

# Big Data Offloading using Smart Public Vehicles with Software Defined Connectivity

Rashmi Munjal, William Liu, Xue Jun Li, Jairo Gutierrez  
School of Engineering, Computer, Mathematical and Information Sciences  
Auckland University of Technology, New Zealand  
{rashmi.munjal, william.liu, xuejun.li, jairo.gutierrez}@aut.ac.nz

**Abstract**—With the explosive increase in the number of mobile devices such as smartphones or laptops, the design of mobile applications becomes increasingly complex, power hungry and resource consuming. Therefore, conventional networks are facing serious problems such as traffic overload and energy consumption due to high traffic demands. As a result, network designers are looking for more options to accommodate numerous data requirements. Aiming to find a promising way to tackle this problem, we are investigating heterogeneous networking architectures, which utilize the existing public transport network as an alternative communication network along with infrastructure-based networks. We propose a heterogeneous network architecture called Software Defined Connectivity (SDC) that utilizes the flow of transport network such as buses, trains, and ferries to start the forwarding process from nearby parking/offloading spots to disseminate data along with conventional networks. Results show that the SDC architecture helps in data offloading over public transport vehicles as per the profiles of each user with significant savings of energy.

Keywords: Delay Tolerant Network, Smart Public Transport Vehicles, Big Data, Network Selection

## I. INTRODUCTION

While looking back at the past years, it's crystal clear that our daily lives are getting equipped with modern devices such as smartphones, smart TVs and smartwatches. These rapid developments are making our life easier and smart. Vehicular networks [1] are visualized to satisfy the upcoming demands for improved transportation efficiency and potentially exert influence on others. Moreover, researchers have validated that vehicular networks can share sensory data among vehicles and infrastructure, thus remarkably improving the overall sensing performance.

The third-generation (3G), fourth generation (4G), and fifth generation (5G) technologies are growing fast with many more advancement to fulfill user's streaming requirements and real-time data traffic distribution [2]. These applications are experiencing explosive growth, which is the main cause of burgeoning mobile traffic. Researchers attempt to mitigate traffic flow burden by finding out alternative data transmission network architectures. Several data offloading techniques have been introduced using cellular base stations, T2T approaches, WI-FI hotspots and vehicular networks, respectively.

Vehicular DTNs extend VANETs with DTN capabilities to support long disruptions in network connectivity. It is one subtype of DTN, where nodes (vehicles) store and carry network data while waiting for opportunities to forward it. Komnios et al [3] introduced a CARPOOL+ routing technique of public transport by providing hotspot internet connectivity to all users and prior knowledge of their contacts. Similar work has been presented in [4], the author considered data center as a central point to pick up and deliver significant amount of data using public transport. In their next work, author [5] also proposed urban transport facilities and road infrastructure to complement traditional option for data transmission. UMass DieselNet [6] leverages opportunistic buses as a source and relay node contact to provide end-to-end connectivity. Their projects collected some datasets to characterize features of DTN. Benjamin et al [7] proposed a centralized architecture to offload traffic using a vehicle flow allocation algorithm taking directions of vehicles as well as performance requirements for reliable data transfer. A SDN-based wireless communication solution also helped in minimizing communication costs utilizing schedule of different network resources. Marincic and Foster [8] developed similar work to transfer data by using internet and transport facilities to find out more energy efficient solutions to transfer data in terms of bits as compared to atoms.

On the other hand, rapid population growth of these smart things has led to tremendous increases in energy consumption in ICT worldwide. Thus, it is a major concern for the next-generation network to improve its energy efficiency (EE). Global mobile data traffic [9] is expected to increase sevenfold within the next four years. It has been estimated that by 2021, 63 percent of global mobile data traffic (e.g., cellular data) will be offloaded to Wi-Fi or small cells networks. In the next few years, the data consumption will exceed the network improvements made by high capacity fibre providers to speed up existing network infrastructure and servers. In our previous work [10], we introduced an alternative option for data dissemination in addition to the conventional traditional (wired and wireless) networks. Next, we proposed a Software Defined Connectivity (SDC) [11], approach for network allocation as per the user's profile. We utilized all of the moving vehicles and their fixed schedule for data transmission

along with energy savings. The SDC Controller is in charge of gathering all the data from every user all over the city, analyzing their requirements, and selecting the appropriate network to carry the data by either using the internet or using scheduled vehicles, obtaining a favorable trade-off between the delivery probability and energy savings. In this work, we are doing analytical studies which shows how much data can be carried by public vehicles on predefined routes.

