Contents lists available at ScienceDirect
Vehicular Communications
www.elsevier.com/locate/vehcom

Efficient replication for vehicular content distribution

Da Zhang, Chai Kiat Yeo

Centre for Multimedia and Network Technology, School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore

ARTICLE INFO

Article history:
Received 10 October 2017
Received in revised form 19 February 2018
Accepted 13 April 2018
Available online xxxx

Keywords:
Content distribution
Vehicular networks
Data replication
Road Side Unit (RSU)

ABSTRACT

We consider a vehicular content replication system which makes use of the deployed Access Points (APs) to maximize the vehicular download progress of delay-tolerant contents through replication in the APs' local storage. The replication system for vehicular users is quite different from the traditional system for static Web users. The transient connection period between the vehicle and the AP makes it difficult for the vehicle to download the entire file requested and thus the content retrieval is usually across several APs. Such characteristic poses two problems: (1) replication in units of entire file may be inefficient in terms of resource utilization; (2) the actual contribution of an individual AP can be affected by the other correlated APs during the content retrieval. To deal with these challenges, we formulate the vehicular content retrieval into an offline optimization model that helps establish the performance bounds of replication algorithms in maximizing vehicular download progress. A real vehicular trace is also thoroughly analyzed. Then we propose an efficient and distributed replication algorithm explicitly taking into account the content popularity, vehicle-AP contact pattern and content availability among correlated APs. Simulation based on real vehicular trace proves the effectiveness of the proposed replication system. The performance in terms of download rate and completion ratio has at least 15% to 20% improvement against the algorithms under comparison.

1. Introduction

Under the envision of pervasive computing, various types of mobile devices ranging from laptops to smart phones are expected to access Internet from everywhere, even from vehicles in driving. Content distribution to vehicular users through wireless network access is emerging as a necessity to facilitate better road safety and enhance driving experience. The content distributed to vehicles can be emergency alerts or traffic condition broadcast, which is the basis of modern Intelligent Transportation Systems (ITS). Drivers who receive such content can prepare well in advance instead of being shocked when the vehicle drives too near to the incident. The accident rate and traffic fluency can thus be improved. In another scenario, the distributed content to vehicles can be local information like promotions of nearby gas stations, available parking lots, or new dishes that can be taken away from restaurant passed by later. Those information can be updated automatically to drivers who drive into the region of interest and have no need to pull over to search them.

We envision that, in the future, vehicles will steer in an environment full of connections not only with cellular networks but also with Road Side Units (RSU) that are short range and lightweight. While government can deploy dedicated RSUs at key areas, the ubiquitous infrastructure of WiFi-based Access Points (APs) can also be exploited for the purpose of vehicular content distribution. Such feasibility has been proved in several publications [1–3]. WiFi-based APs are characterized by short-range coverage (hundreds of meters), cheap and easy deployment, and high data access rate with latest gigabits data rate in IEEE 802.11ac [4]. WiFi-based APs can be a complementary and integral part of 4 G/5 G cellular networks, where the data retrieval from vehicular users are optimized adaptively based on data transmission rate and latency requirement. Mobile data offloading [5] is one promising integration of the proposed WiFi-based algorithm with existing 4 G or emerging 5 G cellular technology. Cellular network operators can deploy low-cost WiFi-based APs in dense area to offload delay insensitive data from cellular network, which could utilize the saved bandwidth to better serve low-latency traffic. Such delay insensitive data, like map/software updates, would usually not cause users' impatience and aversion. Cellular network operators can also incentivize the use of WiFi to encourage offloading. Some main operators already try such incentives like TMobile @Home.

However, network access through WiFi-based APs poses many unique challenges on the system design for effective content distribution to vehicles: (1) a single vehicle-AP contact duration is quite

E-mail addresses: sharetju@gmail.com (D. Zhang), asckyeo@ntu.edu.sg (C.K. Yeo).

https://doi.org/10.1016/j.vehcom.2018.04.004
2214-2096/

limited (typically tens of seconds) due to the fast vehicle speed and the AP's short coverage range, thus limiting the data transfer opportunities; (2) response latency of remote data-origin server on the Internet can waste the valuable contact duration, especially for the heavily loaded server and congested or long-delay route; (3) wireless bandwidth between APs and vehicles can be bottlenecked by APs' backhaul links to Internet, due to the heterogeneity of AP devices whose backhaul link capacity vary widely, e.g. low-cost devices that are easily deployed but with low rate backhaul.

To solve the above challenges in vehicular content distribution, one of the effective ways is to replicate the data from remote data-origin server into the local storage of APs. This approach makes the requested data immediately available at the first-hop connection of vehicles, and thus obviates the need to resort to the remote data-origin server. The performance gain will be significant since the path delay from the AP to the server is eliminated and the transmission bandwidth between the AP and the vehicle is no longer constrained by the AP's backhaul links. The basic idea here is similar to that of replication in Web proxy servers but the latter is mainly designed for static or low-mobility Web users and not for highly mobile vehicular users. In this paper, we aim to capture the unique characteristics of vehicular content replication, and design an efficient and scalable replication strategy. Our proposed replication system is distributed, and each AP replicates a set of content items into its local storage, based on the long-term historical statistics including the content request rate as well as the contact pattern of requesting vehicles.

