Temporal Coverage Based Content Distribution in Heterogeneous Smart Device Networks

Wei Peng[*], Feng Li[†], and Xukai Zou[*],
[*]Department of Computer and Information Science
[†]Department of Computer, Information, and Leadership Technology
Indiana University-Purdue University Indianapolis,
Indianapolis, IN, U.S.A.

Abstract—**The present work studies content distribution in heterogeneous smart device networks, in which all smartphones/tablets can communicate through proximity channels such as Bluetooth/NFC/Wi-Fi Direct when they are in proximity, but only some devices have the cellular data communication capability. In the context of recent applications of content distribution in smart device networks such as mobile offloading and enterprise network defense prioritization, we propose a temporal coverage based scheme that exploits nodes' encounter regularity and content's delivery delay tolerance to reduce content delivery costs. Using kernel-density estimation (KDE) on the readily available proximity encounter records, we propose a network structural property, T-covering set, and a corresponding localized algorithm that distributedly elects a T-covering set from the underlying network. Using real Bluetooth encounter traces, we demonstrate that temporal coverage based content distribution using T-covering set can significantly reduce content delivery cost with minimal delay and no sacrifice in coverage.**

Index terms—**heterogeneous smart device network, temporal coverage, content distribution, T-covering set, kernel density estimation, temporal coverage quality**

I. Introduction

In this work[1], we consider *content distribution* in *heterogeneous* smart device networks. By "heterogeneous smart device network," we mean a collection of smartphones/tablets in which the *cellular data channel* is available on only *some* devices, but *all* devices can transfer data through the *proximate channel* such as Bluetooth/NFC/Wi-Fi Direct. In other words, such networks consist of a few constantly available links to a set of nodes (backed by the cellular channel) and many intermittently available links between (potentially) all the nodes in the network (backed by the proximate channel and defined by the mobility of the nodes). The heterogeneity models the common scenario that the cellular data channel is not available on many tablets (due to the lack of 3G/4G cellular transceivers) or some smartphones (explicitly disabled by their users due to cost or security concerns). On these devices, when the (infrastructural) Wi-Fi channel is not available, the proximate channel is the only means of data communication.

We consider the common scenario in which the mobility of the network nodes, although cannot be predicted precisely, nevertheless have regularity [1, 2, 3]. Examples of such networks are all the smart devices of regular students and faculty/staff members on a university campus or of employees on an enterprise site. In fact, given a densely populated and frequently visited area, the set of smart devices owned by frequent visitors often exhibit such encounter regularity.

Content distribution in smart device networks have multiple applications, two of which are mobile cellular data offloading [4, 5, 6] and prioritized defense deployment in enterprise networks [7]. In these applications, a piece of data (e.g., user-subscribed content in mobile cellular data offloading or vulnerability patches in prioritized defense deployment) is injected into or collected from the network through the cellular channel and is propagated among the nodes in the network through the proximate channel. A common objective is to minimize monetary or energy costs by reducing the number of times the content is downloaded through the cellular channel or duplicated through the proximate channel. Moreover, due to objective (e.g., the high costs or absence of a central coordination mechanism [6]) and subjective (e.g., privacy concerns [7]) constraints, it is desirable that the content distribution process emerges from the collective effect of *localized* forwarding decision made by intermediate nodes without central coordination. These settings are formulated in Section II.

The key problem addressed in this work is how to exploit content's delay tolerance for more cost-effective content distribution (i.e., fewer copies over the proximate channel are considered to be more cost-effective) in a heterogenenous smart device network. The key ideas of this work towards addressing this problem are proposing:

- a temporal-spatial structural property (*temporal coverage*) of heterogeneous smart device networks that exploits the temporal regularity of proximate encounters in such networks for effective content distribution.
- algorithms that distributedly (i.e., each device runs the algorithm without central coordination) elect a temporal covering "backbone" from such networks, based on devices' local proximate encounter records.

To the best of our knowledge, novelties and contributions of our work include:

- Unlike existing applications of content distribution in smartphone network [4, 5, 6, 7, 8], in which the network nodes are homogeneous with regard to cellular data communication capability (i.e., all nodes can push/pull data from cellular links at will), we consider the more challenging heterogeneous setting, in which nodes without cellular links can only receive/send data through proximate

channels such as Bluetooth/NFC/Wi-Fi Direct.

