UNITED STATES PATENT APPLICATION

For

PREDICTIVE MACHINE LEARNING FOR ENERGY OPTIMIZATION IN 6G NETWORKS

INVENTORS:

SAYANDEV MUKHERJEE

BERNARDO HUBERMAN

Predictive Machine Learning for Energy Optimization in 6G networks

August 30, 2023

Introduction and Background

As stated in [1], the upcoming 6G networks, now in the early stages of design, are supposed to offer 10X the capacity at 1/10th the latency for 10X as many users as existing 5G networks. Clearly, the obvious approach to realize the capacity and latency goals for such 6G networks is to continue the "cell densification" philosophy that has characterized cellular wireless network design for the last 40 years. In other words, the idea is to reuse the available wireless spectrum as intensively as possible, i.e., by shrinking cell sizes more and more.

While cell densification is a long-running trend in wireless systems design, there are two other, more recent, trends that are expected to revolutionize the design and deployment of future wireless networks: virtualization and AI/ML. Virtualization was initially conceived to allow for quick time-to-market deployment of new features in a future software-defined network, while also saving on equipment costs by using commoditized components. Over the last decade, as AI/ML began to show promise as a general-use technology in all aspects of network design, operations, maintenance, and business planning, it was also observed that virtualization offered a readymade platform for the deployment of such AI/ML models throughout the breadth and depth of a network.

It is therefore expected that 6G networks will be characterized by greater intelligence at the edge, in the form of AI/ML models deployed on virtualized access points that are interconnected through a high-speed backhaul.

Problem statement

Given the above context, it follows that upcoming 6G networks will be software-defined networks employing heavy use of virtualization to facilitate running AI/ML models over the network. One consequence of this is that access points are essentially going to be implemented as software running on CPUs and GPUs. This marks a break from today's access points, which are mostly deployed on custom ASICs. This in turn has an important consequence, to be discussed next.

As pointed out in [1], the greatly increased densification of access points envisaged in 6G networks will mean that with high probability, there will be parts of the network that, at certain times of the day, will have no users to be served at all by the access points in those regions of the network. If these access points are kept fully operational all the time, there will be wasteful energy consumption in the network, which not only increases operating costs for the network operator but may also run afoul of legislation.

The above energy consumption problem is compounded by the fact that CPUs and GPUs consume more energy than custom ASICs. Moreover, the traditional engineering approach of minimizing energy consumption by inactivating devices via putting them into a “sleep” state with periodic awakenings to sense the environment is increasingly untenable for devices like access points implemented on CPU/GPU hardware whose functionality is primarily software-defined. This is because such devices both consume more energy in their sleep state and take longer to wake up from such sleep than devices built on custom ASICs.

To summarize, the very short sleep time durations and very rapid sleep-followed-by-wake cycles that would be required in the traditional approach to reduce energy consumption are not compatible with the stated capacity and latency goals of 6G and cannot be supported by the software-defined network implementation of 6G networks.

Brief description of the idea

Our idea is to reduce both *energy consumption and interference* in future 6G networks by keeping inactive (i.e., not serving any users) access points ***switched off*** until a user device is predicted (by machine learning models running on the active parts of the rest of the network) to arrive in the region to be served by a cluster of currently inactive access points. This gives these inactive access points enough time to boot up and be ready to serve the user when it arrives.

In addition to the location prediction of the user, machine learning models will be used to predict the resources required by the user from the network to support a given quality of experience, and an economics-based allocation algorithm will determine the fraction of required resources (to support the user QoE) to be allocated to each access point in the cluster that serves the user.

All these elements of the proposed idea build on previously filed disclosures (details below).

Detailed description

The proposed complete solution has the following components:

1. *ML-aided prediction of the future location of a user device in the 6G network*: We propose to deploy a distributed ML model over the 6G network edge and backhaul (i.e., in access points and a controller unit, all connected via a high-speed backhaul) that will predict the location of a user device at a short time in the future. The algorithms used to do this prediction will be that already described in disclosure https://cablelabs.brightidea.com/D4526 titled “User Mobility Prediction for Position

Awareness in Intelligent Radio Access Networks." We repeat Fig. 1 from that disclosure below for convenience. Note that the proposed estimators in that disclosure, including LMS adaptive filtering, cubic spline interpolation, and polynomial regression, are examples of simple machine learning algorithms. More advanced ML algorithms may always be employed in place of these.

