Vehicular Carriers for Big Data Transfers

Raul Adrian Gorcitz, Yesid Jarma, Prométhée Spathis, Marcelo Dias de Amorim, Ryuji Wakikawa, John Whitbeck, Vania Conan, Serge Fdida

To cite this version:

Raul Adrian Gorcitz, Yesid Jarma, Prométhée Spathis, Marcelo Dias de Amorim, Ryuji Wakikawa, et al.. Vehicular Carriers for Big Data Transfers. 2012 IEEE Vehicular Networking Conference (VNC), Nov 2012, Seoul, South Korea. 2012 IEEE Vehicular Networking Conference (VNC), pp.109-114, 2012, 10.1109/VNC.2012.6407418 . hal-00739361v2

HAL Id: hal-00739361

https://hal.archives-ouvertes.fr/hal-00739361v2

Submitted on 9 Oct 2012

Vehicular Carriers for Big Data Transfers (Poster)

Raul A. Gorcitz*, Yesid Jarma*, Prométhée Spathis*, Marcelo Dias de Amorim*,
Ryuji Wakikawa$^\diamond$, John Whitbeck*, Vania Conan‡, and Serge Fdida*

*UPMC $^\diamond$Toyota ITC ‡Thales

Abstract—**In the latest years, Internet traffic has increased at a significantly faster pace than its capacity, preventing efficient bulk data transfers such as data-center services and high-definition user-generated content applications. In this paper, we propose to take advantage of the existing worldwide road infrastructure as an offloading channel to help the legacy Internet assuage its burden. Our results suggest that piggybacking data on vehicles can easily lead to network capacity in the petabyte range.**

I. INTRODUCTION

Motivated by the need for technical flexibility and cost-effective scalability, large companies, organizations, universities, and governmental agencies constantly move their data and applications within and between data centers to balance workloads, handle replication, and consolidate resources. As a result, the demand for bandwidth-intensive services such as cloud computing, multimedia transfers, data migration, disaster recovery, and online backups has strained the Internet infrastructure to its limits.

Despite the ever-growing demand in bulk transfers, the price of bandwidth remains prohibitively high, especially at the network core. As a result, many edge providers are rate-limiting or even blocking the use of bandwidth-intensive applications. While bulk traffic can be seen as expensive when considering the high bandwidth consumption incurred, data-intensive applications are also less demanding when it comes to the requirements in terms of delay. Compared to most interactive applications that are highly delay-sensitive, the average throughput is the main criterion to evaluate the performance of bulk transfers and also to improve user experience.

Current methods for transferring such data include adaptations to standard file transfer methods or the use of hard drives and DVDs together with a courier service [1]. These solutions, although simple, can be either time consuming, or costly, or both. Recent alternatives propose to schedule data traffic to off-peak hours or using transit storage nodes [2], [3]. Nevertheless, as long as the legacy Internet is used, the underlying technology stays the same and the ISPs are still handling all the traffic. Cho and Gupta suggest to combine the Internet with the postal system to send a part of the data using hard-drives [4]. Nevertheless the bandwidth consumption requirements are on a steady rise and are mostly dictated by demands at peak hours.

We argue that disruptive solutions should be considered when it comes to moving huge amounts of data between geographically distributed sites. In this work, *we propose to exploit the delay-tolerant nature of bulk transfers to deliver data over the existing road infrastructures*. Our work is motivated by the increasing number of vehicles driven and miles traveled in the world. The vehicle fleet in operation worldwide surpassed the 1 billion mark in 2010 and is expected to double in the next two decades. The number of vehicles ownership is forecast to grow to up to 4 billion by mid-century. By leveraging the communication and storage capabilities that will soon equip (if not already) vehicles, *we advocate the use of conventional vehicles as the communication medium for big data migration in an opportunistic manner*.

Given the flow of vehicles daily traveling roads, our system design built on top of the road network can effectively *offload* the legacy Internet infrastructure for massive delay-tolerant data transfers.To evaluate our system, we compare our results with a state-of-the-art bulk data transfer scheme and we show that a vehicular-based solution may lead to significant improvements in terms of bandwidth.

In summary, the contributions of our work are as follows:

- **Offloading scheme.** We describe a novel vehicular-based opportunistic bulk data transfer system, designed to offload the Internet from delay-tolerant content.
- **Capacity improvement.** We show that the system has the potential of moving massive amounts of data in short time periods compared to today's traditional techniques.

