PCV: Predicting Contact Volume for Reliable and Efficient Data Transfers in Opportunistic Networks

Shiraz Qayyum*, Mehrab Shahriar†, Mohan Kumar*, Sajal K. Das† ‡
*Golisano College of Computing and Information Sciences
Rochester Institute of Technology
†Department of Computer Science and Engineering
University of Texas at Arlington
‡Department of Computer Science
Missouri University of Science and Technology
E-mail: sq8010@rit.edu*, mdmehrab.shahriar@mavs.uta.edu†, mjk@cs.rit.edu* , das@uta.edu†

***Abstract*—Exploiting opportunistic contacts between mobile devices to enable deployment of real applications through reliable and efficient data transfers poses a significant research challenge. Indeed, accurate prediction of contact volume, defined as the maximum amount of data transferable during a contact, can improve performance of deployments. However, existing schemes for estimating contact volume that make use of preconceived patterns or contact time distributions may not be applicable in uncertain environments. In this paper, we propose a novel scheme called PCV that predicts contact volume in soft real-time to enable efficient and reliable data transfers in opportunistic networks. An Android Application that learns data rate profiles has been developed to facilitate PCV. In addition, an analytical model has been developed to depict variable data rates between mobile devices. Extensive simulations are carried out on both synthetic and real world mobility traces to validate the usefulness of PCV. Experimental results show the effectiveness of our approach in terms of reliable data transfers.**

***Index Terms*—Opportunistic networks, Contact duration, Contact Volume, Data Rate Profile**

I. Introduction

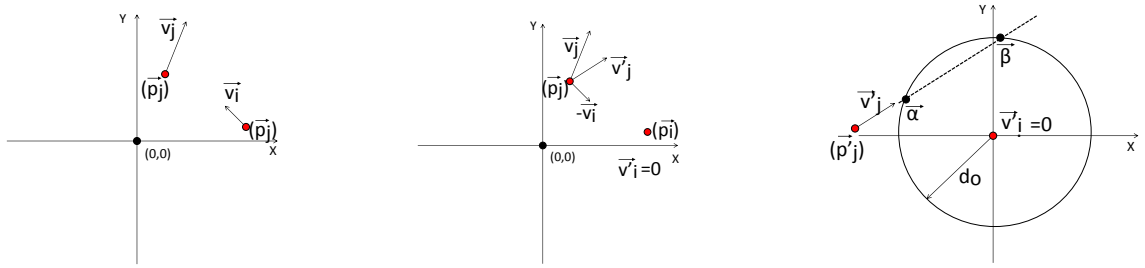
In recent years, the number of smartphone users has exceeded one billion, thus creating the opportunity for developing and deploying useful applications in opportunistic networks. As an example, consider a group of smartphone users running a bandwidth sharing application to download and possibly stream a video. In such an application, a group of nearby users simultaneously download different chunks of a video with the help of a low speed cellular network connection and share it among themselves over a high speed wireless LAN (possibly at no cost) [11]. The advantages of bandwidth sharing include, but are not limited to increased download speeds, better video quality, reliability and reduced delays. In the existing bandwidth sharing schemes [2] [28] [20] [27], it is assumed that the group of users wanting to watch the same video are within proximity of each other and are not mobile. However, situations where a group of mobile users may not be within their radio communication range, but can be connected in an opportunistic network are quite common in such public places as streets, train stations and state fairs. Employing PCV[1], it is possible to exploit contact durations among pairs of intermediate nodes to deliver video chunks to a group of mobile users reliably. When the contact duration is short, it is worthwhile to transfer at least part of the information, as compared to the full video chunk a node may have. The latter has a higher chance of failure than the former in opportunistic networks. In such a case, an accurate prediction of contact volume, defined as the maximum possible amount of data transferable during a contact, can help in making better choices (in terms of cost, delay, and/or fidelity) at the application level. Knowledge of contact volume at the start of an opportunistic contact would facilitate efficient management of resources in challenging network environments to meet the needs of a variety of application and user requirements. It should be noted that the user mobility patterns are haphazard in dynamic and unpredictable public environment. Existing works utilizing cyclic mobility patterns or forecast peer interactions [15] [14] for effective information dissemination in opportunistic networks, are likely to perform poorly in such an environment. Furthermore, the use of contact duration distribution [6] for the prediction of contact volume does not hold promise for similar reasons.

In this paper, we propose a novel scheme called PCV that effectively predicts the contact volume and enhances the usable capacity of an opportunistic network. PCV makes use of the pre-calculated data rate profile expressed as a function of distance, as well as the instantaneous velocity vectors of the mobile nodes. The energy consumption is reduced by PCV by switching on Global Position System on a device only at the time of contact. PCV is the first scheme that also supports bandwidth sharing among mobile users who may or may not be within each other's communication range.

PCV is primarily developed to assist applications in opportunistic networks, by making data transfers reliable and efficient. Conceptually, PCV serves as a middleware interface between the application and the opportunistic network layers.

[1]The work presented in the paper was completed during 2012-13, when all the authors were affiliated to the University of Texas at Arlington.

978-1-4799-0537-9/13/$31.00 ©2013 IEEE



(a) Initial scenario, i.e., before applying transformations

(b) In the Inertial Frame of Reference of node n_i

(c) After applying Euclidean Translation to position n_i at (0,0)

Fig. 1: Series of transformations applied for calculating contact duration of two nodes

PCV estimates contact volume for any pair of nodes meeting opportunistically in soft real-time and subsequently sends this information to the application layer above. Our major contributions are summarized below.

