This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

Querying Blobs in Vehicular Networks[1]

Ouri Wolfson Bo Xu

Department of Computer Science, University of Illinois at Chicago

ABSTRACT

In this paper we study querying binary large objects such as video and voice clips in a network of vehicles communicating wirelessly. We develop a set of query processing strategies and compare them along three dimensions, namely push versus pull, whether or not communication infrastructure is utilized, and whether metadata dissemination is separated from blob dissemination. We analyze these strategies theoretically and experimentally s in terms of answer throughput and communication overhead.

Keywords: vehicular ad-hoc networks, real-time traffic information, multimedia, blobs, query processing, mobile peer-to-peer, cellular communication.

1. INTRODUCTION

A vehicular ad-hoc network (VANET) is a set of vehicles (mobile peers) that communicate with each other via short-range wireless technologies such as WiFi and DSRC. An attractive ITS application of VANETs is sharing of binary large objects (blobs) such as voice and video clips for situation awareness and urban monitoring [16, 8, 3, 4]. For example, currently many taxi-cabs use cameras that capture continuous videos of the traffic ahead. Additionally, driver-recorded, time- and location-stamped audio clips (e.g. providing a short description of an accident) can also be captured. This information can be queried by other drivers to determine traffic-conditions and hazards outside their fields of view, and by first-response vehicles to facilitate law enforcement (e.g. allow police to track wanted cars). In fact, we have conducted field experiments in which videos are automatically captured by dashboard-mounted smart-phones, and are disseminated among vehicles in a peer-to-peer fashion. A 2-second, 116KB, sample video clip can be viewed at [15]. It shows the traffic

[1] This research was supported, in part, by the U.S. Department of Transportation National University Rail Center (NURAIL), Illinois Department of Transportation (METSI), National Science Foundation grants IIS-1213013, CCF-1216096, DGE-0549489.

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

condition of a segment of Ashland Ave. in Chicago. As the clip indicates, 2 seconds are sufficient to see whether or not the traffic is flowing, and at what speed.

Existing studies explored two paradigms of sharing blobs in VANETs, namely push (data-to-query) and pull (query-to-data). In the push paradigm, blobs are proactively disseminated [4]. In the pull paradigm, queries are proactively disseminated; blobs are disseminated as responses to received queries [16]. VANETs have also been studied as an augment to the cellular communication for cellular offloading purpose [9, 10, 12]. In this case, blob sources reside in a fixed network and need to be disseminated to a large number of mobile peers. Instead of every mobile peer downloading the blobs separately, only a small portion of the mobile peers download the blobs via the cellular communication, and they share the blobs with the other peers via the short-range communication. We refer to the VANETs in which the cellular communication is available as *hybrid vehicular networks*.

In this paper, we consider hybrid vehicular networks where blobs are generated by the mobile peers, e.g., a 2-3 seconds video of the surroundings is generated by a vehicle every minute. Thus, blob sources reside at mobile peers rather than in the fixed network. Queries are also generated by the mobile peers. For example, a vehicle may ask for video clips regarding the traffic condition one mile ahead of it. As in a typical VANET, a peer does not initially know the network-id's (i.e. cell-phone numbers) of the other peers in the network. However, a peer can communicate directly with other peers within its WiFi[2] transmission range without knowing their network-id. Furthermore, we do not require or assume a central server or any other form of directory/storage service in the fixed network.

[2] We use the term WiFi for simplicity, but the results of this paper apply to other short-range networking technologies such as DSRC.

