Optimizing Crowd based Monitoring in Large Scale Mobile Networks

Jan Groenendijk
Network Management Lab
LM Ericsson Ireland
Athlone, Co.Westmeath, Ireland
jan.groenendijk(at)ericsson.com

Yangcheng Huang
Network Management Lab
LM Ericsson Ireland
Athlone, Co.Westmeath, Ireland
yangcheng.huang(at)ericsson.com.

Abstract—**This paper presents a method using terminal reports for 24/7 monitoring by selecting an optimum set of subscribers that can be incentified and let their devices become measurement points for a service provider. This method provides benefits for both subscriber and service provider and applies to all services provided over large scale networks. The method presented is work in progress.**

Keywords-component; crowd; monitoring; terminal measurements; large scale network; service quality;

I. Introduction

Service providers are increasingly looking for ways to get feedback from the subscribers on the service delivery [1]. To monitor the service delivery of the network based on terminal measurements, sufficient measurements needs to be available to reflect the network quality. Terminals report measurements that are used to calculate key performance indicators (KPI) to reflect accessibility, retainability and integrity of the delivered services such as video/browsing/streaming/VoIP.

For large-scale network deployments of terminal measurement the industry needs to solve the following challenges:

1) Intrusion introduced by terminal based measurements and reporting of subscriber terminal usage (in terms of resource usages, including battery and bandwidth, and required human intervention);
2) Inaccurate measurements due to complexity of networks (i.e. re-active traffic-aware network in resource allocation) and insufficient considerations on probing traffic;
3) Being unable to handle low or partial penetration coverage of terminal measurement and reporting; existing solutions work on terminals in an isolated way, without coordination between reporting instances, or any holistic view of the penetration coverage of the network. Collected reports from devices cannot be used unless per session correlation with network measurement is possible.

In particular, intrusion introduced by terminal based measurements and reporting can be categorized from the following aspects:

1) Willingness of downloading and activating of terminal reporting application.
2) Resource consumption in in mobile device due to terminal reporting.
3) Charging for network usage by mobile device due to terminal reporting.

These factors make the actual population of potential measurement points (for crowd based terminal reporting) very small and therefore hindering the introduction of network wide service delivery monitoring.

Existing work that discusses the problems with crowd based monitoring and service delivery can be found in [2]. The solution described in this paper takes all the above factors into account and provides a method for optimization and management of the measurement point population. The proposed solution for 24/7 monitoring, includes:

- A method of automatically identifying subscribers as measurement points by correlating network topology, subscriber mobility and service usage, and determining a measurement efficiency factor for each subscriber
- A method of targeted incentive mechanisms by analysing subscribers' service usage in building a network of terminal measurement points.
- A method of scheduling measurements and reporting for identified subscribers based on the identified coverage of each subscriber, and improving measurement accuracy based on a holistic coverage view of the terminal based reporting activities across multiple measurement points

II. Solution Overview

The proposed methods are aimed to identify devices that are potential measurement points by analyzing the subscriber's mobility, the service usage and the measurement efficiency. From this activity a potential list of terminals is generated and those subscribers are targeted with an incentive. Acceptance of the incentive will result in classification of devices as measurement point.

A measurement point will be activated and subject to a measurement schedule optimized to the (at that time) known behavior of that device and subscriber. Figure 1 shows an architectural overview of the solution, and their relations with existing solutions (or products).

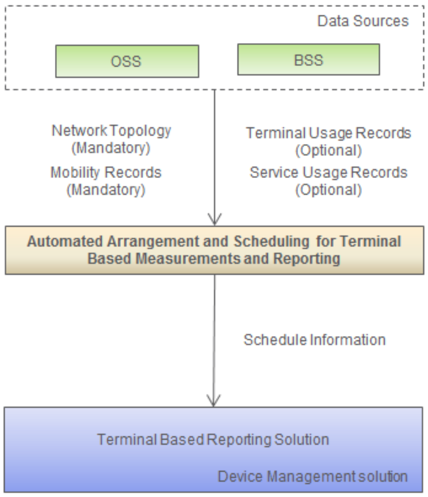

Figure 1. Architecture overview

III. Description of Algorithms

A. Automated Identification of Measurement Points

The identification of potential measurements does not initially require the terminal to be a measurement point. The network information provided by a diverse set of Operations and Business Support Systems (OSS/BSS) can be used to build an individual temporal and geographical coverage picture of all subscriber and terminal combinations.

