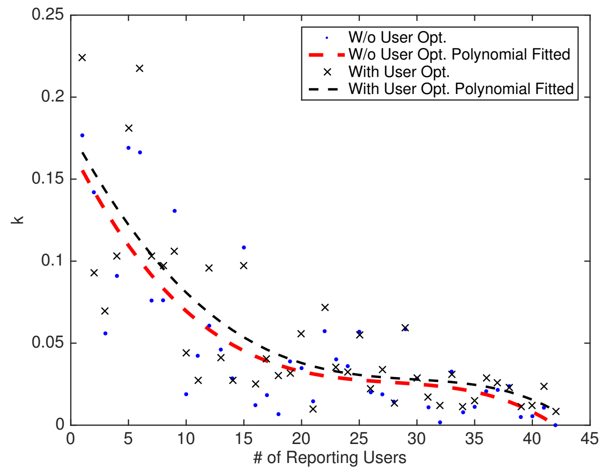

Fig. 6. Performance comparison of the system with both mobile terminal and server optimization frameworks with $k_{th}^{M} = 0.1$.

According to Table II, the size of the collected data at one measurement instant is equal to 44 bytes. In our simulations, each user collects 5000 measurements, and there are 42 users in the cell located at $(1600, 2400)$. Without any optimization, the total data size received by the base station will be # of users $\times$ # of measurements $\times$ data size ≈ 8.8 GB at each reporting interval. As seen in Fig. 6, the same cell can be monitored with Kolmogorov-Smirnov statistic of 0.1 by 7 randomly selected users each sending 55 data points. Hence, the total data size received by the base station becomes approximately 16.5 MB in each reporting period. For $k_{th}^{M} = 0.2$ in Fig. 6, the network can be monitored with 8 randomly selected users each sending 17 data points, and the total data received at the base station becomes 5.8 MB. Therefore, the network tomography data becomes manageable with the proposed optimization framework.

V. Conclusion

In this paper, we propose a novel network monitoring framework for cellular networks. The proposed approach is crowdsourcing-based since the mobile users monitor the channel and send the channel information and data usage to the server. At the mobile terminal, we find energy optimal reporting frequency by employing two-sample Kolmogorov-Smirnov test based approach. At the server, after the server receives all the information related to the channel and data usage from the users, it optimizes the number of mobile users in a base station to estimate the channel conditions within the certain confidence level.

Acknowledgment

This work is funded by Turk Telekom Group under the Grant No: 4020000695.

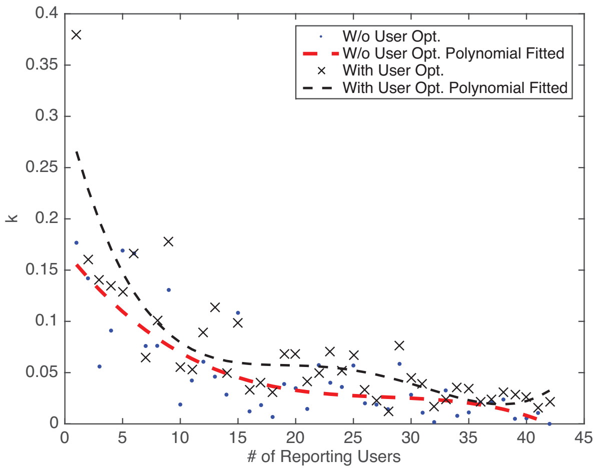

Fig. 7. Performance comparison of the system with both mobile terminal and server optimization frameworks with $k_{th}^{M} = 0.2$.

REFERENCES

[1] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green Cellular Networks: A Survey, Some Research Issues and Challenges," *IEEE Communications Surveys and Tutorials,* vol. 13, no. 4, pp. 524-540.

[2] I.Z. Kovacs, P. Mogensen, B. Christensen, and R. Jarvela, " Mobile broadband traffic forecast modeling for network evolution studies," in Proc. *IEEE Vehicular Technology Conference*, 2011, pp. 15.

[3] Y. Huang, N. Feamster, and R. Teixeira, "Practical issues with using network tomography for fault diagnosis," in *Proc. ACM SIGCOMM Computer Communication Review,* vol. 38, no. 5, pp. 53-58, 2008.

[4] MobiThinking, "*Global mobile statistics 2012 Home,"* 2013, http://mobithinking.com/mobilemarketing-tools/latest-mobile-stats/.

[5] Actix, "*ActixOne,*" http://www.actix.com/. [Online].

[6] Cellwize, "*Elastic-SON,*" http://www.cellwize.com/. [Online].

[7] Vasona, "*SmartVision,*" http://www.vasonanetworks.com/#/platform/smartair. [Online].

[8] Cerion, "*Cerion Optimiser,*" http://www.cerioninc.com/docs/datasheet_optimization.pdf [Online].

[9] Android, "*Telephony API,*" http://developer.android.com/reference/android/telephony/package-summary.html

[10] A. Faggiani, E. Gregori, L. Lenzini, S. Mainardi, and A. Vecchio, "On the feasibility of measuring the Internet through smartphone-based crowdsourcing," in *Proc. of WiOpt12*, pp. 318323, May 2012.

[11] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1205-1211, Jul 1999.

[12] I. T. Young, "Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources," *J. Histochem. Cytochem.*, vol. 25, no. 7 pp. 935-941, 1977.