- **System Model:** Analytical model to compute instantaneous distance and contact duration between nodes. These values are used in the PCV algorithm to compute contact duration and instantaneous distance between the nodes in an opportunistic network.
- **Data Rate Profile:** A custom Android Application to acquire data rate profiles of contacts between nodes in an opportunistic network to facilitate the prediction process. Data rate profile incorporates wireless characteristics for different environments. In order to validate PCV, we have collected data rate profiles for a real world scenario, wherein the users carried Android smartphones.
- **PCV Algorithm:** PCV algorithm predicts the contact volume for pairs of nodes. The algorithm is simple, scalable and runs in linear time in terms of the number of opportunistic contacts a node makes.

The primary focus of PCV is to make file transfers reliable and efficient to improve the usable bandwidth of the underlying opportunistic network. However, we do realize the privacy and security issues that may arise out of sharing data among untrusted nodes. To address this, PCV is designed to have a compartmentalized structure in order to accommodate anonymity and encryption techniques. Moreover, PCV can also use any state of the art routing and forwarding schemes for opportunistic networks [24].

II. System Model

In this section, we develop a model for opportunistic contacts between two nodes in an opportunistic environment. Expressions for contact duration and instantaneous distances between nodes are derived using nodes' positions and velocities at the time of contact. Both contact duration and instantaneous distances of the nodes capture the dynamics of an opportunistic contact, and are critical for the estimation of contact volume.

A. Contact Duration

Let $\mathbf{N} = \{n_1, n_2, n_3, ..., n_l\}$ be the set of l nodes, where all nodes have a homogenous radio communication range d_0. Whenever two nodes n_i, $n_j \in \mathbf{N}$, where $1 \leq i, j \leq l$, come within each others' radio communication range, they are said to be in 'contact'. Let $t_{ij} = t_{ji}$ be the duration of this contact. For the sake of simplicity we assume the nodes to move in a two dimensional space. However, the analysis can easily be scaled to three dimensional space. Instantaneous position and velocity vectors are used to predict the contact duration. Let

$$\vec{p_i} = \begin{bmatrix} p_{ix} \\ p_{iy} \end{bmatrix} \ , \ \vec{p_j} = \begin{bmatrix} p_{jx} \\ p_{jy} \end{bmatrix} \quad \vec{v_i} = \begin{bmatrix} v_{ix} \\ v_{iy} \end{bmatrix} \ , \ \vec{v_j} = \begin{bmatrix} v_{jx} \\ v_{jy} \end{bmatrix}$$

be the instantaneous position and velocity vectors of the nodes n_i and n_j, at the start of the contact.

1) Inertial Frame of Reference and Euclidean Translation: In node n_i's inertial frame of reference, applying Euclidean Translation to position n_i at the origin, the transformed position and velocity vectors of n_i and n_j are,

$$\vec{p_i'} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ , \ \vec{p_j'} = \begin{bmatrix} p_{jx} - p_{ix} \\ p_{jy} - p_{iy} \end{bmatrix} \quad \vec{v_i'} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ , \ \vec{v_j'} = \begin{bmatrix} v_{jx} - v_{ix} \\ v_{jy} - v_{iy} \end{bmatrix}$$

respectively. Fig. 1 illustrates the effect of these transformations in succession. The duration t_{ij} is the time for which n_j stays within distance d_0 of node n_i, which is essentially the time it takes for a node to traverse a chord in a circle of radius d_0, as depicted in Fig.1c. The chord through the circle, defined by $x^2 + y^2 = d_0$ is traced by a line $y = mx + \sigma$, where

$$m = \frac{v_{jy} - v_{iy}}{v_{jx} - v_{ix}}$$
$$\sigma = (p_{jx} - p_{ix}) - (\frac{v_{jy} - v_{iy}}{v_{jx} - v_{ix}} \times p_{jx} - p_{ix})$$

Solving the equation of the line and circle simultaneously leads to a quadratic equation in x, as follows:

$$(m+1)x^2 + 2mcx + \sigma^2 - {d_0}^2 = 0 \qquad (1)$$

Let the solutions of the quadratic Equation (1) be

$$\vec{\alpha} = \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} \quad and \quad \vec{\beta} = \begin{bmatrix} \beta_x \\ \beta_y \end{bmatrix}$$

The effective displacement of chord that node n_j traverses while staying in contact with n_i is expressed as

$$\vec{\lambda} = \vec{\beta} - \vec{\alpha} \tag{2}$$

Therefore, the final contact duration of the nodes is given by

$$t_{ij} = \frac{\lambda_x}{v'_{jx}} = \frac{\lambda_y}{v'_{jy}} \tag{3}$$

B. Instantaneous Distance

In order to make an estimate of the contact volume, the instantaneous distance between nodes n_i and n_j is calculated as a function of time. This will be used in Section III to express the data rate profile as a function of time. In wireless communications, the received signal power P_{rec} at the receiver varies inversely with the distance d from the sender. The relationship is concretely represented as $P_{rec} \propto 1/d^{\psi}$, where the exponent ψ depends on the environment in which the nodes operate. It is also known that the transmission data rate depends on P_{rec}, making the data rate a function of instantaneous distance [9]. In a dynamic environment the nodes are usually mobile during an opportunistic contact. Assuming a constant data rate between these mobile nodes for the entire duration of contact is very likely to lead to incorrect estimations of contact volume. Therefore, our proposed model accounts for the changing distance and hence, variable data rate during the contact period.