A subscriber may have multiple terminals and may make that subscriber an efficient target for providing terminal measurement reports (via incentification). The following paragraphs describe the measurement point identification algorithm, inputs and outputs.

<u>Algorithm Input</u>

The following data sets shall be considered as inputs for the proposed identification method:

- Topology of an operational network (mandatory); network topology gives the monitoring requirements, i.e. a list of cells to be monitored by the terminal based reporting solution.
- Subscriber mobility records (mandatory). 3GPP TS 23.401 defines mobility management procedures; mobility management for access networks keeps track of a User Equipment (UE) in the radio access networks in terms of location areas. The records related to mobility management can be obtained either through an event based monitoring solution, or by tapping the signaling traffic sent on S1-MME interface.
- Subscriber service usage records (optional), depending upon whether service monitoring is considered as a factor in measurement point identification. Service usage records can be obtained via Deep Packet Inspection (DPI) functions running on G_n interfaces

In addition, at least the following monitoring requirements shall be considered as input parameters:

- Minimum number of subscribers a monitored target shall be covered by.

<u>Algorithm Output</u>

A list of measurement points S= $\{S_1, S_2, S_3\ldots S_n\}$ shall be identified, with the cells covered by these measurement points. In particular, a measurement point may be identified for a particular cell, or a group of cells, and for a particular time window. The following table shows an example of the coverage map output:

Cells (or Cell Groups)

	08:00 – 09:00	09:00– 10:00	10:00-11:00	…
Cell_12300	S_1, S_9, S_{11}	S_2, S_9, S_{20}	S_1, S_4, S_9	…
Cell_Group_1	S_2, S_{11}, S_{21}	S_3, S_4, S_{10}	S_5, S_6, S_8	…
Cell_13450	S_2, S_8, S_9	S_3, S_4, S_5	S_3, S_5, S_7	…
…	…	…	…	…

Table 1. Automated Identification of Measurement Points: Coverage Map Output

New dimensions, such as different service/application types video/browsing/streaming/VoIP etc. may be added to the output according to the monitoring requirements. Measurement points shall be identified for a particular cell, or a group of cells, for a particular time window, and for a particular service. Such outputs are used as inputs for the targeted incentive & scheduling steps.

<u>Algorithm</u>

STEP1: Determine frequent visitors per cell (or cell groups) per observation time window.

The purpose of this step is to identify periodic visitors with frequent mobility, whose mobility and service usage patterns are predictable, and therefore feasible as measurement points. Only *predictable* mobility and service usage shall be considered in measurement point determination.

Those subscribers with unpredicted mobility patterns, or the unpredictable mobility of a particular subscriber, shall be filtered out. That is, at least the following criteria are considered in identifying measurement points:

- High mobility, with maximized reporting coverage, i.e. covering as many cells as possible by a single measurement point;
- Periodic patterns in mobility, so that terminal based measurements are repeatable to carry out 24/7 monitoring tasks.

To achieve this, subscribers are evaluated by calculating the following reporting efficiency factor based on the criteria above. Efficiency factor of a subscriber is to quantify the efficiency of using a subscriber as a measurement point for terminal based measurements and reporting. The calculation algorithm is as follows:

```
Input:      mobility records of the subscriber

Output:     efficiency factor of the subscriber

Let C i, j be the counter for the number of visits (of a subscriber) on cell i during the
observation window j (as shown in the table 2);

For each observation window j

            Get the list of cells visited during j;

            For each cell i in the list

                        Calculate the number of new visits m i, j on cell i during j;

                        C i, j <- C i, j + m i, j;

            For each counter Cx, j in the observation window j

                        Cx, j <- Cx, j – delta (delta = 1 by default)

                        If (Cx, j < 0)

                                    Cx, j <- 0;

Keep feeding records into the calculation above;

Sum up all counters of a subscriber as the efficiency factor;
```