Our proposed work aims to do real world assesment of Auckland Public transport and gives detail explanation of alternative communication channel through SDC to save energy consumption. To manage the explosion of data traffic, we focus on metropolitan environments with an ultimate goal to extend delay-tolerant data delivery to under-served sectors of society while consuming less energy. We complement these smart vehicles (storage-enabled) with parking spots to carry data for a particular destination. The SDC controller controls all activities such as user profile, data type, data size, and offloading and downloading spots. In summary, the contributions of our work are as follows:

- **Software Defined Connectivity:** We propose a centralized Software Defined Connectivity procedure that enables scalable and adaptive control of the public transport’s smart vehicles to offload traffic. Our main aim is to utilize the existing public transport network for data dissemination.
- **Optimal Solution in the heterogeneous network:** We will be using the minimum-cost multi-commodity optimization algorithm to identify the optimal solution in terms of energy efficiency. Our objective is to minimize the energy consumption subject to two commodities: traditional and vehicular networks.

The remainder of this paper is organized as follows: In Section II, we will introduce an alternative communication channel and represent it with Software-Defined Connectivity. Section III proposes our system model for energy consumption. Section IV discusses the numerical studies as per Auckland public transport system and energy consumption analysis. Finally, Section V will conclude the paper.

## II. AN ALTERNATIVE COMMUNICATION CHANNEL THROUGH SDC

The traditional communication networks are becoming increasingly overloaded, while the cost of maintaining and expanding the infrastructure to accommodate big data communication remains high. Next we present, Software Defined Connectivity (SDC) approach as an alternative communication channel as shown in Fig. 1. SDC is a paradigm, where a central software program, called a SDC controller takes offloading demands from users as an input in its local storage and stores data there for a certain time to offload onto a vehicle. This approach consists of a SDC controller, roadside unit centers (RSUCs), roadside units (RSUs), BSs, vehicles, and users. Moreover, it also defines connectivity options such as infrastructure-to-infrastructure (I2I) links and vehicle to infrastructure (V2I) links.

- **SDC Controller:** The controller translates the network policy to actual packet forwarding rules. It makes all traffic forwarding decisions based upon the user’s profile. It has many connectivity options to consider for communication. It takes demands of all users to store in flow table and matches the profile with different parameters such as parking spot, data volume, data type, route and destination offloading spot. Then, it follows the entry of all vehicles at the parking spot to arrange the pick up of the data. The scheduled movement of all vehicles with installed storage capacity is stored in a geographic database managed by the controller which then will have knowledge about the fixed route until destination.

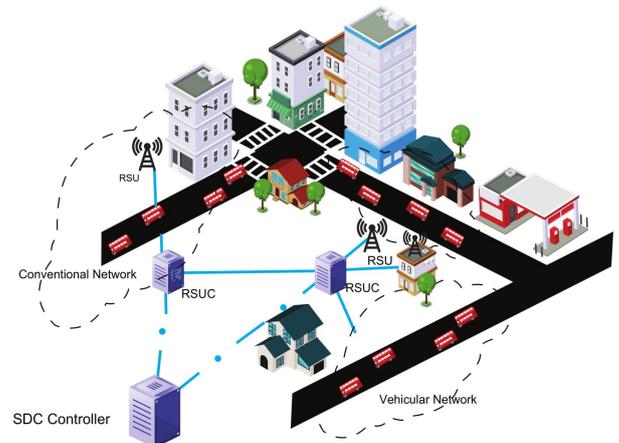


Fig. 1: Software Defined Connectivity

- **Data Collection:** Data are collected from the parking spot to analyse its type. Every Offloading spot/parking area is equipped with local storage room or repository, where data is stored until it is transferred to a parked vehicle or until the internet drop-off or pick-up decision is taken by the SDC controller as per the flow tables’ entry. As shown in Fig. 2, the controller takes the demands and objectives of each user and stores them in local storage, it analyses its data type and accordingly matches the direction of the passing vehicle against the destination offloading spot. However, if the data type is non-delay tolerant, the controller selects an internet option for the successful delivery of messages.