1) Distance as a function of time: In Section II-A we see that the essential dynamics of the model are characterized by $\vec{v'_i}$, $\vec{v'_j}$, $\vec{\alpha}$ and $\vec{\beta}$. In order to exploit the inherent symmetry in the problem, a rotational transformation is applied as follows:

$$\begin{aligned} \mathbf{M'} &= \begin{bmatrix} v''_{ix} & v''_{jx} & \gamma_x & \delta_x \\ v''_{iy} & v''_{jy} & \gamma_y & \delta_y \end{bmatrix} \\ &= \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} v'_{ix} & v'_{jx} & \alpha_x & \beta_x \\ v'_{iy} & v'_{jy} & \alpha_y & \beta_y \end{bmatrix} \\ &= \mathbf{RM} \end{aligned}$$

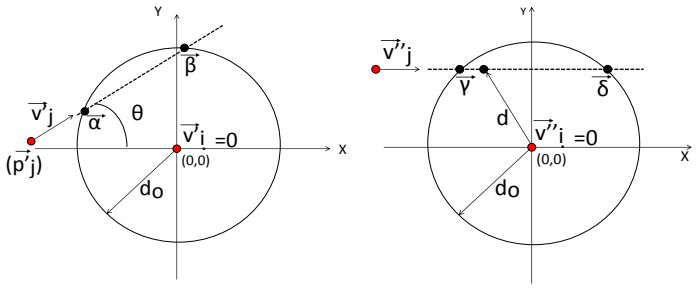
where, $\theta = arctan(\frac{v'_{jy}}{v'_{jx}})$. This rotational transformation moves every vector in $\mathbf{M}$ by the angle θ clockwise about the origin to give their rotated counterparts in $\mathbf{M'}$. Fig. 2 shows the result of this rotational transformation. Note that after rotation, the trajectory of n_j is parallel to the horizontal axis, therefore:

$$\gamma_y = \delta_y; \quad v''_{jy} = 0; \quad \vec{v''_i} = \vec{v'_i} = 0. \tag{4}$$

Using (4), we can trace the trajectory of node n_j after it comes within the communication range of n_i. Thus, for any time t where, $0 \leq t \leq t_0$ (contact duration), the motion of n_j in two dimensional space is given by

$$x = \gamma_x + (v''_{jx} \times t) \tag{5}$$

$$y = \gamma_y \tag{6}$$



(a) Before (b) After

Fig. 2: Effect of rotational transformation by angle θ

From Equations (5) and (6), the instantaneous distance between n_i and n_j is given by,

$$d = \sqrt{(\gamma_x + (v''_{jx} \times t))^2 + \gamma_y^2} \tag{7}$$

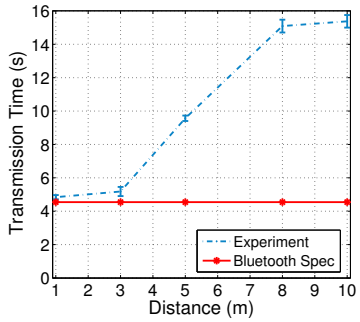
III. Data Rate Profile

The data rate between wirelessly communicating nodes depends on the distance between the nodes and the environment in which they operate. Specifically, the environment characterizes the amount of interference faced by the communicating nodes. For example, environments cluttered with buildings, trees or terrain irregularities will give rise to a higher value of the exponent ψ, i.e., the received power at the receiver drops steeply with the distance from the sender, thereby lowering the data rate at a given distance $d \leq d_0$. It is usual practice to determine the value of ψ empirically for a given environment. Empirical models have the benefit of limited reliance on detailed knowledge of the terrain and speed of execution [1]. We have developed an Android Application that learns the data rate profile for a given environment. Subsequently, in order to interpolate the discrete set of data rates obtained from the application, cubic splines [29] are used.

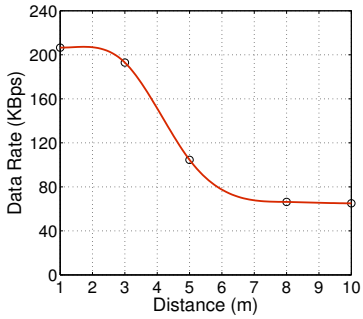
A. Custom Android Application for Learning Data Rates

We developed a custom Android Application called DatPro in order to obtain data rate profiles as a function of distance. DatPro is primarily designed to obtain Bluetooth data rates in a real-world scenario. However, it can be enhanced to work with 802.11n and WiFi Direct [3]. DatPro serves as a proof of concept and aids the application designers to better understand the data rate constraints prevalent in a specific opportunistic network environment.

In this paper, the DatPro is used to collect data rate profile in the outdoor environment of our university campus. The experiments were carried out in the walk-ways beside a multi-storied building and in a parking lot, of the university. Table I shows the different wireless devices used for collecting data rate profiles. As all the devices had similar trends in data profiling, we use results obtained only from Google Nexus One in the following sections. This particular device ran Android version 2.3.6 operating system and Bluetooth version 2.1 EDR (Enhanced Data Rate). Fig. 3a shows the effect of distance on



(a) File Transfer time at different distances between two Nexus One smartphones (*filesize* =1000 KB)



(b) Cubic spline for obtaining data transmission rate at different distances

Fig. 3: DatPro results

the file transmission period. It is evident that with increasing distance, a file requires more time for transmission, implying the drop in data transmission rate with distance.