The detailed contributions of this paper is as follows:

• Formulating the content-replication decision into a offline optimization problem, which provides the upper bound which any online algorithm can ever achieve for the system performance.

• Conducting thorough analysis on a real vehicular trace, and having some important findings that support our system design philosophy. These findings include: (1) non-exponential and diverging distribution of pairwise inter-contact time between the vehicle and the AP; (2) Independent inter-contact times for each vehicle-AP pair; (3) highly skewed contact correlation existing among APs.

• Modeling the inter-contact process of a vehicle-AP pair as a delayed renewal process, in order to calculate the expected data volume a vehicle can download from an AP during the maximal delay tolerance period.

• Recognizing that the contribution of an individual AP is affected by the storage status in correlated APs, and then proposing several algorithms to calculate content availability among the correlated APs.

• Proposing an efficient distributed replication algorithm, in which each AP decides the replication strategy locally without the help of any central controller. The distributed algorithm comprehensively takes into account serval important factors, like content popularity, delivery potential of AP, and content availability on the correlated APs.

In the rest of this paper, we first present an overview of the vehicular content replication system and the unique characteristics faced by such system in Section 2. In Section 3, the system is modeled into a centralized offline optimization problem. A thorough analysis on a real vehicular trace is conducted in Section 4. Based on the analysis results, Section 5 presents a distributed online algorithm, MaxRep. Section 6 proves the effectiveness of MaxRep through extensive simulation.

2. System overview

A typical architecture of WiFi-based vehicular content distribution system, as illustrated in Fig. 1, is made up of a network of connected Access Points (APs), which are geographically deployed near the roads, running the customized protocols for cooperation and also equipped with local storage. The APs can communicate with one another through backhaul links to the Internet or via high-speed LAN. The data-origin server stores the contents the vehicular users can access. A central controller is usually located in the service provider and operates like a system supervisor, monitoring users' demands and preferences as well as the data load and usage of each AP.

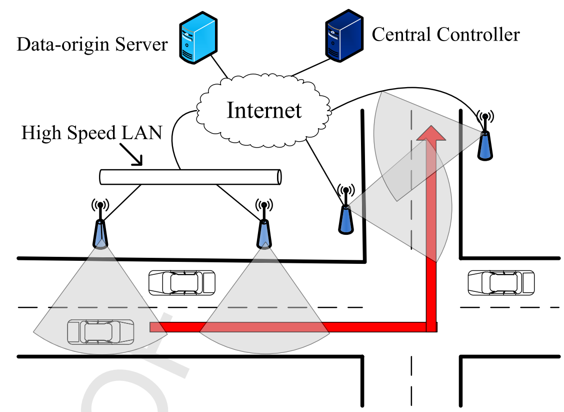

Fig. 1. Architecture of WiFi-based Content Replication System for Vehicular Users.

A vehicular user after requesting a file-download can choose to wait for a time interval, i.e. maximal delay tolerance, before starting the cellular transfer. During the maximal delay tolerance, vehicular users can collect the requested file through the opportunistically encountered roadside APs. After the delay interval, the uncompleted part, if any, would be delivered through the cellular network. To accelerate content retrieval, each AP can replicate a set of content items into its local storage, based on the vehicular demands and access patterns. The *content replication strategy* of an AP consists of determining which set of content items and how much of each item should be replicated in the storage as well as which set of items to be ejected when the storage is used up.

The **system objective** of this proposal is to determine the content replication strategy for each AP, under the constraint of finite capacity of AP storage, so as to maximize the expected *download progress* perceived by all the requesting vehicles within their respective maximal delay tolerance. Download progress here is measured by the percentage of a file that can be completed through AP-access by a requesting user before resorting to cellular download.

2.1. Data model

The system in our proposal encodes the contents by network-coding. Chou et al. [6] proposed a practical network coding method which bridges the gap between initial theory and realistic system. We follow their proposed concepts and methods. The data-origin server divides the original file into N ***generations*** and each generation consists of M packet(s). By network coding, the server generates an encoded packet for one generation by linearly combining the set of packets in the same generation with random coefficients c_j: $p'_{ij} = \sum_{j=1}^{N} c_j p_{ij}$, where p_{ij} is the jth packet in ith generation. For decoding purpose, encoding vector $[c_1, c_2, ...c_M]$ needs to be transmitted together.

The AP replicates a set of encoded packets based on the replication strategy, and then distributes them upon vehicular requests. After collecting M independent encoded packets of a generation together with its encoding vector, the vehicle can recover the original contents of that generation by solving a set of linear equations. The entire file can be recovered after collecting all the generations.

Please cite this article in press as: D. Zhang, C.K. Yeo, Efficient replication for vehicular content distribution, Veh. Commun. (2018), https://doi.org/10.1016/j.vehcom.2018.04.004