Offloading Infrastructure using Delay Tolerant Networks and Assurance of Delivery

Christoph P. Mayer
Institute of Telematics
Karlsruhe Institute of Technology (KIT), Germany
mayer@kit.edu

Oliver P. Waldhorst
Integrated Communication Systems Group
Ilmenau University of Technology, Germany
oliver.waldhorst@tu-ilmenau.de

***Abstract*—Infrastructure-based mobile networks are becoming increasingly overloaded due to strong growing number of mobile devices like smartphones and their communication needs. Especially in urban cities the cost of maintaining and extending infrastructure is high due to increased geographical density of mobile devices. While this increasing density puts high load on infrastructure-based networks, it is an enabler for infrastructure-less networks like Delay Tolerant Networks that perform end-to-end routing through store-carry-forward.**

In this work we present a novel routing scheme for offloading traffic from infrastructure-based networks with the help of Delay Tolerant Networks. Messages are initially routed in the infrastructure-less network, and continuously switch over to infrastructure-based routing when the probability of successful delivery in the infrastructure-less network decreases. We analyze the scheme under different types of heterogeneity by varying the fraction of infrastructure-capable devices, fraction of DTN-capable devices, and message size. Our scheme allows to offload larger parts of traffic from infrastructure networks. For example, using a message time-to-live of 5 h our scheme can offload 36% of traffic from infrastructure networks with acceptable ad-hoc forwarding load and storage requirements on mobile devices.

***Index Terms*—Delay Tolerant Networks, Hybrid Networks, Offloading**

I. INTRODUCTION

The strong growing number of mobile devices—7.1 billion expected 2015—results in an expected 26-fold increase of traffic volume in infrastructure-based mobile networks between 2010 and 2015 [1]. While already today cellular mobile networks are overloaded [2] and costly [3], this trend is expected to continue [1]. On the other hand, the growing number of mobile devices results in increased density in the physical world. This can be exploited to deploy *infrastructure-less* networks that provide end-to-end routing through opportunistic ad-hoc communication. Such *Delay Tolerant Networks* (DTN) [4] perform *store-carry-forward* routing to deliver messages in an end-to-end fashion, although a continuous end-to-end communication path may never exist between sender and destination device. The integration of infrastructure-based networks and Delay Tolerant Networks has shown beneficial as it can boost DTN routing performance [5], [6], and offload traffic from congested infrastructure networks [7], [8], [9].

In this paper we present a novel routing scheme for infrastructure offloading that supports end-to-end unicast communication between mobile devices, in contrast to content dissemination used in related works [7], [8], [9]. Our scheme is decentralized and autonomous, and does not require a central coordination unit. In [10] we presented the *Hybrid Routing System* (HRS) that can integrate infrastructure-based and infrastructure-less networks transparently and autonomously. We use HRS in this work as enabling platform to implement our offloading scheme by preferably routing messages in the DTN. The routing strategy in our offloading scheme is determined on a per-message basis, based on information about the network, and a message's remaining *Time-To-Live* (TTL). Our scheme is applicable for different types of heterogeneity in device capabilities: devices with/without capability to access infrastructure, and devices with/without capability to perform ad-hoc communication. Applications our scheme addresses are unicast user-to-user communication like email, photo/video, or status updates.

The general idea is based on *assurance of successful delivery through the infrastructure*: The higher the probability that a message can be delivered through the infrastructure in case delivery in the DTN fails, the longer DTN routing is performed to deliver the message without generating infrastructure load. In contrast to most DTN protocols that route multiple copies of the same message to increase chances of successful delivery, only a single message copy is routed in our scheme. This reliefs from the need of an acknowledgment system used to detect whether a message has been successfully delivered in the DTN, before a replica is sent through the infrastructure. In this paper we provide the following contributions:

- A novel infrastructure offloading scheme based on assurance of delivery.
- Integration of heterogeneous infrastructure-capabilities, and heterogeneous ad-hoc capabilities of mobile devices.
- Unicast end-to-end routing between mobile devices.
- Extensive evaluation under varying infrastructure-capability, ad-hoc capability, and message size.

In Section II presents related work for offloading infrastructure networks. Section III briefly reviews the Hybrid Routing System [10] that is used as enabling platform for our offloading scheme. The actual offloading scheme is presented in Section IV, and evaluated in Section V. Finally, Section VI summarizes, concludes, and looks at future work.

TABLE I: Overview of offloading schemes.

Work	Message direction	Goal	Problem	Mobile device capabilities
MADNet [7], [8]	Infra. → mobile device	Reduce infra. load	Target-set selection	All infra.-access, all DTN
Push-and-Track [9]	Infra. → mobile device	Reduce infra. load	Target-set, replication, timing	All infra.-access, all DTN
Lee et al. [11]	Infra. ↔ mobile device	Prefer WiFi over cellular	Estim. WiFi avail.	All WiFi and cellular capable
Wiffler [12]	Infra. ↔ mobile device	Prefer WiFi over cellular	Estim. WiFi avail., fast switch	All WiFi and cellular capable
This work	Mobile device ↔ mobile device	Offload infra., prefer DTN	DTN routing decision	Heterog. infra.-access/ad-hoc

II. RELATED WORK

Table I shows related works on infrastructure offloading. Our work differs in two perspectives: First, we look at unicast communication between mobile devices while related work focuses on scenarios of content dissemination or communication between mobile device and infrastructure. Second, our scheme supports heterogeneous infrastructure- and heterogeneous ad-hoc-capabilities, in contrast to related work that assumes all devices infrastructure- and ad-hoc-capable.