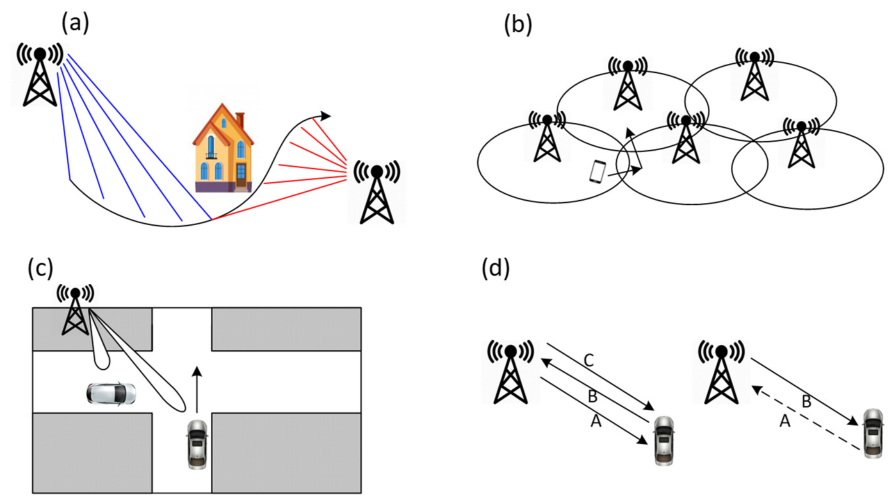

Figure 1. Application scenarios of mobility prediction in mobile networks: (a) positioning for millimeter wave beam sweeping; (b) small-cell hand over; (c) traffic management and self-driving cars; and (d) reductions of communication redundancy and overhead.

2. *Network-initiated boot-up of a cluster of access points near the predicted future user location to serve that user*: this is the action to be taken by the controller unit in the network upon getting the predicted user location. If the user is predicted to be present in a region where all the access points are currently inactive and therefore completely switched off to save energy, then a cluster of these access points will be simultaneously activated by a suitable control signal sent from the controller. We choose to activate a cluster of access points instead of just one because our predicted future user location may not be precisely correct, in which case a neighboring access point to the one nearest to the predicted user location may be the one that actually needs to serve the user. Moreover, it is possible to conceive of a cluster of access points simultaneously serving the user, which helps to minimize the possibility of session outage if the user is moving through the region of coverage of this cluster of access points.

   The above describes the core functionality of the proposed solution. However, given that the 6G network will offer a platform for the deployment of rich AI/ML services and functionality, we can enhance the above basic functionality through the following additional features.

3. *ML-aided prediction of the traffic load incurred by supporting the user at its new predicted location*: we propose to train and deploy an ML model that will predict the traffic load that the user will bring to its local network at its new, predicted, location. In fact, this traffic load will be a function of the predicted application(s) to be run by the user device at that location, and will in turn map into the wireless resources required by that user device in order to maintain the Quality of Experience (QoE) desired by that user, or specified in the Service Level Agreement (SLA) between that user and the network operator.

   Once again, we will employ the traffic/QoS prediction algorithm already described in disclosure https://cablelabs.brightidea.com/D4057 titled “A Machine Learning Algorithm for QoS Assurance of Network Traffic” for this purpose.

4. *Fast handover to and resource allocation by a cluster of newly-rebooted access points*: One aspect of 6G network design we have not touched upon so far is that the present slow and overhead-intensive method of implementing handover in 3GPP cellular systems cannot scale to the much denser and smaller-cell 6G network of the future. The disclosure https://cablelabs.brightidea.com/D4641 titled “Market-based Fast Handover with Predictive CoMP in a Small Cell Wireless Network” was originally conceived to provide a fast allocation of resources to a user device that had suddenly lost its connectivity to a serving base station, in order to maintain its session connectivity until the usual slow handover process concluded with an assignment of a new serving base station. However, we now propose this solution as the default way of supporting the user in the 6G network in its new predicted location.

   As in the original disclosure, the user will be served not by a single access point but by a cluster of neighboring access points. The resources required to support the user at the QoE level desired for the applications it is running (as predicted in the previous step) will be allocated by the access points in the serving cluster. The disclosure also specifies an economics-based allocation algorithm that ensures that no access point in the serving cluster is overloaded in having to serve the requirements of the incoming new user.

Concluding remarks

We propose an energy-saving scheme that is tailored to future highly-densified 6G networks with edge intelligence in the form of AI/ML services running on software-defined virtualized access points. The scheme essentially completely switches off inactive access points that are not currently serving any users, while relying on ML prediction algorithms to predict when users will travel to the region served by these inactive access points in order to give ample time for these access points to be booted up and ready to serve such users when they arrive. For additional efficiency, we also predict the resources required by these users to support their desired QoE, and propose a market-based mechanism for the users to acquire these resources from the access points of their serving clusters.

References

[1] Hewon Cho, Sudarshan Mukherjee, Dongsun Kim, Taegyun Noh, Jemin Lee, "Facing to wireless network densification in 6G: Challenges and opportunities," *ICT Express*, Vol. 9, No. 3, 2023, Pp. 517-524, https://doi.org/10.1016/j.icte.2022.10.001