- We define the concept of temporal coverage based on quantitative metrics of proximate channel's temporal quality using kernel-density estimation (KDE), which preserves certainty about such estimation that is otherwise lost in simpler statistical metrics such as average or expected [7] inter-encounter interval.
- We propose localized algorithms (Section IV) that distributedly elect temporally covering nodes without central coordination.
- We verify the proposed algorithm's effectiveness for content distribution in heterogeneous smart device networks with simulations using real public Bluetooth encounter traces.

II. PROBLEM FORMULATION

The problem discussed in Section I can be formulated as follows. Let U be a set of nodes in a heterogeneous smart device network, U_c and $U_{\bar{c}}$ be the sets of nodes with and without cellular data channel, respectively: $U_c \cup U_{\bar{c}} = U$ and $U_c \cap U_{\bar{c}} = \emptyset$. In this network:

- Content are generated outside the network and injected into the network through the cellular channel, i.e., the nodes with cellular data links U_c (i.e., the "seeds" hereafter) are the interface between network and the outside Internet.
- Content can be forwarded between two devices when they move close enough to establish a proximate channel, in which case we say they encounter with each other.
- Every $u \in U$ records its past encounters with other nodes, say, with $v \in U$: $[s_1^{u,v}, e_1^{u,v}], \ldots, [s_{k_{u,v}}^{u,v}, e_{k_{u,v}}^{u,v}]$ ($s_1^{u,v} < e_1^{u,v} < \ldots < s_{k_{u,v}}^{u,v} < e_{k_{u,v}}^{u,v}$), in which u encounters v (and hence u can send data to v through the proximate channel) during the time windows $[s_i^{u,v}, e_i^{u,v}]$ (for $i \in \{1, \ldots, k_{u,v}\}$); conversely, no content can be forwarded between u and v during $(e_i^{u,v}, s_{i+1}^{u,v})$ (for $i \in \{1, \ldots, k_{u,v} - 1\}$) due to the lack of communication channel.

The objective is to minimize the content distribution cost, defined as the number of times the content is forwarded from one device to another through the proximate channel, without central coordination.

The challenge of achieving cost-effective content distribution in heterogeneous smart device networks can be better understood by considering the following schemes. In each case, node u has obtained the content and is deciding whether to forward the content to the nodes it encounters in the future.

Eager multiple forwarding. u forwards the content once to every node it encounters. This is known as flooding or epidemic routing in literature [9]. The overall delivery delay is minimized. However, the delivery cost can be higher than necessary. Nevertheless, it envelops the proximate-channel-based data propagation process from the outside—no data propagation through the proximate channel can deliver the content faster than the eager multiple forwarding.

Eager k forwarding. u forwards the content once to the first k nodes it encounters [2]. The delivery cost is bounded from above by $k|U|$: Each node forwards the content at most k time. Delay at each intermediate nodes is also minimized. Depending on proximate encounter opportunities and the choice of k,

eager k forwarding's performance ranges from eager single forwarding [3] to eager multiple forwarding [3]. However, since proximate encounter opportunities are often not uniform among nodes, it is difficult, if not impossible, to find a (global) k that performs optimally.

Random forwarding. Upon encountering another node v, u makes a random decision of whether to forward the content to v. u will forward the data to v at most once to avoid duplication. If the random decision is unbiased (i.e., equal chance of forwarding/not forwarding), the delivery cost is halved comparing with eager multiple forwarding, while random forwarding does not suffer from the delivery failure as in eager single forwarding. The relative delivery cost to eager multiple forwarding can be tuned by adjusting the forwarding decision's odds: Lower forwarding odds correspond to lower delivery costs. However, it is not clear how to optimally tune the forwarding odds without global coordination or the ability to detecting in-network content saturation as required by more sophisticated adaptive random forwarding schemes (e.g., the work by Liu and Wu [10]), which are not generally available.

III. TEMPORAL QUALITY METRICS

In light of the schemes discussed at the end of Section II, our key idea of improvement towards cost-effective content distribution (Section III-C) is to apply these forwarding rules to, instead of the full network, a restricted set of nodes that we call a *temporal covering set* (Section III-B). Intuitively, a temporal covering set is a proximate-channel content distribution backbone with *strong* internal connectivity and *full* external coverage of the whole network. Both the internal connectivity and external coverage are defined on quantitative *temporal quality* metrics of proximate channels based on the readily available encounter records.