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

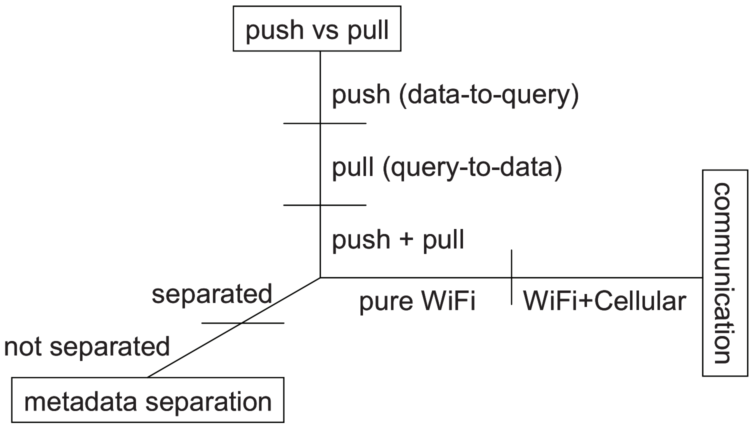


Figure 1. The design space for blobs query processing in hybrid vehicular networks

An environment as described above renders a broad spectrum of possible query processing strategies, along the following three design dimensions. First, the peer-to-peer communication may use purely WiFi or it may use both WiFi and cellular communication (i.e., hybrid). Pure cellular communication is not an option because it requires knowledge of network id of the receiver, whereas a query originator initially only knows the description of the requested blobs but not the network-id of the peers containing these blobs. Second, as aforementioned, the query processing may adopt push or pull or the combination of the two. Third, observe that a blob can be described by a much shorter *metadata* tag (e.g., time and location at which a multimedia clip was produced). The match (yes or no) between a query and a blob can be determined solely based on the metadata. Due to size-differences, the metadata and content of a given blob may be disseminated independently, and by different means (WiFi or cellular). The three design dimensions are illustrated by Figure 1.

In this paper we **define a paradigm** called (WiFi-communication, Match, Communication) or **WiMaC**, **of query processing strategies** based on the above design dimensions. The WiMac paradigm generates a list of 13 possible query processing strategies. Then we define the notion of <u>dominance</u> between query processing strategies. Intuitively, strategy A dominates strategy B when each query returns in A a superset of the set of answers it returns in B, each with a response time that is not higher in A than in B; additionally, the communication cost of A is not higher than that of B. We conduct analysis based on this definition and determine that 4 of the 13 WiMac strategies dominate the others. Finally we **compare these four strategies by simulation** in the context of querying vehicle-captured multimedia traffic information.

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

2. RELEVANT WORK

It has been demonstrated that in a static environment the combination of push and pull is superior to pure push and pure pull (see e.g., [7]). In this case, the dissemination of blobs/queries follows a geometric structure, such as line segments or trees. Such structure based methods do not work in VANETs due to mobility and disconnections.

Many methods have been proposed for resource discovery and data dissemination in MANETs/VANETs (see e.g., [3, 8]). In contrast to the present paper, these methods do not utilize cellular communication. However they provide the insight, used in our paper, that WiFi communication in the WiMaC strategies can be improved by cooperative caching and prioritization.

The authors in [18] propose a protocol with which parking reports are exchanged in a vehicle-to-vehicle fashion. They consider short reports giving the locations of available parking slots, whereas here we address the exchange of much larger messages containing blobs.

The authors in [16] propose a vehicle-to-vehicle live video streaming architecture called V3. The architecture adopts a query-to-data paradigm for query processing and it uses only WiFi communication. Thus the query processing method in V3 is similar to the (Q)-WiFi strategy which will be discussed in this paper. However, the work in [16] does not compare the query-to-data paradigm with other WiFi-only strategies and with WiFi-cellular strategies.

The authors in [2, 17] study reliable multimedia delivery in VANET's. The problem they focus on is how to encode the frames of a video clip communicated via WiFi, so that the receiver can recover from packet losses. Their methods provide the insight that the reliability of multimedia transmission in the WiMaC system can be improved by encoding and error correction.