The remainder of this work is structured as follows. In Section II, we describe the overall system operation and list the assumptions and shortcomings of this work. In Sections III, we evaluate the potential of our solution and compare it to an existing Internet-based technique. We give insights into some interesting open issues in Section IV and postpone the related work to Section V so that the reader has enough material to better understand the positioning of our work with regard to the literature. We finally conclude the paper in Section VI.

II. OFFLOADING ONTO CONVENTIONAL VEHICLES

In this body of work we argue that by taking advantage of the characteristics of future *smart* vehicles, such as data storage capabilities, these can be used to transport massive quantities of data between two geographical locations.

A. System operation

In order to overcome the limitations in terms of capacity and design of the Internet, vehicles are equipped with one or more removable memory storage devices such as magnetic disks or other non-volatile solid-state storage devices. The term "vehicle" refer to both passenger and commercial vehicles;

in the latter case, it may be part of a fleet vehicle owned or leased by a business or governmental agency. We assume that vehicles also embed one or more communication network interfaces and a positioning system. The system we describe below includes vehicles in operation and their users, a service provider, and content providers.

Memory devices can be owned by a party other than the user of the vehicle. Typically, a service provider may own the memory devices and owners of the vehicles can be compensated based on the amount of data transferred. A content provider distributes the data to be piggybacked onto the vehicles through a wide-area data network such as the Internet. The service provider charges the content provider for the amount of data to be transferred along the road infrastructure. A network of *offloading spots* provides the data to be piggybacked on the memory of vehicles. We use the term "offloading spots" to refer to locations that provide the data to be transferred to the memory devices of the vehicles or where on-board memory devices can be exchanged for pre-loaded memory devices that match the destination of the vehicle. The offloading spots can be placed at locations where vehicles may be parked. For example, an offloading spot can be located in a shopping center parking lot, a street parking spot, or at the users' home place. At a higher level, a collection of offloading spots form an entry/exit point to the system.

The service provider selects the offloading spots from the group consisting of the loading stations that transfer the data to the already in-place memory of the vehicle and the memory swap stations that replace the on-board memory of the vehicle. Vehicle memories can be loaded with data while parking at the offloading spot or exchanged for ready-to-ship memory devices so that users can continue their travels without waiting for the data to be loaded. The selection of the offloading spots is based in part on the geographic location of the vehicle and if available, its planned destination. The service provider also monitors the status of the offloading spots which include the available parking space, the memory exchange bays that are free, the destination of the data made already available for shipment. The service provider also periodically queries the vehicles over the data network to determine the current geographic location and destination of the vehicles. The positioning system of the vehicle includes a navigation system that generates routes and guidance between a geographic location and a destination. The historical locations and addresses are stored in a geographic location database managed at the service provider's control center. The service provider also keeps record of the status of the offloading spots in a specific database. The service provider matches the destination of the vehicles to a group of offloading spots selected based on park space availability. If preloaded memory devices are ready to be shipped, the service provider checks for free exchange bays at the offloading spots or contacts the content provider in order to transfer the data to be loaded on the memory devices.

TABLE I
SUMMARY OF THE VARIABLES USED THROUGHOUT THIS PAPER.

symbol	**meaning**
τ	transmission delay for the entire data
$\mathcal{D}$	total data to be transferred
f	vehicle frequency at a point
$\mathcal{S}$	storage capacity of a vehicle
$\mathcal{P}$	penetration ratio of the technology
d	travel distance
$\bar{s}$	average speed
s_f	free-flow speed
K	traffic jam vehicle density
k	actual vehicle density

B. Assumptions

Let us first denote the frequency of vehicles passing an entry point and traveling towards the same exit point (destination) as f. This value will be important to compute the maximum achievable capacity of the system. In practice, we assume that not all vehicles will have enough incentives to take part in the system. We call $\mathcal{P}$ the penetration ration of the technology, i.e., the probability that a vehicle accepts to carry data. We set this value to 20%, which corresponds to the approximate market share value of some of the major car producers in France [5],[6].[1]

We assume that the content provider takes all necessary security precautions to preserve the confidentiality of its own data. The service provider ensures that no physical harm can come to the data storage devices and that data integrity and confidentiality is preserved by limiting the access of the driver to the storage device. The entire set of data handling operations taking place at the offloading spots is automated, requiring no human intervention during normal functioning conditions.