A. Temporal quality of proximate channels

Based on its encounter records with v (i.e., the encounter intervals $[s_1^{u,v}, e_1^{u,v}], \ldots, [s_{k_{u,v}}^{u,v}, e_{k_{u,v}}^{u,v}]$), u can estimate the *temporal quality* of its proximate channel with v, in terms of the proximate channel's potential of forwarding the content timely.

A straightforward idea is to use *average inter-encounter interval*, defined as

$$\frac{1}{k_{u,v}-1} \sum_{i=1}^{k_{u,v}-1} \left(s_{i+1}^{u,v} - e_i^{u,v} \right).$$

A smaller average inter-encounter interval between two regularly encountered nodes indicates that content are more likely can be forwarded from one node to the other timely, and hence their (opportunistic) proximate channel is of a better temporal quality.

However, as will be discussed shortly, average inter-encounter interval fails to capture the *certainty* about proximate channel quality and can lead to counter-intuitive results. Therefore, we propose the following temporal quality metric of proximate channels based on kernel density estimation (KDE).

A KDE[2] $\hat{f}(x)$ of u's inter-encounter intervals to v, with the Epanechnikov kernel $K(x) = \frac{3}{4}(1-x^2)\mathbf{1}_{|x|\leq 1}$ [11], is:

$$\hat{f}_{u,v}(x) = \frac{1}{k_{u,v}-1}\sum_{i=1}^{k_{u,v}-1} K(x-(s_{i+1}^{u,v}-e_i^{u,v})), \quad (1)$$

in which $\mathbf{1}_{|x|\leq 1}(x)$ is the indicator function on the set $\{x | -1 \leq x \leq 1\}$ that equals to 1 when $|x| \leq 1$ and equals to 0 otherwise. Then, the T-*coverage (temporal) quality* $d_u^T(v)$ of u's proximate channel to v is defined as:

$$d_u^T(v) = \int_{-\infty}^{T} \hat{f}_{u,v}(x)\mathrm{d}x. \quad (2)$$

As a special case, if u has never encountered v, or no inter-encounter interval is less than the parameter T, $d_u^T(v)$ is defined to be 0. By Equation (2), $0 \leq d_u^T(v) \leq 1$ and $d_u^T(v) = d_v^T(u)$. Note that $d_u^T(v)$ can be computed locally by u from information readily available to u, i.e., its encounter records with v.

T in Equation (2) is a time-domain *quality threshold* parameter that is used to filter out sporadic or long-delay opportunistic links between nodes. Without T as the integral upper bound, the integration (Equation (2)) of the kernel $\hat{f}_{u,v}$ (Equation (1)) from $-\infty$ to ∞ would always be 1 by the definition of smoothing kernels, and thus cannot be used to compare temporal quality of their proximate channel. In contrast, integration from $-\infty$ up to T in Equation (2) endows the temporal quality metric $d_u^T(v)$ the semantics of an estimation of the probability that u will encounter v at least once within a time window of T. Greater $d_u^T(v)$ translates to a better chance that u can deliver content to v *timely* through their opportunistic proximate channel.

Comparing with average inter-encounter interval, KDE-based proximate channel quality estimation (Equations (2)) is more nuanced. To see this, consider an example with time unit of seconds, $T = 110$, and 10 groups of inter-encounter interval records: Group i ($i \in \{1, 2, \ldots, 10\}$) consists of 2^i pairs of interleaved 100 and 200, i.e., u encounters v with periodic intervals of $100, 200, 100, 200$, etc.. The average inter-encounter interval for all 10 groups has the same value of 150, which is greater than the proximate channel quality threshold $T = 110$. This suggests that the quality of these proximate channels does not meet expectation. However, the fact that "u periodically encounters with v in 100 seconds" suggests otherwise.

In contrast, Figure 1 and Table I show that KDE-based proximate channel quality $d_u^T(v)$ preserves more temporal quality information about the opportunistic proximate channel between u and v (i.e., producing a continuous, rather than binary, degree of satisfying quality expectation with regards to T) and captures the differences in temporal quality of the proximate channel between these groups *in a single number*: Proximate channel temporal quality derived from a group with $2 \times 2^{10} = 2048$ data points (0.410 from Table I) is intuitively better (i.e., more robust) than the estimation that is derived from a group with only $2 \times 2^1 = 4$ data points (0.293 from Table I).

