Performance Evaluation 90 (2015) 18–35

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

Load transients in pooled cellular core network nodes

Åke Arvidsson [a,b,*], Tobias Rydén [c]

[a] Ericsson AB, SE-164 80 Stockholm, Sweden
[b] Department of Computer Science, Kristianstad University, SE-291 88 Kristianstad, Sweden
[c] Department of Information Technology, Uppsala University, Box 337, SE-751 05 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 16 January 2014
Received in revised form 22 April 2015
Accepted 22 April 2015
Available online 4 May 2015

Keywords:
Iu-flex
S1-flex
Pool
Node failure
User redistribution

ABSTRACT

The coverage areas of cellular networks are logically subdivided into service areas. Each service area has a local anchor node which “hides” the mobility *inside* the area and the entire network has a global anchor node which “hides” the mobility *between* areas.

The concept of unique local anchor nodes per service area was invented to simplify routing but has been found to complicate expansion. The rapidly growing demand for cellular access has therefore prompted for alternative solutions with pools of local anchor nodes per service area. Such pools are now deployed by several operators all over the world.

Users in pooled service areas are mapped to specific pool members according to a load distribution policy, but the mapping can change as a result of node failures or operator interventions. Such changes take a certain time to implement and cause additional load on the anchor nodes. We study these processes in detail and derive closed form expressions which allow operators to control the trade-off between rapid changes and acceptable loads.

Finally we show that the key assumptions of our model are in agreement with measured data and demonstrate how the model can be applied to investigate the effects of different network settings (timers) under different user behaviour (traffic and mobility).

Contrary to current solutions to this problem, which typically are slow and/or inaccurate, our results enable fast and accurate analysis of different scenarios thereby enabling operators to maximise utilisation of the existing investments and at the same time avoid potentially dangerous situations of overload.

1. Introduction

Traffic in cellular networks is routed through *anchor nodes* which “hide” user mobility from the rest of the network. In more detail, there are “global” anchor nodes which interface external networks and “local” anchor nodes which interface dedicated subsets of cells. The set of cells allocated to a particular anchor node is known as the “service area” of that anchor node. The concept of one unique anchor node per service area is easy to implement but also imposes limits to the number of users and the volumes of traffic that can be handled in each service area. When these limits are reached the existing nodes can be upgraded or new nodes can be added, but both options have significant drawbacks. As for *upgrades*, the strong growth of subscriber numbers and traffic volumes in cellular networks means that the capacity of new nodes often fills up before they have been depreciated (*i.e.*, an upgrade is economically impossible) or before new versions have been developed (*i.e.*, an upgrade is technically impossible). As for *additions*, the requirement that service areas be unique means that service

* Corresponding author at: Ericsson AB, SE-164 80 Stockholm, Sweden.
E-mail addresses: Ake.Arvidsson@ericsson.com, Ake.Arvidsson@hkr.se (Å. Arvidsson), tobias.ryden@it.uu.se (T. Rydén).