We assume that all the vehicles that enter the highway at entry point A reach exit point B. Therefore, the value of $f \times \mathcal{P}$ for vehicles leaving A is identical for vehicles reaching B. This allows us to calculate the vehicular system performance unhindered by routing errors and data loss.

We denote the total amount of data to be transferred between A and B as $\mathcal{D}$. This amount is chunked and divided among the participating vehicles. For this computation, we denote as $\mathcal{S}$ the storage capacity of each vehicle. Finally, we call d the travel distance between A and B and $\bar{s}$ the average speed of the highway. The summary of the variables used in this paper is shown in Table I.

C. Shortcomings

This work is in its early stages, and several assumptions that might seem too strong will be removed in the future. One of them is the fact that, for the time being, we do not consider the impact of the data loading time on the total transfer delays. Data operations at the offloading point are part of a more complex architecture that will serve as a basis

---

[1] Note that this choice heavily impacts our results. However, we believe that this would be a lower bound, which corresponds to the case where a single car manufacturer accepts to embed this technology on their vehicles. Real values would be hopefully better than the numbers we show in this paper.

for further research. We assume a particular approach where data is loaded using batteries and swap stations [7]. In this approach, data is pre-loaded to batteries before cars enter the station. The loading time is than reduced to the time required to swap batteries (in the order of a minute) [8].

In this paper, we do not consider either the impact of the vehicles that exit the highway prematurely, leading to data loss. As our dataset does not allow tracking an individual vehicle, we will have to develop probabilistic models based on information from the traffic domain. Nevertheless, throughout the evaluation of our architecture, we use quite a conservative value regarding the flow of vehicles on the highway segments in hope of reducing the impact of data losses due to vehicles leaving the highway prematurely. We would like to stress again that the main purpose of this work is to present and demonstrate the potential of performing vehicular-based bulk data transfers in an opportunistic way and therefore we do not discuss data loss and routing related issues. We invite the reader to refer to Section IV for some discussion regarding this subject.

III. VEHICULAR CARRIERS: PERFORMANCE ANALYSIS

We evaluate the performance of bulk data transfers using vehicular carriers as described in the previous section.

A. Dataset

Our results are based on a vehicular frequency dataset spanning a two-year period, made publicly available at the end of 2011 by the French Ministry of Ecology, Sustainable Development, and Energy [9]. In this dataset, vehicular frequency measurements were conducted on multiple segments of the highway, yielding slightly different results at each measurement (as vehicles enter and exit the highway segment). Since we consider that all vehicles starting a journey from point A will reach point B, in order to filter out vehicles exiting the road in between we use the minimum vehicle flow value measured on the trajectory. The results of the considered scenarios are presented later in this section.

The data is provided under the AADT (Annual Average Daily Traffic) engineering standard used for vehicle traffic load on a given section of road. This type of standard is used for transportation planning/engineering and traffic related pollution [10], [11]. The AADT is an average measured for a period of one year and divided by the number of days. This type of data is useful in avoiding traffic differences depending on season or time of the day.

B. Computing transfer delays

The time required to transport data along the highway between two locations can be calculated as a sum of two factors. The first one is the time required for a vehicle to transit the stretch of highway between the starting point and its destination. It is calculated from the average speed of vehicles and the distance to the destination. The second one is the amount of time required for enough vehicles to pass through the offloading point towards the destination. The parameters of this second factor are the storage capacity of each vehicle, the penetration ratio of the vehicular bulk data transfer technology, the frequency of vehicles, and the total amount of data to be transferred.

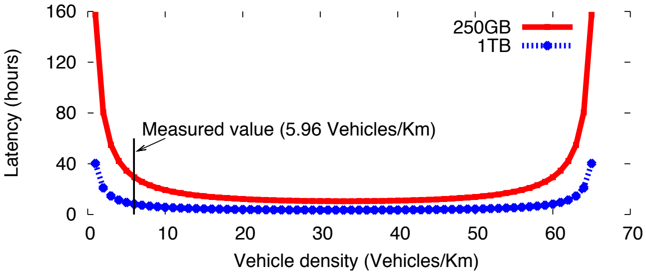

Fig. 1. Total delay required to transfer 1 Petabyte as a function of the highway density. The vertical line represents the measured value obtained from the dataset, for a 118 Km stretch of highway between Orleans and Tours.