[2]We deliberately omit the "smoothing bandwidth" parameter, often denoted by the symbol h, to simplify the (already complex) notation; it is understood that a default smoothing bandwidth (e.g., the R implementation of KDE specifies the algorithm of computing the smoothing bandwidth from inputs) is used here.
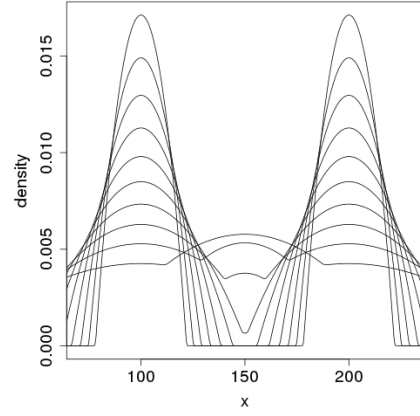

Fig. 1: The kernel density estimation (KDE; Equation (1)) for 10 groups of inter-encounter interval records with 2^i pairs of interleaved 100 and 200 in group i. The emerging dual peaks with increasing i correspond to increasing certainty about the density distribution of the nodes' inter-encounter intervals. KDE preserves u's certainty about the temporal quality estimation of its proximate channel with v that is otherwise lost in average inter-encounter interval.

i in 2^i	$d_u^T(v)$	i in 2^i	$d_u^T(v)$
1	0.293	6	0.346
2	0.303	7	0.360
3	0.312	8	0.375
4	0.323	9	0.391
5	0.334	10	0.410

TABLE I: KDE-based T-coverage temporal quality (with $T = 110$) $d_u^T(v)$ (Equation (2)) of u's proximate channel to v can capture the quality differences of the different groups shown in Figure 1.

B. Temporally covering set

For a pair of nodes $u, v \in U$, if the T-coverage quality $d_u^T(v) > 0$, we define a directed edge from u to v with a weight of $d_u^T(v)$—in this case, we say that u T-*covers* v or, equivalently, v is T-*covered* by u. These edges on U define a directed weighted graph, which we also denote as U when there is no ambiguity in the context. A set of nodes $\mathcal{D}_T \subset U$ is a *temporally covering set with temporal threshold of T* ("T-covering set" for brevity) if:

- (Coverage) For each node $u \in U$, either $u \in \mathcal{D}_T$ or there is a node $v \in \mathcal{D}_T$ such that u is T-covered by v.
- (Connectivity) For each node $u \in \mathcal{D}_T$, either u is a seed (i.e., $u \in U_c$), or there is a seed $v \in U_c$ (i.e., v is equipped with cellular data channel) such that there is a path (i.e., a chain of consecutively T-covered nodes) from v to u.

The nodes $\mathcal{D}_T$ are the T-dominators (or simply "dominators"), and the nodes that are T-covered by other nodes the T-dominatees (or simply "dominatees"), i.e., "dominators T-cover dominatees." By Connectivity, non-seed dominators are also dominatees.

C. Temporal coverage based content distribution

Coverage and Connectivity, coupled with the interpretation of T-coverage temporal quality as the probability of timely encounters, essentially make a T-covering set a virtual backbone for content distribution in a heterogeneous smart device network. More concretely, if we restrict the "eager multiple forwarding"

rule (Section II) to the T-covering set (i.e., only T-dominators will forward data):

- Connectivity dictates that each dominator can receive the content through a chain of dominators from the seeds (where the content is injected or collected).
- Coverage dictates that each non-dominator shall be directly reachable from a dominator timely.

Therefore, this temporal coverage based content distribution scheme (i.e., eager multiple forwarding restricted to a T-covering set) allows content to be delivered from seeds to any node with only timely encountered nodes serving as intermediaries. Intuitively, this allows more cost-effective content delivery (since the T-covering set is a subset of the whole network) than eager multiple forwarding, without incurring delay penalty (for delaying content delivery to destinations too much).

In this scheme, delivery cost is positively associated with the size of the T-covering set (number of dominators and the edge density of the covering set). Therefore, the localized dominator election algorithms presented next make efforts to reduce the number of elected dominators.