Han et al. present *MADNet* [7], [8], an infrastructure offloading scheme based on the target-set selection problem. Given a set of mobile devices reachable through the infrastructure-based network, a subset of devices is selected and content spread to those devices. Mobile devices in the target-set disseminate content further through opportunistic ad-hoc communication to devices not in the target-set. The authors explore three different strategies for target-set selection: greedy, heuristic, and random. Using an information dissemination function, the benefit of adding a mobile device to the target-set is evaluated on a per-device basis for the greedy strategy, based on how active a mobile device is. The heuristic strategy uses the device's regularity in mobility for future estimations, and the random strategy selects the target-set at random. Using trace-driven evaluation the authors show that the greedy strategy works best for offloading infrastructure.

Similar to Han et al., Whitbeck et al. [9] propose an infrastructure offloading scheme named *Push-and-Track*. For disseminating content, Push-and-Track determines how many copies should be disseminated, to which devices the copies should be disseminated, and at what times the copies should be disseminated. Upon receiving content from the infrastructure, mobile devices disseminate content opportunistically in an epidemic fashion and send acknowledgments back to infrastructure upon content receipt. Based on the elapsed TTL of content, an infection-rate objective function is used to determine the number of copies to push from the infrastructure. A similar function is used in our work, we however use the elapsed TTL to determine the routing strategy, not for replication decisions.

Lee et al. [11] analyze the relation between WiFi access and cellular access used by mobile devices like smartphones. Based on real-world experiments, they find that due to automatic switching from cellular to WiFi access in today's smartphones, already a larger fraction of traffic is offloaded from cellular networks to WiFi-based networks. They analyze how much traffic can be further offloaded by accepting a certain delay, by waiting for WiFi access to become available. In [12] Balasubramanian et al. investigate a similar direction for offloading data by intentionally waiting for WiFi access to appear. They present *Wiffler* that implements strategies for intentionally delaying data transfers, and uses low-level link information to quickly switch between 3G and WiFi.

III. HYBRID ROUTING SYSTEM

In [10] we presented the *Hybrid Routing System* (HRS) that is used in this work as enabling platform. HRS integrates infrastructure-based and infrastructure-less networks by an announcement system built up by devices in the infrastructure-based part of the network. Infrastructure-connected mobile devices announce their ability to route in the DTN towards other devices, which we call their *awareness* for other devices. Two distributed announcement systems are described in [10], based upon structured key-based routing overlay networks. Addressing is based on flat identifiers, with *same* identifiers being used for overlay routing, *and* for DTN routing.

A. Destination/Infrastructure Awareness

In HRS every device d_i locally manages a table of awareness for other devices $\mathcal{T}_i = \{(id_j, p_i(d_j)), (id_k, p_i(d_k)), \ldots\}$. An entry $(id_j, p_i(d_j))$ describes device d_i's awareness for device d_j, and therewith its applicability to route in the DTN towards device d_j with identifier id_j. Awareness is described as $p_i(d_j) \in [0, 1]$ and managed by the actually integrated DTN protocol—we e. g. use PRoPHET [13] in this work for evaluation. A higher value $p_i(d_j)$ depicts a higher awareness and therefore better applicability of d_i to route towards d_j.

HRS can integrate devices with heterogeneous infrastructure -access capabilities—i. e. devices with, and devices without infrastructure-access—into hybrid networks. For tracking infrastructure-access, a virtual device $\mathcal{I}$ is integrated and managed like device awareness by the underlying DTN protocol; i. e. $p_i(id_\mathcal{I})$ describes how applicable d_i is for routing messages towards infrastructure. The value $p_i(id_\mathcal{I})$ is managed by the device itself.

B. Distributed Announcement System

Infrastructure-capable devices announce a subset of $\mathcal{T}_i$ in the distributed announcement system. The announced subset is selected from entries with highest awareness. Other devices connected to the infrastructure can query the announcement system for devices with high awareness for a given device in a distributed fashion. The announcement system is implemented using a key-based routing overlay that additionally allows infrastructure-connected devices to directly exchange messages. For details we refer to [10].

C. Mixed DTN Routing Metric

Routing in the DTN is based on a mixed decision metric, made up of awareness for a message's destination device, and awareness for infrastructure-access. A message m_j with destination device d_j currently stored by device d_s is forwarded/replicated from d_s to d_t in communication range, if

$$\left(\alpha \cdot p_t(id_{\mathcal{I}}) + (1-\alpha) \cdot p_t(d_j)\right) > \left(\alpha \cdot p_s(id_{\mathcal{I}}) + (1-\alpha) \cdot p_s(d_j)\right). \quad (1)$$

Parameter $\alpha \in [0,1]$ trades off awareness for infrastructure-access (used for routing towards the infrastructure-based announcement system) against awareness for the destination device (used for routing in the DTN). After traversing the infrastructure-based announcement system, messages are *marked*. Marked messages are in the remainder of this paper *always* routed using $\alpha = 0$ to prevent routing towards the infrastructure a second time.