The vehicle frequency measurement is one of the core elements required for capacity calculations in the highway research area [12], [13]. Highways are designed to avoid traffic jams and facilitate the free flow of traffic by estimating the vehicle frequency in a stretch of highway and the maximum vehicular density (i.e., the number of vehicles per unit of distance). Since for our proposal traffic jams could severely hinder the performance of the system, our calculations are based on the vehicle density. The vehicular frequency f at one point on the highway can be expressed in terms of the vehicular density as follows:

$$f = s_f \times \left(k - \frac{k^2}{K}\right), \tag{1}$$

where s_f is the free-flow speed (i.e., maximum speed), K is the density of vehicles characterizing a traffic jam, and k is the actual density measured in vehicles per kilometer (a traffic jam takes place when the value of k approaches K). Therefore, the total transfer latency τ can be expressed as a function of the vehicular density as follows:

$$\tau = \frac{\mathcal{D} \times K}{s_f \times ((k \times K) - k^2) \times \mathcal{S} \times \mathcal{P}} + \frac{d}{s}. \tag{2}$$

In Fig. 1, we present the transfer latency for transporting 1 PB of data in function of the density of vehicles using Equation 2. We consider vehicles with storage capacities of 250 and 1,000 Gigabytes, with a class A level of service on a rural highway with 3 lanes [12].[2] On the one hand, long delays are obtained with low traffic densities as less vehicles are available to act as carriers. As density increases, we observe a significant increase in performance until we reach the optimal density. On the other hand, as vehicle density grows beyond the optimal value, congestion levels rise and the speed of the

---

[2] The “class A” level of service refers to very good driving conditions where the drivers are unhindered by other traffic participants and are able to maintain the desired speed. Vehicle density values are the main performance metric used for the level of service estimation.

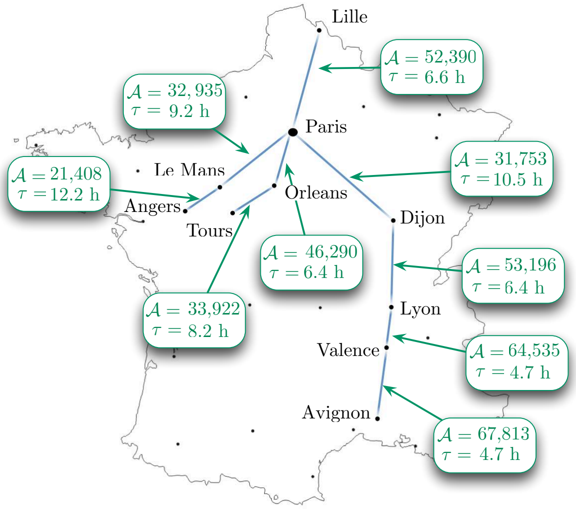

Fig. 2. Average Transfer Delays (τ) obtained using the "annual average daily traffic" ($\mathcal{A}$) on segments of highways connecting several important locations in France. The parameters used here are: average speed = 100 Km/h, total data = 1 PB, per-vehicle storage capacity = 1 TB, and penetration ratio = 20%.

vehicle flow slowly decreases until it reaches a jam zone that causes a steep increase of the latency parameter.

In Fig. 1, the vertical line indicates the density obtained from the dataset (5.96 vehicles/Km), which is the reference value we will use in the following. Note that, in this case, even though the measured flow is largely below the optimal capacity of the highway (approximatively five times lower), the transfer latency values obtained are as low as 8 hours to transfer 1 PB of data on a 118 Km stretch of highway, and have the potential of being lower as the vehicle density approaches the design optimum.

In order to provide clearer insights into the potential of this system, we have extended these calculations to different regions of France. The chosen set of highways was preferred as it links a large variety of points of interest with different characteristics, allowing the computation of a wider range of results. In Fig. 2 we show these results by mentioning the measured Annual Average Daily Traffic values and the calculated Average Transfer Delay obtained on the corresponding segment of highway for transporting 1 PB of data. The calculations were made considering an average vehicle speed of 100 Km/h and a per vehicle data storage capacity of 1 TB.

The range of values obtained for all highways shown in Fig. 2 varies from 4.7 to 12.2 hours, indicating transfer capacities of several orders of magnitude higher compared to the Internet.