IV. ALGORITHM

The core of our solution to cost-effective content distribution in heterogeneous smart device networks is the following *localized* dominator election algorithm, in which the nodes, instead of being coordinated centrally, turn themselve into dominators/non-dominators based on the information they gather from their encounters with other devices. In the algorithm, each node u locally maintains two lists about other nodes: the *upstream list* $L_\uparrow(u)$ and *downstream list* $L_\downarrow(u)$.

Initially: all the seeds U_c (i.e., nodes that are equipped with cellular channel) turn dominators and will remain so throughout the election process; all the non-seeds $U_{\bar{c}}$ (temporally) turn non-dominators and may turn dominators by localized election later. Both $L_\uparrow(u)$ and $L_\downarrow(u)$ are both initially empty (i.e., $L_\uparrow(u) = L_\downarrow(u) = \emptyset$) for every node u in the network.

When u encounters v, u first updates its T-coverage quality $d_u^T(v)$ (which, as discussed after Equation (2), equals to $d_v^T(u)$), and then carries out the following information exchange procedure:

- u sends its seed/dominator status to v.
- u sends $L_\uparrow(u)$ and $L_\downarrow(u)$ to v.
- u receives $\{d_v^T(w) | w \in L_\downarrow(u)\}$ and $\{d_v^T(w) | w \in L_\uparrow(u)\}$ from v.

v follows the same procedure by swapping the symbols u and v. The amount of exchanged information is linear to the number of nodes that they have encountered in the past (rather than the size of the network unless the network is dense) and can be, for example, piggy-backed to periodic beacons[3].

After the information exchange, u has all the information needed to independently carry out Algorithm 1, in which u will turn a dominator if its updated upstream/downstream lists are both non-empty, and will turn a non-dominator if its downstream list is empty. v follows the same procedure by, again, swapping the symbols u and v in Algorithm 1.

---

[3]See, for example, Wu's discussion [12] on the implementation of this information exchange through periodic beacons.

Algorithm 1 u's local decision process on whether to change its dominator status when u encounters v after they have exchanged information.

```
 1:   ▷ only consider quality opportunistic links
 2:  if d_u^T(v) > 0 then
 3:       ▷ u updates L_↑(u) and L_↓(u)
 4:     if v is a dominator then
 5:        if v ∈ U_c or u is a non-dominator then
 6:           L_↑(u) ← L_↑(u) ∪ {v}
 7:           L_↓(u) ← L_↓(u) \ {v}
 8:        end if
 9:        for w ∈ L_↓(u) do
10:           if d_v^T(w) > d_u^T(w) then   ▷ if w is better T-dominated by v than by u
11:              L_↓(u) ← L_↓(u) \ {w}
12:           end if
13:        end for
14:     else if v is a non-dominator then
15:        L_↑(u) ← L_↑(u) \ {v}
16:        x ← TRUE   ▷ x = TRUE if u is v's best T-dominator
17:        for w ∈ L_↑(u) do
18:           if d_v^T(w) > d_u^T(v) then   ▷ if v is better T-dominated by w than by u
19:              x ← FALSE
20:              go to 23
21:           end if
22:        end for
23:        if x = TRUE then   ▷ u is v's best dominator
24:           L_↓(u) ← L_↓(u) ∪ {v}
25:        end if
26:     end if
27:     ▷ u sets its dominator status based on whether L_↓(u) and L_↑(u) are empty
28:     if u ∉ U_c then   ▷ only non-seeds change dominator status
29:        if L_↓(u) = ∅ then
30:           u turns a non-dominator
31:        else if L_↑(u) ≠ ∅ and L_↓(u) ≠ ∅ then
32:           u turns a dominator
33:        end if
34:     end if
35:  end if
```

The essence of Algorithm 1 is that:

- u will include a dominator v as (one of) its upstream $L_\uparrow(u)$ (line 6) if u thinks (based its local information after the exchange) that v connects u to one of the seeds, i.e., for Connectivity. If so, u will in turn consider delegate its downstream $w \in L_\downarrow(u)$ to v (lines 9–13) if v is a strictly better dominator (defined by the relation $d_v^T(w) > d_u^T(w)$ on line 10).
- u will include a non-dominator v as (one of) its downstream $L_\downarrow(u)$ (line 24) if u thinks (again, based on its local information) that none of u's upstream dominates $L_\uparrow(u)$ strictly better than u does (the logic on lines 16–25).
- The "strictly better" comparison (lines 10 and 18) prevents two nodes from mutually delegating the dominator responsibility for a third node w to each other and thus leaves w (wrongfully) uncovered.