Cells (or Cell Groups)

	Observation window 1	Observation window 2	…	Observation window j
Cell_1	$C_{1,1}$	$C_{1,2}$	…	$C_{1,j}$
Cell_2	$C_{2,1}$	$C_{2,2}$	…	$C_{2,j}$
…	…	…	…	…
Cell_i	$C_{i,1}$	$C_{i,2}$	…	$C_{i,j}$

Table 2. Number of visits on cell i during observation window j

After the calculation, top k subscribers (k=1,000 by default) are selected as the measurement points.

As it is shown in the calculation, the algorithm is designed to identify those with high mobility and predictable mobility patterns as measurement points. The efficiency calculation method is the best-mode implementation.

STEP2: Determine a minimum set of frequent visitors for the monitored network, whose mobility (trajectory) covers all of the nodes.

For each measurement point m identified in STEP1, let $S_{m,j}$ be the set of cells the measurement point cover during its movement during an observation window j. This step is to determine a minimum set of measurement points whose movement covers all of the nodes in the monitored network.

This is a minimum set cover problem, one of the classical set covering problems in computer science:

Minimum Set Cover is known to be NP-Hard (Non-deterministic Polynomial-time). There is no fast algorithm to solve it. The problem to be solved is however with relaxed requirements in terms of number of subsets. A greedy approximation algorithm is proposed below as the best mode implementation and suffices for practical purposes.

The proposed algorithm generates the minimum cover set for each observation window. The identified sets over different observation windows shall be merged into one subscriber set as the final output.

```
Input:      cell lists of the subscribers identified as measurement points during each observation window

Output:     a minimum set of the identified subscribers satisfying monitoring requirements

For each observation window j

            REPEAT

                        Select a subscriber from the identified subscribers, satisfying the cell list of
the subscriber (during j) covers the maximum number of uncovered cells in the monitored network

                        Mark those cells in the monitored network as covered

            UNTIL all cells in the monitored network are marked as covered
```

It shall be noted that a redundancy factor may be taken into consideration in the minimum coverage algorithm above, i.e. each cell shall be covered at least n times. This can be simply done by adding a counter to each cell.

STEP3: Repeating STEP1 and STEP2 until the identified measurement points fully cover the monitored network, or satisfying the monitoring requirement (for example 95% of the network monitored)

It is possible that the identified measurement points in STEP1 can NOT cover all of the cells in the monitored network. In such situations it is necessary to go back to STEP1 to identify more subscribers as the measurement points (by lowering the efficiency factor threshold in identification).

B. Targeted Incentive Mechanism in Crowdsourcing

The step is to provide targeted incentives to the identified measurement points, to encourage subscribers to participate in the terminal based measurement and reporting program. The targeted incentive mechanism is based on analysis of service usage records.

Targeted Incentives

Before being classified as measurement point the subscriber has to give approval and the accepted way to gain approval is to give the subscriber an incentive.

In particular, service usage records of the identified measurement point are analyzed to determine the most frequent accessed services. Other data sources, such as billing information, may also be fed into the analysis to help determine the incentives. For example, the incentive may take account in the account type (private subscriber vs. company subscriber) and charging information to identify the most attractive factors for the measurement point. For example, a truck driver may have an account with free voice call service (i.e. his company paying all voice call charges) but have to pay for all SMS he sends; in such circumstances, SMS is the targeted incentive for this particular measurement point.

Delivery of Incentives

For example the targeted subscriber could be sent an SMS with an offer for free text messages during the time the terminal becomes a measurement point. This offer could be used for both post and prepaid subscriptions.

An offer like this could have the beneficial side effect of generating more traffic on one hand but also skew the user behavior and therefore the measurement accuracy after the subscriber has completed its obligation of his/hers terminal becoming a measurement point.