B. Cubic Splines for Estimating Data Rates

Once the data rates are measured at different distances of separation between the wirelessly communicating devices, a curve is fitted through these discrete set of points to compute the data rate function denoted by $R(d)$. Cubic splines are used for curve fitting, as they have the advantage of giving good estimates without the danger of over fitting. Moreover, as the first and second derivatives are continuous on the break points, the curve is nicely smooth [29]. Fitting the data rate points with cubic splines defines the data rate profile as,

$$R(d) = \begin{cases} R_1(d) & d_1 \leq d \leq d_2 \\ R_2(d) & d_2 \leq d \leq d_3 \\ . & \\ . & \\ . & \\ R_k(d) & d_k \leq d \leq d_k + 1 \end{cases}$$

where the kth cubic polynomial is represented as

$$R_k(d) = a_k(d - d_k)^3 + b_k(d - d_k)^2 + c_k(d - d_k) + d_k$$

and $d_1, d_2, ..., d_k$ are the distance values at the break points. Using cubic spline on the set of data presented in Fig. 3a, we plot the changing data rate against distance in Fig. 3b. This data rate profile is used to measure the final contact volume

TABLE I: Wireless devices used to gather data rate profiles

Device Name	Platform
Google Nexus	Android 2.3.6
Motorola Milestone	Android 2.1
Acer Iconia A500	Android 4.0
Nook Color	Android 2.3.7

Data: $R(d)$,$\vec{v_i}$, $\vec{v_j}$, $\vec{p_i}$ and $\vec{p_j}$
Result: V_c'
forall the *Opportunistic contacts* **do**
  Transform velocity and position vectors to $\vec{v_i'}$, $\vec{v_j'}$, $\vec{p_i'}$ and $\vec{p_j'}$;
  t_{ij} ←Evaluate eq. (1), (2) and (3);
  $d \leftarrow$ eq. (7);
  Substitute d in $R(d)$ to get $R(t)$;
  $V_c' \leftarrow$ Evaluate eq. (8);
  return V_c' ;
end

Algorithm 1: Contact volume estimation by PCV

possible when node n_i meets n_j. It is also noted that the actual data rates obtained in this environment are considerably lower ($R \leq 220KBps$) than the ones mentioned in the official Bluetooth 2.1 EDR (Enhanced Data Rate) specifications.

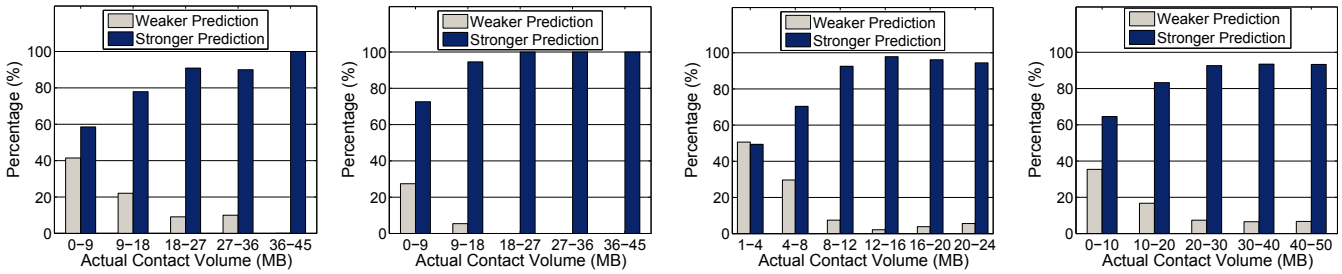
IV. Contact Volume

In Sections II and III we have derived the important parameters that PCV uses to estimate the contact volume, defined as, the maximum amount of transferrable data during an opportunistic contact. This section glues together contact duration, instantaneous distances and data rate profile in an algorithm that forms the backbone of PCV.

Let dV_c be the small amount of data that can be transferred during a contact. Then the instantaneous data rate is represented as, $\frac{dV_c}{dt} = R(t)$ where dt is an infinitesimal time unit. To derive an upper bound on the total amount of data that can be transferred, i.e., the contact volume V_c', the above expression is integrated over the contact duration t_0.

$$V_c' = \int_0^{t_0} R(t)\, dt. \tag{8}$$

PCV estimates the contact volume at the start of an opportunistic contact as detailed in Algorithm 1.



(a) Disney World (b) State Fair (c) TLW (d) SLAW

Fig. 4: Prediction strength for different ranges of actual contact volume

V. Performance Evaluation

In this section, we present results of exhaustive simulation experiments pertaining to both real world mobility traces [21], [23] as well as synthetic mobility traces [13], [22]. Data profiles from DatPro are used to predict the contact volume.

A. Simulation Setup

The real world traces used for simulation were acquired from two groups of volunteers who visited the NC State Fair [23] and Disney Land in Orlando, [21] respectively, where each of the volunteer's positions were recorded at 30 seconds intervals. There are 19 and 41 log files for the State Fair and Disney Land traces respectively. We consider each file as a mobility trace of one independent smartphone bearer agent or node. The State Fair and Disney Land traces span over a period of less than 3 hours and less than 12 hours, respectively.

Additionally, we use two synthetic traces based on the Truncated Levy Walk (TLW) [22] and Self-Similar Least Action Walk (SLAW) [13], to emulate the statistical features observed in human mobility.

An opportunistic *contact* begins when the distance between the nodes is $\leq$d$_0$ and ends when the distance is $>d_0$. The contact volume calculated through PCV is called *predicted contact volume*. We also define the *state* of a node, as its velocity at the beginning of a contact.