Thus, Connectivity (i.e., content coverage) is maintained and the number of dominators (i.e., delivery cost) is also reduced.

V. VERIFICATION

We verify the effectiveness of the proposed temporal coverage based content distribution scheme by simulating content distribution processes with real public Bluetooth encounter traces and comparing it to the alternative schemes listed in Section II.

A. Dataset and setup

We use the publicly available Bluetooth encounter dataset sigcomm2009 [13, 14], downloaded from the CRAWDAD
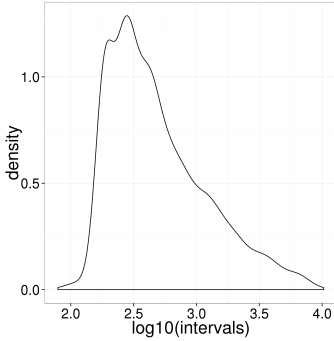
Fig. 2: Smoothed density distribution of intervals between consecutive encounters in the sigcomm2009 dataset.

wireless dataset archive[4]. The raw trace (the "proximity.csv" trace in the dataset) consists of timestamped periodic Bluetooth proximity device discovery records of 76 users during the SIGCOMM 2009 conference.

Based on the meta-data, we filter out sporadic devices (those with ID over 100 in the dataset), and transform the periodic scanning records into encounter events ("sessions"). Specifically, since the devices make a scanning every 120+/-10.24 seconds (randomized) for 10.24 seconds, we combine consecutive scanning records between a pair of devices within that time window into the same session. Moreover, only 48 out of the 76 nodes regularly meet each other up to the trace timestamp of about $12,500$ (out of timestamp of up to about $35,000$, after which the recorded encounter are sporadic and the performance curves shown below go flat). Therefore, we zoom in to that segment of trace to show the details of the results below.

Figure 2 shows the (smoothed) density distribution of inter-encounter intervals (delay between two consecutive sessions for a pair of nodes) of the devices. Note that the x axis is in logarithmic scale. In the results shown below, we set the temporal quality threshold T to $1,000$ (corresponding to the x axis value of 3.0 in Figure 2) to include enough temporal-spatial links without obliterating the quality value: As briefly discussed in Section III-A, choosing a too small or too small threshold T would lead to the quality metric $d_u^T(v)$ to be all 0 (for too small T) or 1 (for too large T) and hence cannot capture the temporal quality of different proximate channels. The results below show that our choice of $1,000$ does a fair job in capturing such quality differences. A general heuristics is left for future work.

We simulate the content distribution processes with the data forwarding rules discussed in Section I: eager multiple forwarding (emulti), eager single forwarding (esingle), and random forwarding with a 50% forwarding chance at each encounter (random 50). As for the proposed algorithm, we consider the T-coverage-based forwarding (Section III-C; tdom) and a variant of T-coverage-based forwarding with the dominator has a 50% chance of forwarding at each encounter (tdom50) to be comparable with random50.
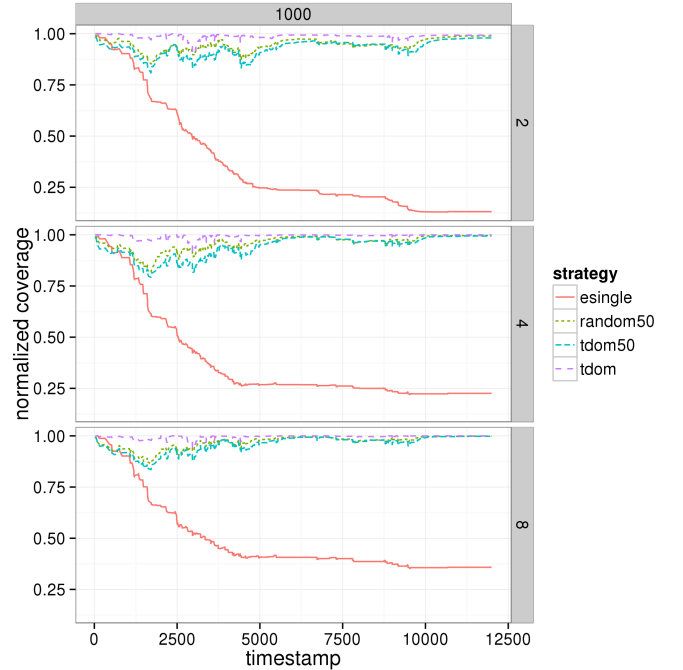