3. THE MODEL

3.1 The Environment

The environment is a *system* consisting of a set of *mobile peers*, or *peers*. This set of peers may change over time. Each peer (e.g. a vehicle) is equipped with the following capabilities: (i) producing a blob (binary large

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

object) data such as video, voice, or multimedia clips; (ii) short-range wireless communication such as WiFi; and (iii) infrastructure based communication such as 3G cellular. Via the infrastructure, a peer is able to transmit messages to another peer by MMS (Multimedia Message Service) or TCP/IP communication. This is referred to as the *cellular-channel* or *cellular* communication. Each peer has a *network-id* that is used as its address for cellular communication, and this id is required in order to send a message to the peer via the cellular channel. The network-id can be a cell-phone number or an IP address. In addition, peers can communicate via the WiFi channel if they are within transmission range. Knowledge of the network id is not necessary for this purpose. Cellular and WiFi communication can be anonymized by decoupling the network-id of a peer from personal information. But a detailed privacy and security analysis is an orthogonal issue left for future work.

3.2 Reports and Reports Databases

Each peer periodically produces blob reports. Formally, a *blob report* R, is a couple *<Meta(R), Blob(R)>*, where *Meta(R)* and *Blob(R)* are the *metadata* and *blob* sub-reports, respectively. The metadata sub-report contains attributes describing the blob such as Time when the blob was produced, the Location at which it was produced, the Network-id of the producing peer, etc. *Blob(R)* is the blob itself, e.g., the music or video file.

A peer also produces queries that are stored and disseminated in the form of reports called *query-reports*. A query requests both sub-reports of each satisfying blob-report, but it refers only to the metadata of the blob-report. Thus, whether or not there is a match between a query-report and a blob-report can be determined solely based on the query and the metadata of the blob. For example, a query requests a song by its title, thus the match between the query and the song can be determined solely based on the metadata of the song-report, but the query asks for both the metadata and the song itself to be returned.

A peer is called the *producer* of the query- and blob-reports that it produces. Each query and each metadata sub-report contains the network-id of the producer of the report. Each query Q has a *query expiration time*, which indicates that the processing of Q stops when its expiration time is reached. The query expiration time is attached to Q when Q is created and remains fixed during the dissemination of Q. Similarly, each blob report has a fixed *blob expiration time*, beyond which it becomes invalid. In this sense query processing in a

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

VANET is “best effort”, since delivery of all the answers is not guaranteed. Each peer maintains a reports database that stores the metadata sub-reports, blob reports, and query reports produced by the peer or received from other peers. To deal with the storage limit, reports relations are managed by a cooperative-caching method such as the one introduced in [11].

4. The WiMaC Query Processing Strategies

As mentioned above, some or all reports that satisfy a query *Q* may reside on peers that are different than the query producer, *Qp*. Since *Qp* does not normally have the network id of such peers, and does not even know how many reports satisfy the query, all query processing strategies start with a stage of WiFi dissemination to neighboring peers. The dissemination may be of the query, the blob reports, the metadata sub-reports, or some combination. Blob reports and metadata sub-reports propagate differently by WiFi; the metadata report propagates faster, thus possibly meets more queries within a given time period. When a match is found, it may be followed by a second stage of additional cellular or WiFi communication. For example, assume that the match is between a query and a metadata report, and that the blob sub-report is located at another peer. Then the blob has to be transferred to the query producer by additional communication. Thus, this is the (WiFi-communication, Match, Communication) paradigm, called *WiMaC*, and all query processing strategies discussed in this paper are special cases of WiMaC.

4.1 Strategies Design-Space

The structure of the design space is depicted in Table 1. There are 13 WiMaC strategies. Each strategy is denoted as follows. The denotation consists of the strategy number as defined in Table 1 and the strategy name. The strategy name is formed as follows. If there is not the second stage, then the strategy is named by the first stage, i.e., (blob). If there is a second stage, then the strategy is named by the two stages connected by a “-“. For example, 2b (meta)-cell denotes the 2b strategy which disseminates metadata sub-reports in the first stage and uses cellular communication in the second.