In our simulations, during each opportunistic contact, the objective is to transmit a file. At the instance when two agents come in contact of each other, the PCV algorithm makes a prediction of the contact volume. If the file size (S_F) is less than the predicted contact volume (V_c), then PCV initiates transmission. Otherwise the transmission is delayed. If the predicted volume is in the range of 70%-100% of the actual contact volume, we term the prediction as *Stronger Prediction*. Otherwise, we call it a *Weaker Prediction*. The choice of range for differentiating between stronger and weaker prediction is justified in Table III, which demonstrates the results obtained from all the test scenarios. The range is chosen carefully based on the minimization of the false negatives. Through experimentation, we observed that the percentage of false negatives falls with decrease in PCVR range as illustrated in Table III. For example, for Disney World the percentage of false negatives drops from 7% at a PCVR range of 50% to 100%, to 5% at a range of 70% to 100%, and remains at 5% for the range of 80% to 100%. As this trend is consistent across all traces, we have chosen 70% to 100% as a safe choice for stronger prediction. It should be noted that a weaker prediction is not synonymous with a wrong prediction, which we elaborate in section V-B.

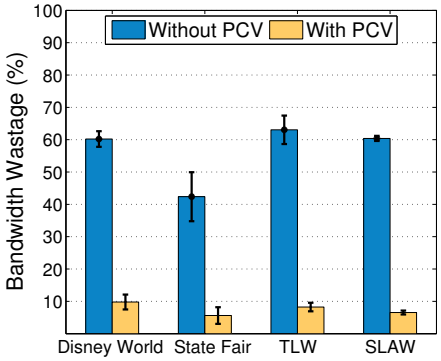


Fig. 5: Bandwidth wastage with and without prediction.

Simulation details and Relevant Metrics: We developed a custom simulator in MATLAB. To retrieve each node's position every second in real world traces, we use linear interpolation between two sample points from the trace, which are temporally 30 seconds apart.

We generate synthetic traces of 50 persons for 6 hours, using TLW [22] by modifying the implemention suggested in [26]. We set the Levy exponent for flight length distribution as α_l = 1.6 and Levy exponent for pause time distribution as β_l = 0.8. The choice of α_l and β_l is also inspired from [26], where the authors successfully generated the inter-contact time distribution of the UCSD trace [16]. The simulation

TABLE II: Distribution of files transferred

File Type	Percentage among all files	(mean, standard deviation) in megabytes
Image	48	(0.5,0.4)
Text	35	(0.75,0.6)
Video	3	(15,12)
Audio	3	(6.3,5.5)
Others	11	(2.3,2)

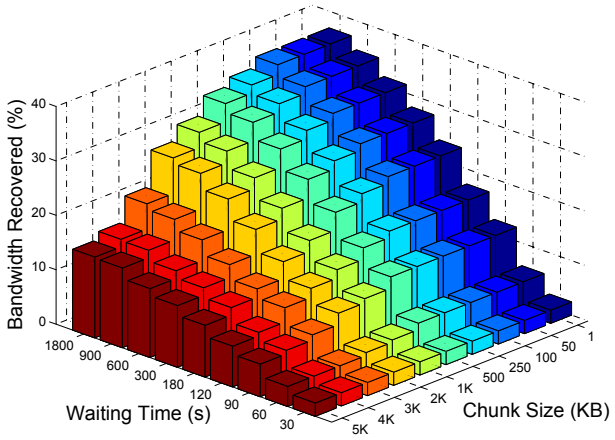


Fig. 6: Effects of chunk sizes and waiting times on bandwidth recovery for Disney World trace

area is chosen as 1 km x 1 km, and the pause times are kept within the range of 30 seconds to 5 minutes. Position information is logged after every 10 seconds. We generate the intermediate poisitions between two sample positions using linear interpolation. In addition, synthetic mobility traces of 50 people for 6 hours, are generated using SLAW [13], also by modifying the implementation suggested in [25]. Thus 600 waypoints are generated within a square area of side length of 1500m. The distance parameter α_s is set to 3, Levy exponent for pause time β_s is set to 1, Hurst parameter for waypoints (H) is set to 0.75. A node's pause time varies from 10 seconds to 5 minutes. We record each node's position at 30 second interval and retrieve the intermediate positions through linear interpolation.

In order to determine the file sizes, to be transferred in the simulation, we consider PCV's prospect in potential human centric applications aimed at online social network (OSN) and Pocket Switched Network (PSN). This makes the distribution of the downloaded multimedia contents by smartphones as a feasible candidate for determinining the sizes of our synthetic files.

To this end, results from [19] and [10] are used for assessing the downloaded contents both in terms of the number of files and bytes and the average file size. In addition, using the results from [8], we use log-normal distribution with separate parameters for each type of files. The combined decisions are elaborated in Table II.

TABLE III: Percentage of false negatives at different estimated contact volume ranges [Contact Volume Ratio (Predicted vs Actual) = PCVR]

PCVR	Disney World	State Fair	TLW	SLAW
$0.5-1.0$	7%	9%	7%	6%
$0.6-1.0$	6%	7%	6%	5%
$0.7-1.0$	5%	7%	5%	4%
$0.8-1.0$	5%	7%	5%	4%

The following performance metrics are considered in the simulation studies:

- *Prediction Strength*: Reliablity and success potential of the predictions with PCV, for different ranges of actual contact volume.
- *Bandwidth Saving*: The range of possible bandwidth savings through the implementation of PCV.
- *File Chunk size and Waiting Time Frame*: Impact of variable chunk sizes and variable waiting time frame, on bandwidth savings, after initial file transmission distruption.
- *Initial State*: Influence of nodes' initial state, defined by their velocity, on the prediction accuracy.
- *False Negative*: Impact of false negative predictions in PCV.