Fig. 3: Average content distribution coverage normalized by the eager multiple forwarding scheme with different numbers of seeds over 100 random runs. The row headings show the number of seeds and the column heading shows the temporal quality threshold T for the temporal coverage based schemes. Scheme notation: esingle (eager single forwarding), random50 (50% random forwarding), tdom50 (50% random T-coverage-based forwarding), tdom (T-coverage-based forwarding).

	esingle	random50	tdom50	tdom
2	5577	271	397	81
4	5530	199	306	29
8	4725	173	241	25

TABLE II: Average content delivery delay comparing to the eager multiple forwarding scheme with the same settings and notation as in Figure 3.

B. Simulations and results

Figures 3 and 4 show the average coverage (the number of nodes that receive the content) and delivery cost (the number of times the content get sent from one node to another) normalized with emulti (i.e., by arithmetic division of the raw numbers) throughout the content distribution processes for different numbers of seeds over 100 random runs; Table II shows the average content delivery delay comparing to emulti with the same settings. In computing the delays, nodes that have not received the content by the end of the content distribution process (there are many such instances for esingle) are considered to receive the content at the last timestamp; otherwise, the indefinite delay could not be used for computing the average delay.

Since emulti envelops the proximate-channel-based content distribution process from the outside (as discussed in Section II), normalizing the results with emulti in Figures 3 and 4 and in Table II clearly show how each scheme exploits content's delay tolerance to improve content delivery costs. The results indicate that, by restricting the eager multi forwarding rule to the locally
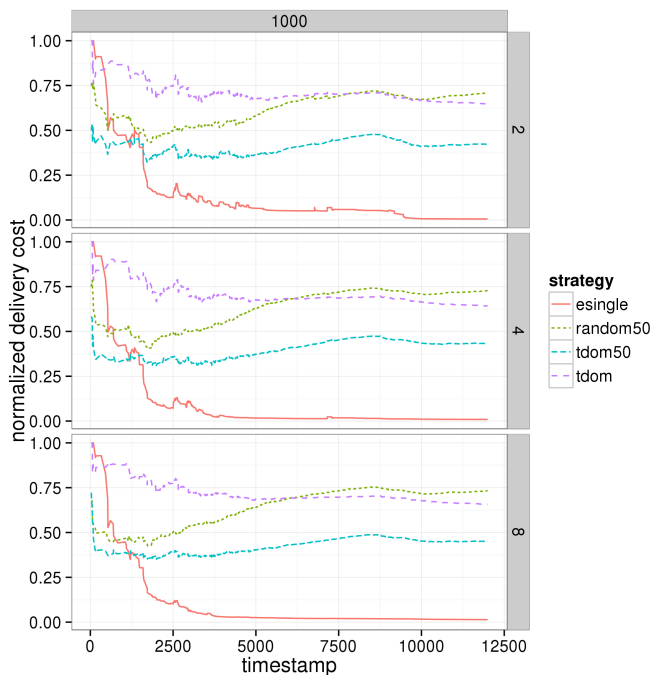
Fig. 4: Average content delivery cost normalized by the eager multiple forwarding scheme with the same settings and notation as in Figure 3.

elected temporal covering set, tdom reduces content delivery cost by 25% (Figure 4) with minimum delays (Table II) and little sacrifice in coverage (Figure 3) comparing with alternatives such as random50 and esingle. Moreover, if modest delays and coverage loss are allowed, tdom50 can be applied to reduce the delivery cost of random50 by another 25%. In summary, these results show that the temporal covering set is a cost-effective (virtual) content distribution backbone for heterogeneous smart device networks.