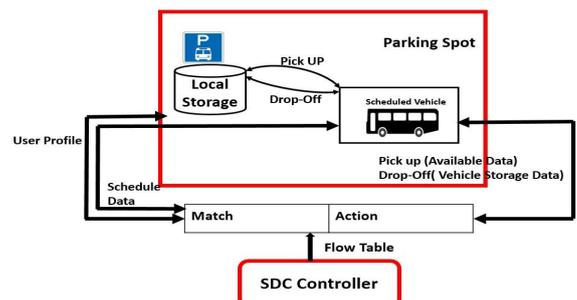


Fig. 2: Data collection at parking spot from every user

Flow ID	Data Size(GB)	Delay Tolerant Indicator(hrs)	Distance(km)
249	500	2	5
276	1000	10	10
676	2000	7	7
267	4000	9	8
432	8000	10	15
254	16000	14	13

TABLE I: Flow Table to match and action

- **Analyse Flow Table:** The flow table helps the controller in making the forwarding decision as per the profile of each user. The flow table consists of a list of parameter such as the data type, data size, and delay tolerant indicator (DTI) and distance travelled. They match the data volume and data type as per the possibility of the scheduled vehicle to move into the direction of a destination spot. The flow table consists of the flow id number with all other details needed to take action. The controller regularly updates the flow table with new demands and information about completed transfers.
- **Network Selection Forwarding Process:** The forwarding process is the network selection procedure between both networks which depends upon the urgency of data. After matching each and every entry of the flow table, the controller selects a network and forwards the data. This process first determines the local storage for data volume to upload onto the parked vehicle as per the flow table entry and direction. If none of the entries matches, data gets uploaded onto the internet, if it matches then it looks for the data type and confirms it to upload onto the smart vehicle as per its capacity. The same forwarding procedure is followed for each incoming vehicle and each offloading spot.

### III. SYSTEM MODEL

#### A. Energy Model (Traditional and Public Transport Network)

In this section, we present a case study to provide a proof-of-concept on the utilization of scheduled vehicles and traditional communication networks for data delivery. We want to send data from Auckland to Henderson considering both options. This case study aims to validate that utilizing smart vehicles as an energy efficient option for data transmission can have significant savings on energy consumption when the data to be delivered has delay tolerant features.

- 1) **Data Dissemination through traditional communication channels:** The controller follows the forwarding process and chooses this option if data is non-delay tolerant. Data gets uploaded onto the conventional network and all packets get forwarded through regular network equipment such as LAN switches, Core and Edge routers. Energy consumption can be calculated from Eq 4, which is the sum of energy consumption by nodes and incremental energy used. Uploading band-

width is 0.1 Mbits/s and downloading bandwidth is 1 Mbits/s for each node in sending data through internet.

$$E_{Nodes} = \max\left(\frac{n}{B_{up1}}, \frac{n}{B_{down2}}(\Delta P_{Node1} + \Delta P_{Node2})\right) \quad (1)$$

$$E_{bit} = \frac{P_{max}}{B} \quad (2)$$

$$E_{inc} = n * E_{bit} \quad (3)$$

$$C_E = E_{Nodes} + \sum_{i=1}^k E_{inc1} \quad (4)$$

- 2) **Data offloaded onto Smart Public Vehicles:** This option consider data communication to offload data onto scheduled public vehicles as per their nature such as delay tolerant or non-real-time.

$$E_{Nodes} = \max\left(\frac{n}{b_{up1}}, \frac{n}{b_{down2}}(\Delta p_{Node1} + \Delta p_{Node2})\right) \quad (5)$$

$$E_{travel} = \alpha_{fuel} \times W_{package} \times \sum_{i=1}^k E_{loading} \quad (6)$$

$$E_{loading} = \frac{d_i}{f_{ueleconomy_i} \times W_{vehicleload_i}} \quad (7)$$

$$C_E = E_{Nodes} + E_{travel} \quad (8)$$

TABLE II: Data forwarded through both networks

Auckland CBD → Henderson	Total Energy (Internet) 2.267857142(MJ)	Total Energy (Smart Public Transport) 0.161 (MJ)
L3 switches	6	N/A
Edge Routers	2	N/A
Core routers	5	N/A
$\Delta P_{Node}$	20 W	N/A
$B_{up}$	1 Mb/s	24 Mb/s
$B_{down}$	0.1 Mb/s	24 Mb/s
Distance	17km	17 km
Weight	N/A	0.95(2TB)
$\alpha_{diesel}$	N/A	38,290,237.52 J/L

Table II gives a detailed description about energy calculations from both networks in sending data Auckland to Henderson using Public Transport. It proves that the energy consumed by vehicles is much less in comparison with a traditional network.