B. Simulation Results

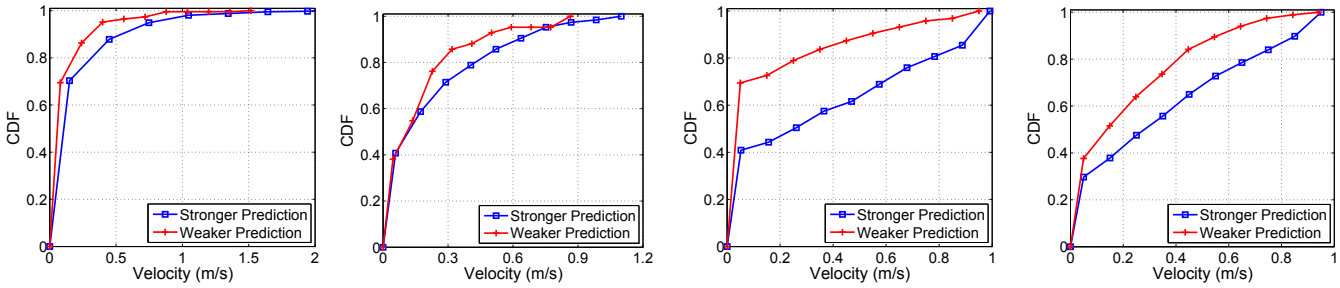
The maximum amount of transferable data during a contact volume is computed using the DatPro results plotted in Fig. 3b. Table IV shows the comparison between the practical higher limit of total data volume calculated through DatPro and theoretical higher limit mentioned in Bluetooth specifications. However, without a sophisticated prediction mechanism, even the practical higher limit is difficult to achieve.

Fig. 4 shows the percentage of stronger and weaker predictions for different ranges of actual contact volumes. As evident from the charts, the predictions are stronger for larger contact volumes. This gives an insight about the PCV's effectiveness in reducing bandwidth wastage, as the system suffers most in terms of energy and bandwidth due to its failure to transmit larger files. Again, recall that a weaker prediction does not imply a wrong prediction. Both over-prediction and under-prediction may eventually lead to successful transfer of many files. However, over-prediction runs the risk of initiating the transfer of a file larger than the actual contact volume, which is destined to fail, thereby resulting in bandwidth wastage. On the other hand, under-prediction may lead to disallowance of the initiation of eligible file transfers, resulting in unused resource. In our experiments, the average percetage of successful transmission in case of weaker predictions are 67.3%, 62.1%, 68% and 73% in terms of data volume for Disney World, State Fair, TLW, and SLAW respectively.

Fig. 5 compares bandwidth wastage in two cases: with PCV and without PCV. From the graph, it is evident that over 50% of bandwidth is wasted when nodes try to transmit during oppportunistic contacts, without predicting. However, if they make decisions based on PCV, the wastage is reduced to less than 10%. This is demonstrates a significant achievement of PCV.

TABLE IV: Percentage of maximum possible data transmission in real experiments (compared to Bluetooth Specification)

Disney World	State Fair	TLW	SLAW
36.4%	41.9%	32.8%	37%



(a) Disney World (b) State Fair (c) TLW (d) SLAW

Fig. 7: Cumulative distributive function (CDF) of stronger and weaker predictions against nodes' initial state

Transmitting data in chunks can further reduce the bandwidth wastage. It can be coupled with a waiting time, which means that the transmitting node will wait for a specific time frame to re-establish contact with the recieving node, after the initial contact is distrupted. Trying to transmit the whole file together, results in wastage of bandwidth. In Fig. 6 we plot the percentage of bandwidth, recovered from this wasted bandwidth, against different chunk sizes and waiting times in Disney World scenario[2]. Since more files transmit successfully with smaller chunks and larger waiting times, the bandwidth wastage is reduced. On the other hand, very small chunk sizes incur additional overhead whereas very large waiting times increases buffer consumption. Considering this trade-off, the applications can decide on the best operating point for a suitable chunk size and waiting time frame.

From Fig. 7 it is evident that the state of the node, which is characterized by its velocity, also affects PCV performance. Velocities are plotted along the x axis and their respective cumulative distribution function (CDF) is plotted along the y axis. As we can see from the graph, for weaker predictions, the CDF grows much faster at the beginning, meaning that when the nodes have relatively smaller velocity, it is difficult to predict their movement, which is intuitive. For example, if both nodes remain static, the predicted contact volume can be extremely large, which may not be true in reallity.

One major concern about any prediction scheme can be the false negative generation rate, that blocks the otherwise eligible transmissions. In a worst case scenario, the available bandwidth may remain unutilised, thus offsetting the bandwidth savings. To quantify this aspect, we plotted the percentage of false negatives with respect to the number of files as well as the unused bandwidth. As illustrated in Fig. 8, the portion of the files affected by false negatives is quite small to have an impact on the overall gain of the system. More importantly, in terms of bandwidth less than 10% , implying that smaller files are more likely to be affected.

[2]Results for Disney World, TLW, and SLAW are similar to State Fair scenario, so we do not report them due to lack of space

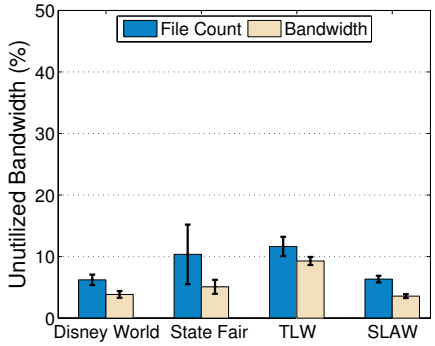


Fig. 8: Effect of false negatives

VI. Discussion

A. Conserving Battery Power

GPS and accelerometer readings on contemporary smart phones give a measure of position and instantaneous velocity. Using accelerometer readings instead of GPS has been shown to consume less energy [12] [31]. Moreover, as accelerometers are usually active on smartphones to detect screen orientations, their usage does not incur additional cost.