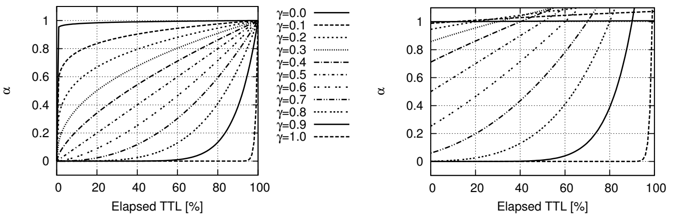
IV. OFFLOADING SCHEME

Our main idea for offloading infrastructure is to initially try to deliver messages in the DTN, and switch to infrastructure-based communication if probability of message delivery in the DTN becomes unlikely. This strategy and the switchover time are implemented through Equation 1 using an $\alpha(\cdot)$ function—in contrast to a static α value—that is evaluated in every DTN routing decision. Initial routing focuses upon awareness for the destination device using DTN routing. If the scheme assumes the message will not be delivered in the DTN within its lifetime, more focus is put onto routing towards infrastructure-capable devices with the goal of delivering the message through infrastructure.

In our scenarios, messages are *generated* by mobile devices, and messages are *destined* for mobile devices. We consider two types of heterogeneity in the following, however only one type at a time: First, devices with/without capability of infrastructure-access, and second, devices with/without capability of ad-hoc communication. Mix of devices is limited to one kind of heterogeneity: If the fraction of infrastructure-capable devices is varied, all devices are assumed capable of ad-hoc communication. If the fraction of devices being ad-hoc-capable is varied, all devices are assumed infrastructure-capable. The fraction of devices being infrastructure-capable is denoted $\gamma \in [0,1]$, the fraction of devices being ad-hoc capable is denoted $\delta \in [0,1]$.

In its simplest form where all devices are infrastructure-capable ($\gamma = 1$), a message is routed in the DTN until its TTL "almost" elapsed, and is then delivered through the infrastructure. How well delivery in the DTN works depends—besides mobility and the DTN protocol—on δ. If initial delivery over the DTN was not successful, $\gamma = 1$ *assures* that delivery through the infrastructure will be successful, as all devices are infrastructure-capable.

Let $t_i(m)$ be the initial TTL of message m, and $t_e(m)$ the message's lifetime, i. e. $t_e(m)/t_i(m) \in [0,1]$ is the elapsed fraction of lifetime. A message has expired if $t_e(m) > t_i(m)$. Let $\gamma \in [0,1]$ be the fraction of infrastructure-capable devices in the network, i. e. $\gamma = 1$ in the prior example. Let $\delta \in [0,1]$ be the fraction of ad-hoc capable devices. For instance, the offloading strategy for $\gamma = 1$ is as follows: Use $\alpha = 0$ if $t_e(m)/t_i(m) < 1$ to route the message in the DTN. The message is sent through the infrastructure if its TTL has "almost" elapsed, i. e. $t_e(m)/t_i(m) = 1$. In this example of $\gamma = 1$ the used $\alpha(\cdot)$ function is then a step function with $\alpha(\cdot) = 0$ if $t_e(m)/t_i(m) < 1$, and $\alpha(\cdot) = 1$ if $t_e(m)/t_i(m) \geq 1$.



(a) Original $\alpha(\cdot)$ function, Equation 2.

(b) Final $\alpha(\cdot)$ function with γ-based shift, Equation 3.

Fig. 1: Dynamic $\alpha(\cdot)$ function.

Depending on γ, the $\alpha(\cdot)$ function requires different behavior, as will be described in Section IV-B. Generally, $\alpha(\cdot)$ determines:

- The routing "direction" for the mixed decision metric described in Section III-C.
- Whether a message should be sent through the infrastructure if it reaches an infrastructure-capable device.

A message is sent through the infrastructure if $\alpha(\cdot) \geq 1$, and routed in the DTN if $\alpha(\cdot) < 1$. If it is routed in the DTN, the value of $\alpha(\cdot)$ is used for the mixed DTN metric in Equation 1 to determine the routing "direction". The value of δ does not influence the routing strategy, it however impacts the probability that a message can be successfully delivered in the DTN.

A. Sampling Infrastructure Capabilities

The value of γ is sampled locally by each device upon contact with other devices. While the locally sampled and estimated value $\hat{\gamma}$ does not necessarily reflect the global network view, it represents the local view of the mobile device. As routing decisions are performed locally by each device, the local view is of actual interest. Different mechanisms for sampling γ are possible, e. g. simple averaging over time, or weighted moving average. We show in Section V-A that the value of γ can be sampled by devices quickly within hours using simple averaging, even without taking transitive propagation into account.

B. Assurance of Delivery

The value of γ impacts the *assurance* that a message can be routed to, and delivered through the infrastructure. If $\gamma = 1$ and all devices are infrastructure-capable, at any time a message can be sent through the infrastructure to the destination device. If $\gamma < 1$, additional time is required to route a message in the DTN towards a device that is infrastructure-capable, and after the infrastructure transfer towards the final destination device through DTN. We use the following restrictions to reduce complexity: $\gamma = 1$ if $\delta < 1$, and $\delta = 1$ if $\gamma < 1$.