---

**Algorithm 1:** Network Selection Procedure
 

---

**Input:** Graphs  $G^I = (N^I, E^I)$ ,  $G^V = (N^V, E^V)$  and set of requirements  $R = \{(b_i, C_{i,j}, U_{i,j})\}$

**Output:** Network Selection with min energy consumption

```

1  $C_E^I \leftarrow f(G^I, R)$ 
2  $C_E^V \leftarrow transshipment(G^V, R)$ 
3 for  $r_i \in R$  do
4   if  $C_E^I > C_E^V$  then
5      $mode_i = \text{SDC}$ 
6   else
7      $mode_i = \text{Conventional Network}$ 
8   end
9 end
10 return  $mode$ 

```

---

### B. Energy Consumption Optimization in heterogeneous network

The key idea of this paper is to refer to the concept of choosing the best optimal solution to traverse data over a network while considering requirements of each user as per their profile. Algorithm 1 outlines the network selection procedure among both networks with minimum energy consumption.

We use a network flow model to formulate our problem to find more sustainable approach in the heterogeneous network. Given that a network  $G = (N, E)$ , where  $N$  = set of nodes =  $(i, j, k, l, m, n)$ ,  $E$  = set of links =  $(i, j), (i, k), (l, m)$  nodes and a collection of  $E$  edges, where each edge is a pair of nodes from  $N$ . In the context of our problem, we consider a set of two commodities  $d \in D$  such that  $d = I, V$ . We can define a primary network (internet network) flow as  $G^I = (N^I, E^I)$  and secondary network as  $G^V = (N^V, E^V)$  for a vehicular network. Thus, given in table below:

TABLE III: Symbols and Notation Used

Term	Description
$G = (N, E)$ ,	Graph modeled for a network
$N = (i, j, k, l, m, n)$	Set of nodes
$E = (i, j), (i, k), (l, m)$	Set of links
$G^I = (N^I, E^I)$	A subset graph of Internet Network
$G^V = (N^V, E^V)$	Vehicular Network
$(d = I, V) d \in D$ .	Set of two commodities
$(b^I + b^V) = b^i$	$b^I$ = demand for internet network $b^V$ = demand for vehicular network.
$F_{ijd}$	Amount of flow produced between arc $(i, j)$ node
$C_{ijd}$	Energy Cost for both commodity
$f_{i,j}$	Fixed charge for the use of each arc $(i, j)$
$U_{i,j}$	Capacity of internet network
$T$	max time to get delivered the data
$t_{i,j}$	total delayed time in the vehicular network
$y_{ij}$	Network Selection $y_{ij} = 1$ if Vehicular Network $y_{ij} = 0$ if internet network

**Objective:** To minimize

$$\sum_{i,j \in N} \sum_{d \in D} C_{ijd} F_{ijd} + \sum_{i,j \in N} f_{ij} y_{ij} \quad (9)$$

**Constraint:** Subject to

$$\sum_{i,j \in N} F_{ijd} - \sum_{i,j \in N} F_{jid} = b_{id}, \quad \forall i, j \in N(s, t), \forall d \in D \quad (10)$$

$$\sum_{i,j \in N: n==t} b_{id} - \sum_{i,j \in N: n==s} b_{di} = 0, \quad \forall n \in N(s, t) : n \neq s, n \neq t \quad (11)$$

$$\sum_{i,j \in N: n==t} b_{id} - \sum_{i,j \in N: n==s} b_{di} = -1, \quad \forall n \in N(s, t) : n = s \quad (12)$$

$$\sum_{i,j \in N: n==t} b_{id} - \sum_{i,j \in N: n==s} b_{di} = 1, \quad \forall n \in N(s, t) : n = t \quad (13)$$

$$F_{ijd} \leq U_{ijd}, \quad \forall i, j \in N(s, t), \forall d \in D \quad (14)$$