In order to minimize battery consumption, PCV, senses a node's position and velocity only at the time of contact. At the start of a contact, to detect whether a user is walking versus other types of movement, such as when the user is stationary but moving the device, techniques like TransitGenie [30] can be used. Furthermore, once it is known that the user is actually on the move at the start of the contact, schemes like Triggered Sensing [17] can be put to use. Here a low-power sensor e.g. accelerometer that detects the movement of the user, can trigger the GPS to gather velocity and position measurements. PCV then uses this information to make a prediction of the contact duration.

B. Bandwidth Sharing Application

From the simulation experiments, it is observed that single opportunistic contacts in most cases are inadequate to transfer considerably large volume files. As a result, users have to rely on the cellular networks to download files of large volume. However, wireless data plans are still very expensive, which

makes the large data bandwidth consumption unaffordable for many. Experimental results illustrated in Fig. 6 demonstrate that large files can be successfully transferred in multiple small chunks.

For the bandwidth sharing scenario depicted in Section I, a set of mobile nodes interested in a particular large file, first download separate chunks of the file on the high cost cellular network and then employ distributed sharing techniques [11] to procure the remaining chunks, on the low (or no) cost wireless network. PCV can further augment the distributed sharing process by fine tuning chunk sizes to available contact volume and efficient usage of available bandwidth.

VII. Related Work

There exist several algorithms for predicting future contact times among nodes. The main purpose of these predictions is to increase message delivery to the destinations with smaller bandwidth consumption. Long Vu et al. [32] constructed a model called Jyotish which provides prediction of location, stay duration, and contact of a person altogether with a considerably high accuracy. Marchini et al. [14] developed a mechanism for successfully detecting cyclic movements and predicting the next appearance of the mobile node. As proposed by Yuan at el. [33], nodes themselves determine the probability distribution of future contact times and choose a proper next-hop in order to improve the end-to-end delivery probability.

Regarding the measurement of contact volumes, Chowdhury et al. [7] proposed a framework based on IEEE 802.11n wireless to give an analytical model on contact volumes. Neto et al. [18] used the findings from [7] and predicts the contact volume as a function of received signal strength indicator (RSSI). Banerjee et al. [4] came up with the idea of using energy efficient throwboxes. These throwboxes are battery-powered stationary nodes with radios and storage, in an opportunistic nework setting. These nodes can predict the opportunistic contact volume by using a long range radio equipped with a GPS to detect mobile nodes and to measure their speed and direction.

However, none of the the above mechanisms exploit the mobility vectors coupled with data profiles at variant distance to optimize contact volumes. Cyclic pattern detection is not feasible for environments with high dynamics. Coming up with a probability distribution function as a predictor of future contact demand sufficient historical information, which is unavailable in many scenarios. . Although being inexpensive, throwboxes can not be deployed everywhere to cover areas where opportunistic networks are formed once in a while. Finally, the idea of calculating of contact volume as a function of RSSI has been proven to be inadequate [5].

VIII. Conclusion

We presented the design and evaluation of a novel scheme called PCV that predicts contact volume to enable reliable and efficient data transfers in opportunistic networks. PCV is the first scheme to exploit user mobility for bandwidth sharing applications in dynamic environments. PCV's architectural design can incorporate existing works on security, privacy, forwarding and routing. To demonstrate its flexible structure, we show how PCV uses data rate profile learnt with DatPro Application. Extensive simulation results demonstrate the efficacy of our approach in terms of reducing bandwidth wastage. Moreover, through the results we address the issue of increasing bandwidth utilization, by introducing chunks and waiting times dynamically. In future, we plan to design human centric opportunistic applications deploying PCV. We believe that accurate prediction of contact volume paves the way for real-world deployment of interesting and useful applications, such as, content management and filtering, resource scheduling, bandwidth sharing, P2P gaming and video streaming in dynamic environments.

Acknowledgment

We would like to thank Umair Sadiq and Giacomo Ghidini for their invaluable suggestions in improving the quality of the paper. This work was supported in part by the US National Science Foundation under grants CSR 0834493, IIS-1064460, and CNS-1150192. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] V. Abhayawardhana, I. Wassell, D. Crosby, M. Sellars, and M. Brown, "Comparison of empirical propagation path loss models for fixed wireless access systems," in *IEEE Vehicular Technology Conference (VTC) 2005*, vol. 1, 2005, pp. 73–77.

[2] G. Ananthanarayanan, V. N. Padmanabhan, L. Ravindranath, and C. A. Thekkath, "Combine: leveraging the power of wireless peers through collaborative downloading," in *Proceedings of the 5th ACM International Conference on Mobile Systems, Applications and Services (MOBISYS)*, 2007, pp. 286–298.

[3] Android, "Wifi direct."

[4] N. Banerjee, M. D. Corner, and B. N. Levine, "An energy-efficient architecture for dtn throwboxes," in *26th IEEE International Conference on Computer Communications (INFOCOM 2007)*, 2007, pp. 776–784.

[5] K. Benkic, M. Malajner, P. Planinsic, and Z. Cucej, "Using rssi value for distance estimation in wireless sensor networks based on zigbee," in *Proceedings of the 15th IEEE International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2008, pp. 303–306.