Refining Measurement Points

Accepted offers will turn a subscriber's terminal in a measurement point. A measurement point added to the overall schedule of measurement points and activated. Once activated the measurement point is subject to a measurement schedule involving both reporting of passive measurement and reporting of active measurements.

The identified measurement points, i.e. the outputs from the previous step, shall be updated based on the results of the targeted incentive step.

In case that the monitoring requirements are not satisfied by the measurement points, it is necessary to iteratively go back to the identification step and the incentive step until enough measurement points are confirmed.

C. Scheduling of Terminal Based Measurements and Reporting

This is to schedule terminal based measurements and reporting according to the coverage map (Table 1). The scheduler may simply extract scheduling information from the coverage map and propagate the schedule towards the terminal agents through configuration management service. The following paragraphs describe further improvements, with extra benefits in terms of less intrusion to terminal usage. The scheduling method is to determine the following for each measurement point:

- When to carry out measurements (& reporting). The scheduler may take the usage records of terminals and services as the input and analyze the usage patterns. Based on the usage patterns, the scheduler may remove the measurement activities overlapping with heavy service/terminal usages). This would significantly reduce the intrusion to terminal usage.
- Where to carry out measurements (& reporting). This may not be necessary, since mobility patterns have already been considered in the coverage map calculation.
- What to measure (& report). The coverage map provides a holistic view of the measurements. This is to use this to determine the measurement traffic profile for each subscriber, so that the measurement traffic is randomized. By randomizing the traffic profile, the accuracy of the measurements can be significantly improved. See also reference [3], [4] describing accuracy improvement when using random payload and random start time.

For this purpose, the coverage map is further enhanced with measurement traffic profile information. The following table shows an example (simplified without losing generality).

The measurement observation window may be reduced to provide better granularity of traffic randomness. In case passive measurements are not sufficient to cover all of the cells and/or observation windows, active probes are scheduled to cover those gaps. As a result, a measurement plan is established across measurement points for the same measurement target; randomness of measurement profiles is achieved.

	08:00-09:00	09:00-10:00	…
Cell_12300	S_1 (100bytes per packet, TCP), S_9 (200bytes per packet, TCP), S11 (500bytes per packet, TCP) …	S_2 (100bytes per packet, TCP), S_9 (200bytes per packet, TCP), S20 (500bytes per packet, TCP) …	…
Cell_Group_1	S_2 (100 bytes per packet, TCP), S_{11} (200 bytes per packet, TCP), S_{21} (500 bytes per packet, TCP)	S_3 (100 bytes per packet, TCP), S_4 (200 bytes per packet, TCP), S_{10} (500 bytes per packet, TCP)	…
…	…	…	…

Table 3 Scheduling of Terminal Based Measurements and Reporting: Enhanced Coverage Map

IV. Deployment

The described solution shall be data-centric and context-aware, agnostic in terms of the implementation of the terminal reporting. That is, the solution shall be applicable, no matter (1) What KPIs are measured; (2) what measurement methods are used; and (3) what reporting protocols are used in transmitting reports. It achieves such technology-agnostic by introducing the following interfaces; as long as existing terminal reporting solutions supports such interfaces, the proposed scheduling algorithms shall be applicable

- Measurement collection
- Scheduling

V. Conclusions

With the non-intrusive solutions proposed in this paper, complemented with privacy-conserving and subscriber-friendly features, terminal reporting can achieve a reasonable penetration ratio among subscriber in the near future.

Further work includes analysis of the performance of the greedy approximation algorithm.

References

[1] Ericsson White Paper 284 23-3150 Uen "Keeping the Customer Service Experience Promise", January 2011.

[2] Crowdsourcing Service-Level Network Event Monitoring, David R. Choffnes, Fabián E. Bustama, Zihui Ge, SIGCOMM'10, August 30–September 3, 2010, New Delhi, India

[3] Fabini, J.; Wallentin, L.; Reichl, P. The Importance of Being Really Random: Methodological Aspects of IP-Layer 2G and 3G Network Delay Assessment. ICC 2009.

[4] Advanced Stream and Sampling Framework for IPPM, draft-morton-ippm-2330-update-00. IETF draft