$$F_{ijd} \geq 0, \quad \forall i, j \in N(s, t), \forall d \in D \quad (15)$$

$$y_{ij} \in \{0, 1\} \quad (16)$$

$$\sum_{i,j \in N} F_{i,j} t_{i,j} \leq T \quad (17)$$

Equation 9 is the objective function to minimize the cost among both commodity and fixed travel cost of the vehicle. Eq 10 defines the flow conservation demand constraint of each node. Eq (11, 12, 13) define the network flow constraints according to source and sink node for both networks, where  $s$  and  $t$  are source and terminal, For each node  $i, i = 1, 2, \dots, n$ , an (integer) number  $b_i$  is given, representing the amount of flow produced (if  $b_i < 0$ ) or consumed (if  $b_i > 0$ ) at  $i$ . The nodes that produce flow are sometimes referred to as a source, and  $b_i$  as supply. Nodes that consume flow are called sinks, and  $b_i$  as demand. If  $b_i = 0$ , node  $i$  do not consume nor produce flow, i.e., it is a transit node. Eq 14 illustrates that flows on each link is always less than link capacity. Eq 15 states that network flow is a non-negative value. Eq 16 is the decision variable which network will be chosen. Eq 17 describes that the traversed time delay of the each edge should not exceed maximum delivered time. We want to minimize the energy consumption cost by transferring a huge range of data volume in network  $N(s, t)$  from source node  $s$  to the terminal node  $t$  considering two commodities. We solved our problem using IBM ILOG CPLEX Optimization studio solver to find the optimal network solution as per user's demands.

#### IV. NUMERICAL STUDIES

##### A. Auckland Public Transport System

We will be focusing on the Auckland city public transportation system for evaluation. Auckland is the larger city in New Zealand and its public transport system is well known for its reputation, punctuality, synchronised timetables, efficiency, frequency and high quality of service and innovation. Auckland Transport (AT) works closely with Auckland Council and the NZ Transport Agency (NZTA) on all the prospective plans. AT includes buses, trains and ferry as public transport services as shown in Fig. 3. This map clearly shows all of the Auckland equipped with so many bus stops, bus services and train services moving around Auckland. Auckland bus route datasets has been acquired from <https://opendata.arcgis.com/datasets>

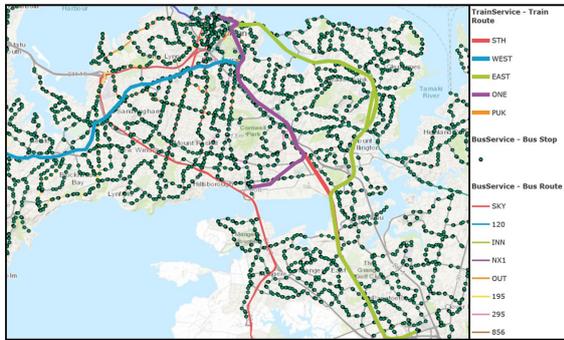


Fig. 3: Auckland Public Transport System

There are four type of bus services in Auckland which connect all parts of the city and they include: local services, connector services, peak services and frequent services. Local services operate within the local central area. There are 12 local services that operate in central Auckland. Next, 29 connector services, which are connected in all directions of Auckland. These services operate at least every 30 minutes, 7am – 7pm, 7 days a week. Peak services are operated only in peak hours on weekdays like morning and afternoon peaks and there are 7 of them. Frequent services operates every 15 minutes 7am – 7pm, 7 days a week as shown in Fig. 4(a). and we can analyze that there are maximum number of connector services, which connects different suburbs of Auckland and can help to carry data around Auckland. There are 498 bus services in total which covers 5474 bus stops. Fig. 6 shows the cumulative number of active buses and routes covered by them during different hours of the weekday. If we see the paterens of active buses during day, these buses cover maximum of the Auckland routes and there are more probabilities to go into the area of selected destination to deliver data. In such way, we can utilize existing infrastructure of public transport system and their services to disseminate data all over the Auckland region.