Figure 1a shows[1] the $\alpha(\cdot)$ function, defined as

$$\alpha\big(t_i(m), t_e(m), \gamma\big) = \left(\frac{t_e(m)}{t_i(m)}\right)^{\gamma/(1-\gamma)}. \qquad (2)$$

for $\gamma \in [0, 1)$. For $\gamma = 1$ we define $\alpha(\cdot) = 0$ if $t_e(m) < t_i(m)$, and $\alpha(\cdot) = 1$ if $t_e(m) \geq t_i(m)$. The x-axis of Figure 1a describes the fraction of elapsed TTL, i. e. $t_e(m)/t_i(m)$. The $\alpha(\cdot)$ function results in higher preference on infrastructure if message lifetime becomes older, and even stronger increase under decreasing fraction of infrastructure capable devices γ. Note that offloading infrastructure makes only sense with a large fraction of infrastructure-capable devices, i. e. $\gamma \to 1$. If only a small fraction of devices are infrastructure-capable, the goal would be to provide communication between the devices, irrespective of load.

Equation 2 does not yet account for time required to route messages *towards* infrastructure-capable devices. This required times directly depends on γ, therefore we shift the x-axis by γ and get

$$\alpha\big(t_i(m), t_e(m), \gamma\big) = \left(\frac{t_e(m)}{t_i(m)} + (1-\gamma)\right)^{\gamma/(1-\gamma)}. \qquad (3)$$

In case $\gamma = 1$ we do not need to account additional time, in case of a small value γ we need to account more time and have a stronger shift. Figure 1b shows the resulting function where a small value of γ results in $\alpha(\cdot) \geq 1$ at earlier points. This way, a message is routed towards infrastructure-capable devices earlier, and sent through the infrastructure more quickly.

The term $+(1-\gamma)$ in Equation 3 accounts for additional time required to route *towards* infrastructure-capable devices. Sending of messages *through* the infrastructure, however, takes additional time in the order of few seconds or minutes—depending on bandwidth and message size. To ensure that during this last routing step a message's lifetime does not expire, the elapsed lifetime used for Equation 3 must be increased artificially.

C. Protocol

Algorithm 1 shows pseudo-code run locally on each device to implement the offloading scheme. If the device is not ad-hoc capable, routing in the DTN is not possible and the message m sent directly through the infrastructure, by settings $\alpha(\cdot) = 1$. If infrastructure-access is available, and for m it holds $\alpha(t_i(m), t_e(m), \gamma) \geq 1$, the message m is sent through the infrastructure. If $\alpha(\cdot)$ evaluates to < 1, the message is forwarded/replicated through DTN routing; initially towards devices that are good forwarders to reach the message destination device directly in the DTN, and, if $\alpha(\cdot)$ approaches a value of 1, towards infrastructure-capable devices.

V. EVALUATION

We perform simulative evaluation to analyze the offloading capabilities of our proposed scheme using the ONE [16] simulator with custom extensions. Table II gives an overview of simulation parameters. As DTN routing protocol we use PRoPHET [13], [15] by Lindgren et al. due to its maturity[2]. We extend PRoPHET for single-copy routing in that a message is not *replicated* to a device with higher applicability for the destination, but rather *forwarded*. Every scenario is simulated with at least 30 statistically independent seeds. Figures show mean values with 95% confidence intervals.

Algorithm 1: Offloading algorithm run by every device.

```
local device $d_i$;
local message queue $q_i$;
infrastructure awareness $p_i(id_{\mathcal{I}})$;
sampled infrastructure fraction $\gamma$;
while running do
    for $m \in q_i$ do
        if ad-hoc capable then
            $\alpha = \alpha(t_i(m), t_e(m), \gamma)$;
            $\beta = 1 - \alpha$;
        else
            $\alpha = 1$;
            $\beta = 0$;
        if $\alpha \geq 1$ and infrastructure connectivity then
            send $m$ through infrastructure;
            continue;
        if $\alpha < 1$ and in ad-hoc range with $d_j$ then
            message $m$ destination is $d_t$;
            local quality = $\alpha \cdot p_i(id_{\mathcal{I}}) + \beta \cdot p_i(d_t)$;
            $d_j$'s quality = $\alpha \cdot p_j(id_{\mathcal{I}}) + \beta \cdot p_j(d_t)$;
            if $d_j$'s quality > local quality then
                forward/replicate $m$ to $d_j$;
```

TABLE II: Evaluation parameters.