VI. RELATED WORKS

The work is motivated by extending applications of content distribution in homogeneous smart device networks (in which *all* nodes have cellular data capability that can be activated on demand) such as mobile cellular [4, 5, 6] and enterprise network defense prioritization [7] to heterogenenous networks (in which only *some* nodes have cellular data capability). One challenge of the heterogeneous setting is the requirement of Connectivity to seeds. In particular, the concept of temporal coverage is inspired by the work on enterprise network defense prioritization [6], which extends previous works [15, 16, 17] on (spatially) connected dominating set (CDS) based routing in ad hoc network (MANET) to the temporal dimension, by exploiting the regularity [1, 2, 3] exhibited by many proximity-channel-based smartphone networks, as a prominent application of delay-tolerant networks (DTNs) [18, 19] that have received significant research in the past decade.

VII. CONCLUSION

We propose temporal coverage based content distribution to effectively exploits content's delay tolerance for reducing content distribution costs in heterogeneous smart device networks. KDE is used to process readily available encounter records to capture the temporal quality of the proximate channel that eludes simpler measurements such as the average inter-encounter interval. Using real Bluetooth encounter traces, we demonstrate that temporal coverage based content distribution significantly cuts content delivery cost with minimal delay and no sacrifice in coverage.

REFERENCES

[1] B. Jiang, J. Yin, and S. Zhao, "Characterizing the human mobility pattern in a large street network," *APS Physical Review E*, vol. 80, no. 2, p. 021136, 2009.

[2] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: the multiple-copy case," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 1, pp. 77–90, 2008.

[3] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The single-copy case," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 1, pp. 63–76, 2008.

[4] B. Han, P. Hui, and A. Srinivasan, "Mobile data offloading in metropolitan area networks," *ACM SIGMOBILE Mob. Comput. Commun. Rev. (MC2R)*, vol. 14, no. 4, pp. 28–30, 2011.

[5] B. Han, P. Hui, V. Kumar, M. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Trans. Mob. Comput. (TMC)*, vol. 99, no. 5, pp. 821–834, 2012.

[6] W. Peng, F. Li, X. Zou, and J. Wu, "The virtue of patience: Offloading topical cellular content through opportunistic links," in *Proc. of IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)*, 2013.

[7] W. Peng, F. Li, K. J. Han, X. Zou, and J. Wu, "*T*-dominance: Prioritized defense deployment for BYOD security," in *Proc. of IEEE Conference on Communications and Network Security (CNS)*, 2013.

[8] B. Han, P. Hui, V. Kumar, M. Marathe, G. Pei, and A. Srinivasan, "Cellular traffic offloading through opportunistic communications: a case study," in *Proc. ACM CHANTS*, 2010.

[9] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks*, vol. 51, no. 10, pp. 2867–2891, 2007.

[10] C. Liu and J. Wu, "An optimal probabilistic forwarding protocol in delay tolerant networks," in *Proc. of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC)*, 2009.

[11] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Probability & Its Applications*, vol. 14, no. 1, pp. 153–158, 1969.

[12] J. Wu, "Extended dominating-set-based routing in ad hoc wireless networks with unidirectional links," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 13, no. 9, pp. 866–881, 2002.

[13] A.-K. Pietiläinen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, "Mobiclique: Middleware for mobile social networking," in *WOSN'09: Proceedings of ACM SIGCOMM Workshop on Online Social Networks*, August 2009.

[14] A.-K. Pietiläinen and C. Diot, "Dissemination in opportunistic social networks: The role of temporal communities," in *MobiHoc'12: Proceedings of the 13th International Symposium on Mobile Ad Hoc Networking and Computing*, June 2012.

[15] J. Wu, F. Dai, and S. Yang, "Iterative local solutions for connected dominating set in ad hoc wireless networks," *IEEE Trans. Comput. (TC)*, vol. 57, pp. 702–715, 2008.

[16] S. Yang, J. Wu, and F. Dai, "Efficient directional network backbone construction in mobile ad hoc networks," *IEEE Trans. Parallel and Distrib. Syst. (TPDS)*, vol. 19, no. 12, pp. 1601–1613, 2008.

[17] K. Sakai, S. Huang, W. Ku, M. Sun, and X. Cheng, "Timer-based CDS construction in wireless ad hoc networks," *IEEE Trans. Mob. Comput. (TMC)*, vol. 10, no. 10, pp. 1388–1402, 2011.

[18] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proc. of ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2003.

[19] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay-tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 11, pp. 1576–1589, 2011.