C. Comparison with NetStitcher

In this section we compare the vehicular-based bulk data transfer system with an Internet-based bulk data transfer solution named NetStitcher [3] (see also Section V). NetStitcher achieves its best performance by transferring 1.15 Terabytes in 3 hours between cities in the same time-zone. These values will be used as a hard constraint during our comparison scenario. Here we are asking the following question: *is our vehicular-based bulk data transfer system able to transfer a comparable amount of data in the same period?* To answer this question, we developed a scenario that uses the segment of highway between two cities in the same time-zone, namely Orleans and Tours. We have varied the data storage capacity of the vehicles between 250 GB and 1 TB, obtaining two distinct performance values for the amount of data transferable in a 3-hour time interval. The results are presented in Fig. 3. The vehicular-based system outperforms by far the Internet-based bulk data transfer system by transferring up to 200 times more data in the same amount of time.

Fig. 3. Comparison between an Internet-based solution (NetStitcher [3]) and the proposed vehicular-based solution (V- 250GB and V-1TB), depicting the amount of data transferable in a delay of 3 hours between two cities in the same timezone.

IV. DISCUSSION

The goal of this section is to point out several research topics that could contribute to the development of the internet-based bulk data transfer system.

Routing and data delivery. One of the most pressing issues that requires attention is data delivery over longer distances. In the case of a highway linking three consecutive cities, it is not reasonable to make the assumption that all vehicles leaving the first city will reach the third city. This motivates the development of specialized routing protocols that limit data loss and ensure efficiency when drivers stop at intermediary points. The limited battery capacity would force the driver to stop for example at recharge stations or at *battery swap stations*. Such "points of interest" could act as routers where data could be swapped from one vehicle to another, heading in the right direction.

Scheduling and transfer planning. Vehicular traffic has observable diurnal patterns with dramatic increases in vehicular frequency during rush hours contrasting with much smaller values during night time when most of the traffic is represented by commercial freighters. This pattern of movement is dictated by constraints that characterize the driver's behavior (like working hours or driving preference) and it can be anticipated to a certain extent. In order to use efficiently the opportunistic vehicular-based system, data transfers must be properly scheduled and thus appropriate methods should be developed.

Data loading. Today's technology is advancing and is offering new alternatives to loading data on to vehicles. If battery swap stations are deployed at a large scale, vehicles could be "fed" with data when they get a newly charged battery, and this in a very short period of time (about one minute [14]). Another option would be to use state-of-the-art microchips that allow wireless transmission speeds up to 1,000 times faster than current technology [15].

Incentives and business plans. Vehicles participating in the system would be private property and as such, they cannot be used to transfer data without the consent of the owner. A mutual beneficial business plan needs to be developed to motivate car owners to participate, while a partnership between multiple companies could potentially reduce the servicing costs of such a system.

V. RELATED WORK

With the cloud technology gaining ground, data centers are increasing in size demanding larger aggregate bandwidth requirements. To improve the services offered to users, data needs to be moved closer to the consumer along the provision based infrastructure. It can also require data restoration functions in a disaster scenario, keeping services running while problems are being solved or a distribution scheme that requires multiple geographic locations to work efficiently. In [16] the authors find that provisioning high levels of bandwidth with today's existing techniques has a determinant impact on the development and maintenance budget of a company.

Though bulk data transfers are at the basis of some popular services such as high-definition multimedia content delivery, many ISPs are using scheduling, traffic shaping, and queue management techniques to limit the rate of bandwidth-intensive applications. An example of service affected by the ISPs policies is Netflix, who is responsible for about 29% of North American fixed internet access bandwidth utilization [17]. Initially launched as a DVD rent-by-mail company, Netflix began streaming movies online in 2007 in an attempt to avoid the postal costs for delivering DVDs by mail. To address the ISPs rate-limiting policies and the costs of serving content, recommendation algorithms are expected to be used in combination with peer-to-peer networking by big data service providers.

To address the shortcomings faced by bulk data applications, Laoutaris et al. proposed NetStitcher, a proposal that exploits diurnal patterns of network traffic to schedule bulk transfers at times of low link utilization depending on the time zones [3]. Although this technique allows a good return on investments for dedicated lines, the physical medium is still limited due to the high cost and limited capacity of today Internet's infrastructure. Furthermore, efficient scheduling decisions require real-time information regarding the network load and a transport layer needs to be designed so as to be able to send all delay tolerant traffic when spare bandwidth is made available.