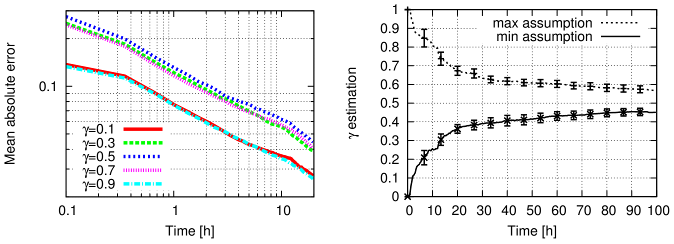
Category	Value
Mobile devices	100
Movement speed	1–3 m/s
Ad-hoc comm. range	15 m
Bandwidth	2.1 MBit/s (Bluetooth v2 EDR)
Mobility model	SWIM [14], wait-time slope 1.45, cutoff 12 h
Message generation	every 15 min–20 min per device process
Message destination	selected uniformly at random over all devices
Message TTL	1 h, 3 h, 5h, 10 h
Message size	250 kByte in Section V-B, Section V-C 250 kByte up to 8 MByte in Section V-D
infrastructure-capable	γ between 0%–100% of mobile devices
Ad-hoc capable	δ between 0%–100% of mobile devices
Capability restriction	if $\gamma \in [0,1) \to \delta = 1$, if $\delta \in [0,1) \to \gamma = 1$
DTN routing protocol	PRoPHET [13], [15]
Playground	map of Karlsruhe, Germany, 2×2 km
Seeds per scenario	30
Simulation initialization	7 days
Simulation reporting	1 day

First, the efficiency of sampling γ is evaluated in Section V-A. Cost and performance under heterogeneous infrastructure-capability γ are evaluated in Section V-B. Scenarios with heterogeneity in ad-hoc capability δ are evaluated in Section V-C. Finally, we evaluate the impact of message size in Section V-D, as it impacts DTN routing due to finite contact durations.

---

[1] Figure 1a shows asymptotic behavior for $\gamma = 0$ and $\gamma = 1$, actually, the behavior is a non-continuous step function.

[2] PRoPHET has been analyzed extensively and is available as IRTF draft.



(a) Mean absolute error of $\hat{\gamma}$ over time (log-log axis).

(b) Worst min/max estimation over time for $\gamma = 0.5$.

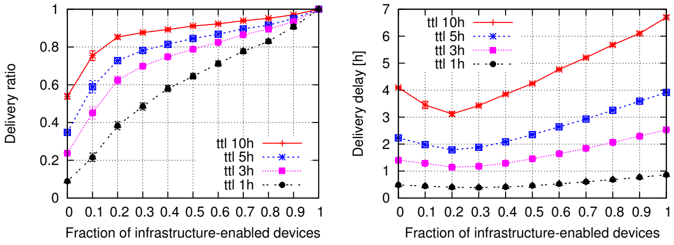
Fig. 2: Behavior of $\hat{\gamma}$ estimation.

A. Sampling of γ

Devices sample the network heterogeneity in infrastructure-capability γ, as described in Section IV-A. For evaluation of estimation speed and accuracy, a simple averaging mechanism is run on each device. Initially devices assume γ equaling their own infrastructure-capability—i. e. $\gamma = 0$ if the device is not infrastructure-capable, $\gamma = 1$ if the device is infrastructure-capable—and average over infrastructure capabilities of encountered devices in the DTN. Figure 2a shows the mean absolute error $(\sum_{d_i} |\hat{\gamma}_{d_i} - \gamma|)/\#devices$ over all locally estimated $\hat{\gamma}$ value over time for networks with different heterogeneity γ. This error falls quickly within few hours, i. e. $\hat{\gamma}$ converges to γ. We believe that an estimation error of < 0.1 is acceptable for our scheme. Worst estimation is for $\gamma = 0.5$ where heterogeneity is highest. Figure 2b shows the worst-case estimation at every point in time for a scenario with $\gamma = 0.5$. While convergence is slowest for $\gamma = 0.5$, still a worst-case error < 0.1 is reached after ≈40 h.

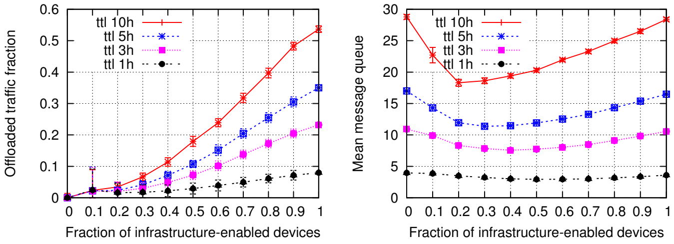
B. Heterogeneity in Infrastructure Capability

We evaluate offloading when all devices are ad-hoc capable ($\delta = 1$) under varying fraction of infrastructure-capable devices ($\gamma \in [0, 1]$). Figure 3a shows the delivery ratio, depending on the fraction of infrastructure-capable devices on the x-axis. Of interest are only scenarios with γ close to 1 where a large fraction or all mobile devices are infrastructure-capable, otherwise the goal of offloading is not appropriate. In case $\gamma = 1$, all mobile devices can communicate through the infrastructure, and all messages are delivered successfully. In this case, DTN routing is performed most aggressive until the TTL almost elapsed, and $\alpha(\cdot)$ becomes a 0/1 step function. The initial slope in delivery ratio results from high utility in infrastructure support, and can be observed for other metrics discussed in the following around the same point of γ on the x-axis. Delay resulting from DTN routing is shown in Figure 3b. Note, that only successfully delivered messages are taken into account for calculating mean delay. At $\gamma = 1$, the mean delay is about 70%–90% of the defined TTL bound. Note, that our scheme provides guaranteed delay bounds through the given TTL. Figure 3c shows the respective fraction of messages that were offloaded from infrastructure through delivery in the DTN. This fraction of offloaded messages is calculated as offloaded fraction, multiplied by γ. Normalization using γ is required, as increase of γ results in messages being successfully delivered that otherwise would not have been delivered, and therewith impact offloading negatively. Increasing TTL allows for higher delivery ratio in the DTN, and more messages offloaded from infrastructure. For example, using a TTL of 1 h results in 9% offloaded messages, TTL of 10 h results in 54% offloaded messages.