http://dx.doi.org/10.1016/j.peva.2015.04.001
0166-5316/

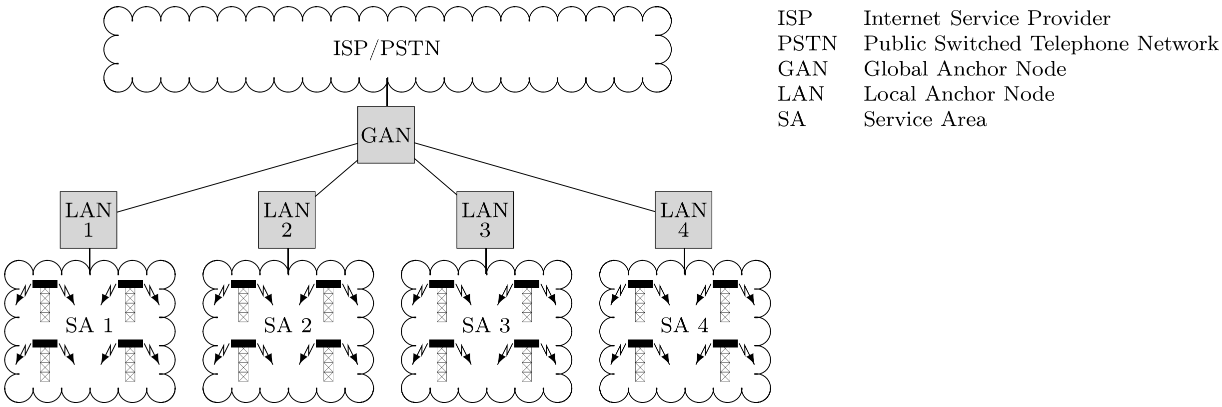

Fig. 1. Simplified view of a cellular network.

areas in which the limits have been reached must be split between the old node and the new one. As will be discussed more below, such splits are complicated to implement and result in reduced efficiency.

Given this background, the existing standards have been extended to allow for "pools" of more than one anchor node per service area. The load in a pooled service area is shared between the nodes by allocating different users to different nodes. Such allocations are in effect from the moment users enter the service area to the moment they leave it. The capacity in a pooled service area is thus the sum of the capacities of all pool members, hence the growing demand in a service area can be handled simply by adding more nodes to its pool.

Besides being a convenient way to handle growth, pools also offer the possibility to *redistribute* users between nodes in case of, *e.g.*, node failures (improved redundancy), planned outages or recent additions (simplified management). However, node changes require data updates in all affected entities (the old local anchor node, the new local anchor node, the global anchor node and the user terminal) hence such redistributions may result in noticeable load peaks. The subject of this paper is to develop a model whereby these peaks can be understood and controlled such that operators safely can take full advantage of pools.

To the best of our knowledge, we are not aware of any previous attempts to study this problem. On the contrary, despite the fact that several operators around the world already have implemented pools, cf., *e.g.*, [1–3], and that many others are planning to do so, the current practical approaches to these problems are insufficient (*e.g.*, steady state analyses which provide no information about progress rates or load transients) or impractical (*e.g.*, measurements or simulations which may be both complex and slow).

The rest of the paper is organised as follows: Section 2 summarises the technical background; Section 2.1 provides a brief introduction to cellular networks while Sections 2.2 and 2.3 describe the problems with unique nodes and the advantages with pooled nodes respectively after which we formulate our problem in Section 2.4. Next, the mathematical models of user redistribution are derived in Section 3 with some preliminaries in Section 3.1 and the models related to node failures and operator interventions in Sections 3.2 and 3.3 respectively. We then verify and apply our results in Section 4; in Sections 4.1 and 4.2 we introduce a measurement based traffic model and a realistic load model respectively, in Section 4.3 we show that the measurements support the key assumptions of our model and in Sections 4.4 and 4.5 we apply our results to illustrate the impact of different user habits and different network parameters under node failures and operator interventions respectively. Finally in Section 5 we summarise our work and discuss various extensions.

2. Technical background

2.1. Cellular networks

A simplified view of a cellular network according to the Third Generation Partnership Project (3GPP) is given in Fig. 1.

The network consists of radio cells with transmitters (towers), the transmitters are connected to local anchor nodes (LANs) and the LANs are connected to a global anchor node (GAN) which interfaces an internet service provider (ISP) or the public switched telephone network (PSTN).

Traffic *from* a user can be routed hierarchically through the current cell and its unique, topologically given LAN and to the topologically given GAN, but traffic *to* a user must be routed by location tables in the GAN (to find the current LAN) and in the LAN (to find the current cell). Users continuously associate themselves with the cell with the currently strongest radio signal and a procedure known as location update is used to update the location tables as required. We shall refer to such events as *re-registrations* and note that only some *cell changes* trigger *LAN changes* since the service areas (SAs) of LANs contain many cells.

The users in an SA are identified by local temporary identifiers (TIs) assigned by the LAN. Users without TIs (who, *e.g.*, power up for the first time) or with invalid TIs (who, *e.g.*, arrive or return to the SA and thus have TIs from other SAs) will be