We consider Auckland bus service routine operation to utilize it as an alternative communication channel. These data set contains number, assignment, agency name and schedule of busses as well as their routes and the location and number

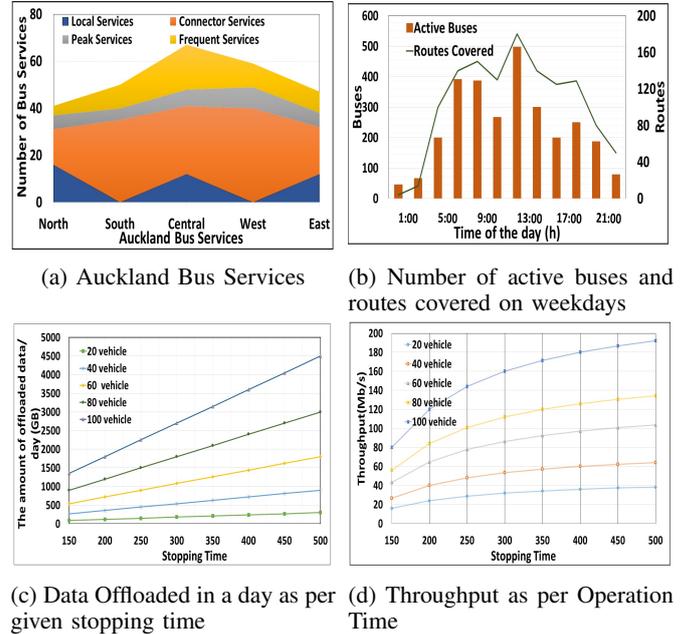


Fig. 4: Auckland Public Transport Assessment

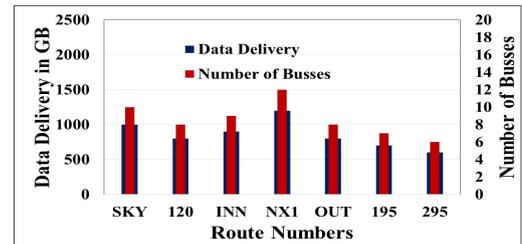


Fig. 5: Data Delivery using Public Buses

of traversed bus stops. Hence, it allows us to analyze the number, duration, and location of contacts between bus and bus stops as the basis for the contact network and subsequently its communication characteristics. To use public transport as an another communication network, we assume all busses and bus stops to be equipped with an IEEE 802.11 module with the capacity of 100 GB storage. To elaborate more, let's consider a bus service from Auckland to Henderson with 17 km distance from Auckland central and three possible public transport services such as bus no 133, 134 and the WEST train service. Let's take details of one possible route 133 to see how much data it can deliver. The GPS report of bus 133 states bus frequency, stops covered, operation hours, route pattern etc which help us to take selection of bus as per profile of user's and their demands given in flow table. These buses operate from 5.50 am until 11.55 pm for 7 days a week.

Data get offloaded onto buses when they stop at a bus stop and vary according to stopping time, according to a transit timetable. The amount of data offloaded [12] onto each bus per day is calculated as follows: The amount of data at parking spot in a day (MB) = (Effective transmission data rate (Mb/s) × Stopping time (s) × the number of vehicles in a day) / 8. Furthermore, The throughput during operation is

calculated as follows: The amount of data offloaded in a day (MB)  $\times$  8 / Operation time (s). Figs. 4 (c) and 4 (d) show the amount of data offloaded onto vehicles and throughput as per operation time. If a bus stops for 500 seconds, for a total of 54 buses, the offloading capacity of a local storage is 64.8 GB/day with an effective throughput of 22.03 MB/s. In addition to it, the transmission performance highly depends on stoppage time at a bus stop and the number of buses in a day. Fig. 5 demonstrates data delivery being done by Auckland public transport buses from central uckland to other areas which shows significant saving of energy while using public transport as an alternative communication channel.

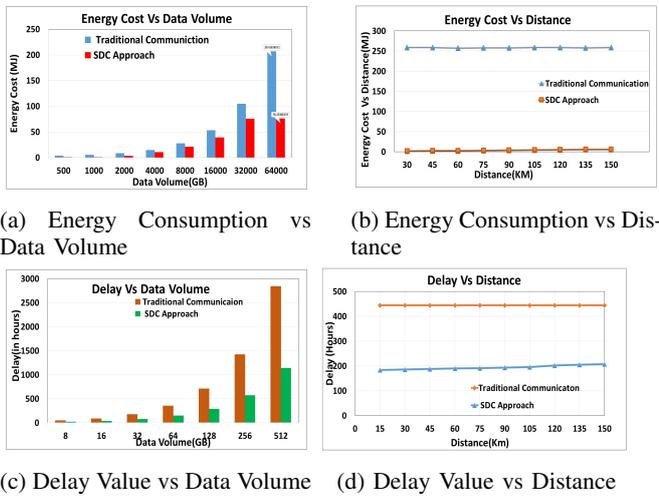


Fig. 6: Performance analysis

We plot the energy consumption and delay value as per the data volume and distance in Fig 6. The results show that our system outperforms the traditional network when we increase the distance and volume. Energy consumption is less in SDC in comparison with traditional network Data gets uploaded onto the data carrier and uploading/downloading times vary as per the size. These calculations also prove that scheduled vehicles are a better option to disseminate data for long distance transmissions.