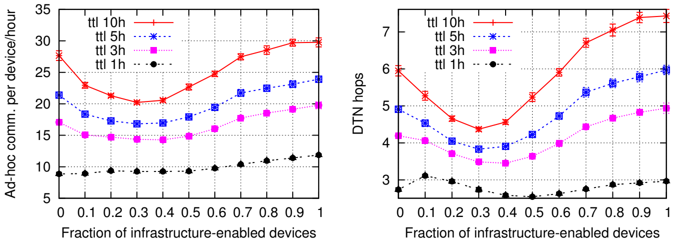
(a) Delivery ratio

(b) Delivery delay



(c) Offloaded fraction

(d) Message queue size



(e) Ad-hoc overhead

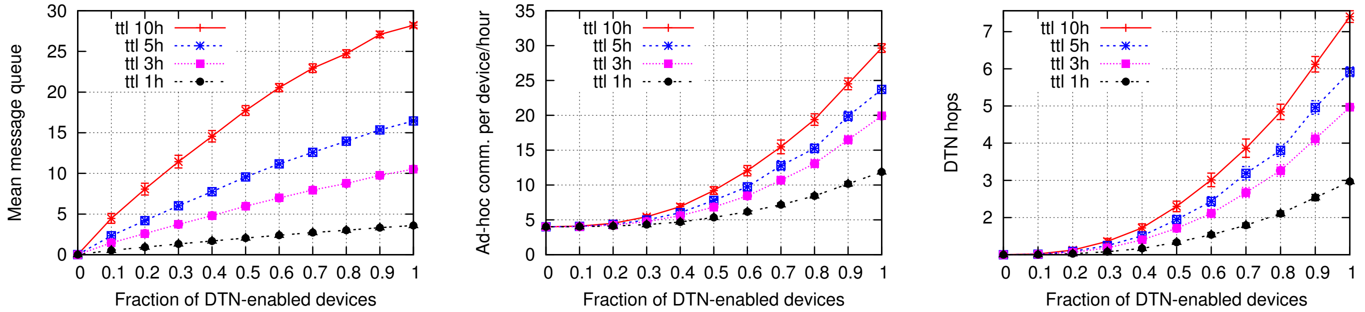
(f) DTN hops

Fig. 3: Performance/cost metrics for variation of infrastructure-access $\gamma \in [0, 1]$, all devices ad-hoc capable $\delta = 1$.

DTN routing requires local message storage. Resulting queue size—in number of messages—is shown in Figure 3d. The queue size is relatively stable and correlates with delivery delay in Figure 3b. Note that queue size strongly depends on the frequency of message generation which is in average 4 messages per device per hour in our evaluation. Figure 3e shows the ad-hoc communication overhead per mobile device per hour, defined as number of ad-hoc message transfers. The number of traversed mobile devices in the DTN routing is shown in Figure 3f. Higher TTL results, as expected, in higher hop count as DTN routing has more time to deliver a message, and therefore forwards the message more often. Generally, cost metrics show decrease where delivery ratio shows steepest slope at the point where infrastructure support is most beneficial.

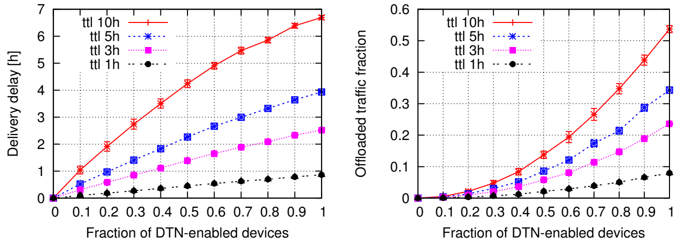
C. Heterogeneity in Ad-hoc Capability

Section V-B evaluated a scenario where all devices are capable of ad-hoc communication to forward messages in the



(a) Message queue size (b) Ad-hoc overhead (c) DTN hops

Fig. 5: Cost metrics for variation of ad-hoc capability $\delta \in [0, 1]$, all devices infrastructure-capable $\gamma = 1$.

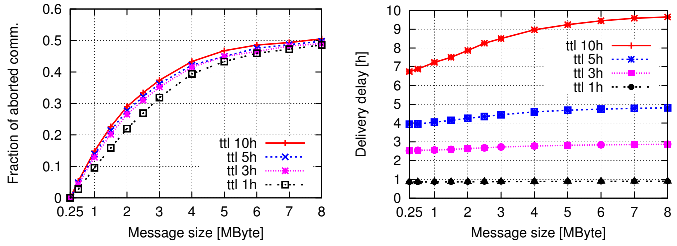


(a) Delivery delay (b) Offloaded fraction

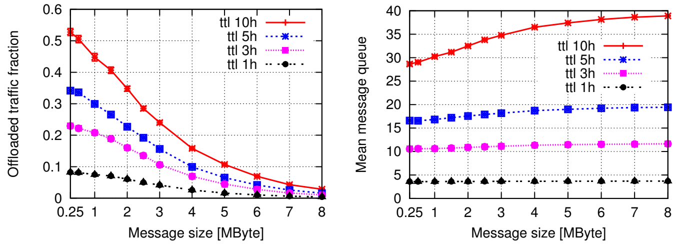
Fig. 4: Performance metrics for variation of ad-hoc capability $\delta \in [0, 1]$, all devices infrastructure-capable $\gamma = 1$.