[6] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding," Technical Report UCAM-CL-TR-617, University of Cambridge, Computer Laboratory, Tech. Rep., 2005.

[7] H. Chowdhury, J. Lehtomäki, J.-P. Mäkelä, and S. Kota, "Data downloading on the sparse coverage-based wireless networks," *Journal of Electrical and Computer Engineering*, vol. 2010, p. 34, 2010.

[8] A. B. Downey, "The structural cause of file size distributions," in *9th IEEE Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, 2001, pp. 361–370.

[9] J. D. Gibson, *Mobile communications handbook*. CRC press, 2012, vol. 45.

[10] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: a view from the edge," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 2007, pp. 15–28.

[11] L. Keller, A. Le, B. Cici, H. Seferoglu, C. Fragouli, and A. Markopoulou, "Microcast: Cooperative video streaming on smartphones," in *Proceedings of the 10th ACM International Conference on Mobile systems, applications, and Services (MOBISYS)*, 2012, pp. 57–70.

[12] M. B. Kjærgaard, J. Langdal, T. Godsk, and T. Toftkjær, "Entracked: energy-efficient robust position tracking for mobile devices," in *Proceedings of the 7th ACM International Conference on Mobile Aystems, Applications, and Services*, 2009, pp. 221–234.

[13] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "Slaw: A new mobility model for human walks," in *IEEE INFOCOM*, 2009, pp. 855–863.

[14] M. Marchini, M. Tortonesi, G. Benincasa, N. Suri, and C. Stefanelli, "Predicting peer interactions for opportunistic information dissemination protocols," in *IEEE Computers and Communications (ISCC)*, 2012, pp. 000 512–000 517.

[15] A. Mazzini, C. Stefanelli, M. Tortonesi, G. Benincasa, and N. Suri, "Disservice: Network state monitoring and prediction for opportunistic information dissemination in tactical networks," in *IEEE MILITARY COMMUNICATIONS CONFERENCE (MILCOM)*, 2010, pp. 555–560.

[16] M. McNett and G. M. Voelker, "Access and mobility of wireless pda users," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 2, pp. 40–55, 2005.

[17] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: rich monitoring of road and traffic conditions using mobile smartphones," in *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, 2008, pp. 323–336.

[18] J. P. Neto, E. Nascimento, E. Mota, E. Cerqueira, P. Almeida, and R. Rojas, "A model for contact volume prediction in dtns," in *IEEE Symposium on Computers and Communications (ISCC)*, 2012, pp. 000 199–000 202.

[19] I. Papapanagiotou, E. M. Nahum, and V. Pappas, "Smartphones vs. laptops: comparing web browsing behavior and the implications for caching," in *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems*, 2012, pp. 423–424.

[20] M. Ramadan, L. El Zein, and Z. Dawy, "Implementation and evaluation of cooperative video streaming for mobile devices," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*. IEEE, 2008, pp. 1–5.

[21] I. Rhee, M. Shin, S. Hong, K. Lee, and S. K. . S. Chong, "CRAWDAD Trace," http://crawdad.cs.dartmouth.edu/ncsu/mobilitymodels/GPS/Disney_World, Jul. 2009.

[22] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the levy-walk nature of human mobility," *IEEE/ACM Transactions on Networking (TON)*, vol. 19, no. 3, pp. 630–643, 2011.

[23] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong, "CRAWDAD Trace," http://crawdad.cs.dartmouth.edu/ncsu/mobilitymodels/GPS/NC_State_Fair, Jul. 2009.

[24] U. Sadiq and M. Kumar, "Proximol: Proximity and mobility estimation for efficient forwarding in opportunistic networks," in *Proceedings of the 8th IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, 2011, pp. 312–321.

[25] Seongik, K. Hong, and Lee, "SLAW trace generator," http://research.csc.ncsu.edu/netsrv/?q=content/human-mobility-models-download-tlw-slaw, 2009, [Online].

[26] Seongik, M. Hong, and Shin, "Truncated Levy Walk (TLW) generator," http://research.csc.ncsu.edu/netsrv/?q=content/human-mobility-models-download-tlw-slaw, 2009, [Online].

[27] P. Sharma, S.-J. Lee, J. Brassil, and K. G. Shin, "Handheld routers: Intelligent bandwidth aggregation for mobile collaborative communities," in *Broadband Networks, 2004. BroadNets 2004. Proceedings. First International Conference on*. IEEE, 2004, pp. 537–547.

[28] M. Stiemerling and S. Kiesel, "A system for peer-to-peer video streaming in resource constrained mobile environments," in *Proceedings of the 1st ACM Workshop on User-provided Networking: Challenges and Opportunities*, 2009, pp. 25–30.

[29] G. Strang, *Computational science and engineering*. Wellesley-Cambridge Press Wellesley, MA, 2007.

[30] A. Thiagarajan, J. Biagioni, T. Gerlich, and J. Eriksson, "Cooperative transit tracking using smart-phones," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, 2010, pp. 85–98.

[31] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, 2009, pp. 85–98.

[32] L. Vu, Q. Do, and K. Nahrstedt, "Jyotish: A novel framework for constructing predictive model of people movement from joint wifi/bluetooth trace," in *Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on*. IEEE, 2011, pp. 54–62.

[33] Q. Yuan, I. Cardei, and J. Wu, "Predict and relay: an efficient routing in disruption-tolerant networks," in *Proceedings of the 10th ACM International Symposium on Mobile Ad hoc Networking and Computing*, 2009, pp. 95–104.