## V. CONCLUSION AND PERSPECTIVES

In this paper, we proposed a new data offloading approach called SDC handled by a controller for sustainable data dissemination. For delay tolerant data needs, our approach aims to best utilize the existing smart public vehicles and their parking spots with local storage to offload and upload data. Our scheme utilizes vehicles' patterned movement and it matches the profile of users from a flow table to make network selection decisions for message forwarding. Accordingly, a public vehicle going in a suitable direction can carry data along with it and there is not extra fuel consumption while transmission. Our optimisation technique performed during this work helps in network selection while proving that there is significant energy saving with the vehicular network. For future work, we plan to extend our architecture by introducing data offloading models with transmission decision as per

different attributes such as delay tolerant data, non-delay tolerant data.

## REFERENCES

- [1] T. RAY, "What kind of computer will you drive?" 2014.
- [2] J. Li, Z. Peng, S. Gao, B. Xiao, and H. Chan, "Smartphone-assisted energy efficient data communication for wearable devices," *Computer Communications*, vol. 105, pp. 33–43, 2017.
- [3] I. Komnios and E. Kalogeiton, "A dtn-based architecture for public transport networks," *annals of telecommunications-Annales des télécommunications*, vol. 70, no. 11-12, pp. 523–542, 2015.
- [4] S. Naseer, W. Liu, N. I. Sarkar, P. H. J. Chong, E. Lai, and R. V. Prasad, "A sustainable vehicular based energy efficient data dissemination approach," in *Telecommunication Networks and Applications Conference (ITNAC), 2017 27th International*. IEEE, 2017, pp. 1–8.
- [5] S. Naseer, W. Liu, N. I. Sarkar, P. H. J. Chong, E. Lai, M. Ma, R. V. Prasad, T. C. Danh, L. Chiaraviglio, J. Qadir *et al.*, "A sustainable marriage of telcos and transp in the era of big data: Are we ready?" in *International Conference on Smart Grid Inspired Future Technologies*. Springer, 2018, pp. 210–219.
- [6] X. Zhang, J. Kurose, B. N. Levine, D. Towsley, and H. Zhang, "Study of a bus-based disruption-tolerant network: mobility modeling and impact on routing," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. ACM, 2007, pp. 195–206.
- [7] B. Baron, P. Spathis, H. Rivano, M. D. de Amorim, Y. Viniotis, and M. Ammar, "Centrally-controlled mass data offloading using vehicular traffic," *IEEE Transactions on Network and Service Management*, vol. 14, no. 2, pp. 401–415, 2017.
- [8] I. Marincic and I. Foster, "Energy-efficient data transfer: Bits vs. atoms," in *Software, Telecommunications and Computer Networks (SoftCOM), 2016 24th International Conference on*. IEEE, 2016, pp. 1–6.
- [9] C. V. N. Index, "Global mobile data traffic forecast update, 2016–2021 white paper," *Cisco: San Jose, CA, USA*, 2017.
- [10] R. Munjal, W. Liu, X. J. Li, J. Gutierrez, and P. H. J. Chong, "Telco asks transp: Can you give me a ride in the era of big data?" in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2017, pp. 766–771.
- [11] R. Munjal, W. Liu, X. J. Li, J. Gutierrez, and M. Furdek, "Sustainable massive data dissemination by using software defined connectivity approach," in *2017 27th International Telecommunication Networks and Applications Conference (ITNAC)*. IEEE, 2017, pp. 1–6.
- [12] S. Kashihara, M. Sanadidi, and M. Gerla, "Mobile, personal data offloading to public transport vehicles," in *Proceedings of the 2012 IEEE International Conference on Mobile Computing and Ubiquitous Networking*, 2012.