DTN. In this section all devices are assumed infrastructure-capable ($\gamma = 1$), but only a fraction of devices ad-hoc capable ($\delta \in [0, 1]$). This can be seen as a scenario with uncooperative devices that are selfish, and do not take part in the DTN. The routing strategy is unchanged as it does not depend on δ. Rather, we analyze the impact of δ on offloading. As $\gamma = 1$, assurance of delivery through the infrastructure is complete. If a message was not delivered in the DTN within its lifetime, successful delivery through infrastructure is assured, resulting in delivery ratio of 1. Note, that messages generated by non ad-hoc capable devices are never routed in the DTN, as messages can only flow inside the DTN, and from DTN to infrastructure. Depending on fraction of ad-hoc capable devices, delay of successfully delivered messages increases, as shown in Figure 4a. Message delay is bound by TTL, and within 65%–80% of allowed TTL. Similar, message queues shown in Figure 5a, ad-hoc overhead per device in Figure 5b, and DTN hops of delivered messages in Figure 5c increase with δ on the x-axis.

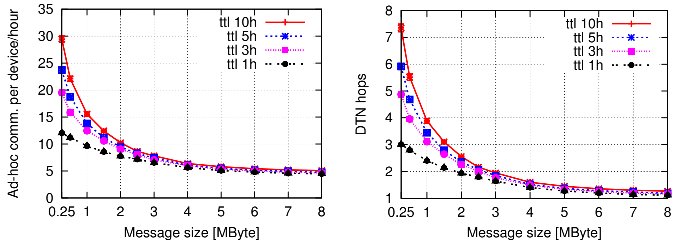
The fraction of offloaded messages is shown in Figure 4b. Offloading is performed through delivery in the DTN, therefore offloaded traffic increases with more ad-hoc capable devices. For $x = 1$ the same scenario as in Section V-B results, with $\gamma = 1$ and $\delta = 1$ where for our scenario the fraction of offloaded messages is 8% for TTL of 1 h, 24% for TTL of 3 h, 36% for TTL of 6h, and 54% for TTL of 10 h.



(a) Average ad-hoc abort rate (b) Delivery delay



(c) Offloaded fraction (d) Message queue size



(e) Ad-hoc overhead (f) DTN hops

Fig. 6: Performance/cost metrics for variation of message size, all devices ad-hoc and infrastructure-capable ($\delta = 1$, $\gamma = 1$).

D. Impact of Message Size

Contacts in opportunistic networks are predominantly short. Under a finite bandwidth model this results in a limited number of bytes that can be transferred between two mobile devices in mutual communication range. At worst, contacts are too short to transfer even a single message, obviously depending on message size. As our scheme relies on successful delivery of messages in the DTN, we are interested in how message size impacts the ad-hoc communication and reduces offloading effectiveness. Increasing message size introduces additional complexity in that not only a DTN path over multiple devices must be found, but parts of this path fail unpredictably. We do not use fragmentation of messages that can help cope with finite bandwidth [17], as we want to explore the limits of non-fragmented offloading.

We use Bluetooth v2 EDR with 2.1 MBit/s, resulting in transfer of ≈250 kByte per second. Message size is varied between 250 kByte up to 8 MByte, while prior simulations in this paper used constant message size of 250 kByte. Under the smallest message size of 250 kByte no ad-hoc message transfers fail, as 1 second is the smallest contact duration that is actually realized by devices in our simulation, i. e. under shorter contact duration all message transfers "fail" in a sense that the contact is not detected by devices.

Figure 6a shows the fraction of ad-hoc message transfers that fail, depending on message size. Increasing message size quickly increases the fraction of failed transfers. As, however, very long contact durations do exist, the fractions growth decreases. Delivery delay in Figure 6b increases as DTN forwarding becomes more complex as more paths fail arbitrarily. However, the increase is marginal for short TTL, and up to 40% for TTL of 10 h. As the DTN forwarding paths fail, the DTN performance decreases and fraction of offloaded traffic vanishes, shown in Figure 6c. Mean message queue behaves very similar to message delay, shown in Figure 6d. As parts of the DTN path fail, the ad-hoc communication in terms of successful transfers decreases, and similarly the number of DTN hops of successfully delivered messages decreases quickly, shown in Figure 6e and Figure 6f.

VI. Summary and Conclusion

In this paper we proposed an infrastructure offloading scheme for the use case of mobile-to-mobile unicast communication. Our main idea is based on assurance of message delivery, calculated using information about the network, and remaining message lifetime. Different types of heterogeneity are supported: heterogeneity in infrastructure-capability, and heterogeneity in ad-hoc-capability. Simulations show that our scheme can offload larger parts of traffic from infrastructure, for instance 36% of traffic under a TTL of 5 h in a scenario where all devices are ad-hoc and infrastructure-capable. Message delivery delay is guaranteed through TTL bound and—depending on γ—a guaranteed or probabilistic delivery ratio. Load introduced on mobile devices is acceptable, both in terms of message queue and ad-hoc communication overhead.