A work that is is similar to ours in essence is the idea of exploiting airplane passengers boarding airline flights to bridge remote airports [18]. Messages to be delivered are loaded onto the passengers' mobile devices at the airport depending on their destination while they are waiting for their flight. Unlike to our work, airplanes follow regular prescribed schedules. Simulation results show that under certain conditions, carrying data over scheduled flights can achieve a similar throughput as a single TCP connection as long as the amount of data to be transferred is equal to capacity of three DVDs.

Another work suggested the combined use of the Internet together with the postal system to send a part of the data using hard-drives [4]. Even though this system offloads some of the data onto a carrier other than the Internet, it still relies on a methodology that requires detailed scheduling, and can only be used for data able to tolerate delays of up to several days. Furthermore data on physical media such as tapes or removable disk drives have the downside of requiring manual handling at arrival.

VI. CONCLUSION

In this paper, we proposed a novel technique for big data transfers by offloading data from the Internet onto a network composed of conventional vehicles in an opportunistic fashion. After evaluating and comparing our solution to another existing Internet-based technique, we have shown that using vehicles as data carriers has a huge practical potential in today's context, where the amount of data to be transferred is increasing at a fast pace, and current technologies are becoming obsolete. We are confident that a more complete evaluation of the system considering the load time will confirm the advantages of our proposal over classical Internet-based solutions.

REFERENCES

[1] W. Allcock, J. Bester, J. Bresnahan, S. Meder, P. Plaszczak, and S. Tuecke, "Gridftp: Protocol extensions to ftp for the grid," http://www.ggf.org/documents/GFD.20.pdf, 2003.

[2] N. Laoutaris, G. Smaragdakis, P. Rodriguez, and R. Sundaram, "Delay tolerant bulk data transfers on the internet," in *ACM Sigmetrics*, Seattle, WA, USA, Jun. 2009.

[3] N. Laoutaris, M. Sirivianos, X. Yang, and P. Rodriguez, "Inter-datacenter bulk transfers with netstitcher," in *ACM Sigcomm*, Toronto, Ontario, Canada, Aug. 2011.

[4] B. Cho and I. Gupta, "Budget-constrained bulk data transfer via internet and shipping networks," in *International Conference on Autonomic Computing*, Karlsruhe, Germany, Jun. 2011.

[5] "PSA-Peugeot-Citroen, Market Share 2012," http://www.psa-peugeot-citroen.com/en/media/psa-peugeot-citroen-to-produce-in-france-its-future-light-commercial-vehicle.

[6] "Renault Group, Market Share 2012," http://media.renault.com/global/en-gb/renaultgroup/Media/PressRelease.aspx?mediaid=33009.

[7] S. Fdida, P. Spathis, and M. D. de Amorim, "Massive data migration on wheels," in *IFIP/IEEE WONS*, Bardonecchia, Italy, Jan. 2011.

[8] Better place. [Online]. Available: http://www.betterplace.com

[9] "Bison Futé," http://www.bison-fute.equipement.gouv.fr/.

[10] *Traffic Monitoring Guide*. U.S. Department of Transportation, Federal Highway Administration, 2001.

[11] T. Wright, P. S. Hu, J. Young, and A. Lu, "Variability in traffic monitoring data," http://wwwcf.fhwa.dot.gov/ohim/flawash.pdf, U.S. Department of Transportation, Federal Highway Administration, Tech. Rep., Aug. 1997.

[12] *Highway Capacity Manual 2010*. U.S. Transportation Research Board, 2010.

[13] F. L. Mannering and S. S. Washburn, *Principles of Highway Engineering and Traffic Analysis*. Wiley, Third Edition, 2004.

[14] "Better Place," http://france.betterplace.com/.

[15] "The virtus chipset," http://www3.ntu.edu.sg/CorpComms2/Documents/2012/May/Fact%20Sheet_120524_Virtus%20Microchip.pdf.

[16] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," in *ACM Sigcomm*, Seattle, Washington, USA, Aug. 2008.

[17] "Sandvine," http://www.sandvine.com/.

[18] A. Keränen and J. Ott, "Dtn over aerial carriers," in *Workshop on Challenged Networks*, Beijing, China, Sep. 2009.