In case the DTN is not able to physically carry messages due to distance, the initial DTN routing before infrastructure delivery is unnecessary. Initial delay estimation would enables to decide more quickly whether DTN routing is feasible at all. We only looked at single-copy DTN routing in this work, as in case of multi-copy routing timely feedback is required whether a message was delivered through the DTN to not unnecessarily burden the infrastructure. We did not analyze energy consumption aspects of mobile devices which are interesting to consider. Further, we did not take the actual overload conditions in the infrastructure into account, which can make delivery in the DTN more probable.

Acknowledgment

This work is supported by the Landesstiftung Baden Württemberg gGmbH within the SpoVNet project, and the "Concept of the Future" of KIT within the framework of the German Excellence Initiative. We thank the anonymous reviewers for their valuable comments.

References

[1] Cisco Systems Inc., "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010-2015," http://newsroom.cisco.com/dlls/ekits/Cisco_VNI_Global_Mobile_Data_Traffic_Forecast_2010_2015.pdf, Feb. 2011.

[2] New York Times, "iPhones Overload AT&T's Network, Angering customer," http://www.nytimes.com/2009/09/03/technology/companies/03att.htmls, May 2009.

[3] AT&T, "AT&T Strengthens 3G Wireless Coverage in New York and New Jersey Areas," http://www.att.com/gen/press-room?pid=4800&cdvn=news&newsarticleid=27069, Sep. 2009.

[4] Z. Zhang, "Routing in Intermittently Connected Mobile Ad Hoc Networks and Delay Tolerant Networks: Overview and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 8, no. 1, pp. 24–37, 1st quarter 2006.

[5] P. Hui, A. F. Lindgren, and J. Crowcroft, "Empirical Evaluation of Hybrid Opportunistic Networks," in *Proceedings of International Conference on COMmunication Systems and NETworkS (COMSNETS)*, Bangalore, India, Jan. 2009, pp. 1–10.

[6] A. F. Lindgren, C. Diot, and J. Scott, "Impact of Communication Infrastructure on Forwarding in Pocket Switched Networks," in *Proceedings of ACM SIGCOMM Workshop on Challenged Networks (CHANTS)*, Pisa, Italy, Sep. 2006, pp. 261–268.

[7] B. Han, P. Hui, M. V. Marathe, G. Pei, A. Srinivasan, and A. Vullikanti, "Cellular Traffic Offloading Through Opportunistic Communications: A Case Study," in *Proceedings of International Workshop on Challenged Networks (CHANTS)*, Chicago, IL, USA, Sep. 2010, pp. 31–38.

[8] B. Han, P. Hui, and A. Srinivasan, "Mobile Data Offloading in Metropolitan Area Networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 14, no. 4, pp. 28–30, Oct. 2010.

[9] J. Whitbeck, Y. Lopez, J. Leguay, V. Conan, and M. D. de Amorim, "Relieving the Wireless Infrastructure: When Opportunistic Networks Meet Guaranteed Delays," in *Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Lucca, Italy, Jun. 2011.

[10] C. P. Mayer and O. P. Waldhorst, "Routing in Hybrid Delay Tolerant Networks," under submission, available at http://www.tm.kit.edu/~mayer/papers/tmc2011-hybriddtnrouting-submitted.pdf.

[11] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile Data Offloading: How Much Can WiFi Deliver?" in *Proceedings of International Conference on Emerging Networking EXperiments and Technologies (CoNEXT)*, Philadelphia, PA, USA, Dec. 2010, pp. 26:1–26:12.

[12] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting Mobile 3G Using WiFi," in *Proceedings of International Conference on Mobile Systems, Applications and Services (MobiSys)*, San Francisco, CA, USA, Jun. 2010, pp. 209–222.

[13] A. F. Lindgren, A. Doria, and O. Schelén, "Probabilistic Routing in Intermittently Connected Networks," in *Proceedings of Service Assurance with Partial and Intermittent Resources (SAPIR)*, ser. Lecture Notes in Computer Science, vol. 3126, Fortaleza, Brazil, Aug. 2004, pp. 239–254.

[14] A. Mei and J. Stefa, "SWIM: A Simple Model to Generate Small Mobile Worlds," in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2106–2113.

[15] A. F. Lindgren, A. Doria, and O. Schelén, "Probabilistic Routing in Intermittently Connected Networks," *ACM SIGMOBILE Mobile Computing Communications Review*, vol. 7, no. 3, pp. 19–20, Jul. 2003.

[16] A. Keränen, T. Kärkkäinen, and J. Ott, "Simulating Mobility and DTNs with the ONE," *Journal of Communications*, vol. 5, no. 2, pp. 92–105, Feb. 2010.

[17] M. Pitkänen, A. Keränen, and J. Ott, "Message Fragmentation in Opportunistic DTNs," in *Proceedings of WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC)*, Newport Beach, CA, USA, Jun. 2008, pp. 1–7.