Table 1. Design space of the WiMaC paradigm

Notation:

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

blob = blob-report, meta = metadata sub-report, Q = query, cell = cellular.

Strategy Names: 1 is (blob), 2a is (meta)-WiFi, 3b is (Q)-cell, etc.

Strategy No.	Type of reports disseminated in the first stage (always via WiFi)	Communication medium in the second stage
1	(blob)	No second stage
2a	(meta)	WiFi
2b	(meta)	cell
3a	(Q)	WiFi
3b	(Q)	cell
4a	(blob, meta)	WiFi
4b	(blob, meta)	cell
5a	(blob, Q)	WiFi
5b	(blob, Q)	cell
6a	(meta, Q)	WiFi
6b	(meta, Q)	cell
7a	(blob, meta, Q)	WiFi
7b	(blob, meta, Q)	cell

4.1.1 WiFi-only Strategies

1 (blob): In the first stage of WiMaC blob reports are disseminated via WiFi. Queries are kept at the producer peer, and a match occurs when a disseminated blob report arrives at a matching query. There is no second stage. The 1 (blob) strategy corresponds to the push (data-to-query) paradigm.

2a (meta)-WiFi: In the first stage of the WiMaC paradigm metadata sub-reports are disseminated via WiFi. When a metadata sub-report B reaches a matching query Q producer, the producer peer disseminates Q

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

via WiFi. When the B(lob)-producer receives Q, the B-producer disseminates the blob report via WiFi, to reach the Q-producer and provide an answer to Q.

3a (Q)-WiFi: In the first stage of WiMaC queries are disseminated via WiFi. When a query Q reaches the producer peer of a matching blob report B, the B-producer disseminates the blob report via WiFi to reach the Q-producer. The 3a (Q)-WiFi corresponds to the pull (query-to-data) paradigm.

4a (blob,meta)-WiFi: In the first stage of WiMaC metadata and blob reports are disseminated separately via WiFi. If the producer of a matching query Q receives a blob report, then there is no second stage. If the producer of a matching query Q receives a metadata sub-report, the Q-producer disseminates Q via WiFi. When a peer Z that has a matching blob report B receives Q, Z disseminates B via WiFi to reach the Q-producer.

5a (blob,Q)-WiFi: In the first stage of WiMaC blob and query reports are disseminated via WiFi. When a blob report and a matching query Q collocate at a peer Z, Z disseminates the blob report via WiFi to reach the query producer.

6a (meta,Q)-WiFi: Metadata sub-reports and queries are disseminated via WiFi. When a metadata sub-report B and a matching query Q collocate at a peer Z, Z disseminates Q via WiFi. When the B-producer receives Q, the B-producer disseminates the corresponding blob report via WiFi to reach the Q-producer.

7a (blob,meta,Q)-WiFi: This strategy is a combination of 4a (blob,meta)-WiFi and 6a (meta,Q)-WiFi.

4.1.2 WiFi-cellular Strategies

In the WiFi-cellular strategies, after a match is discovered, the answer B is communicated from a peer P to the query producer via the cellular channel. However, P first inquires via the cellular channel whether the producer has already received B (from other peers); if so, the transmission of B is suppressed.

2b (meta)-cell: In the first stage of WiMaC metadata sub-reports are disseminated via WiFi. When a metadata sub-report B reaches a matching query Q producer, the Q-producer peer sends Q to the B(lob)-producer via the cellular channel. In response, the B-producer sends B and all the other matching blob reports that it has to the Q-producer, via the cellular channel.

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

3b (Q)-cell: In the first stage of WiMaC queries are disseminated via WiFi. When a query Q reaches a matching blob report B, the B-producer sends B to the Q-producer via the cellular channel.

4b (blob,meta)-cell: This strategy is identical to (meta)-cell (2b), with the following addition. Blob reports are disseminated as well in the first stage of WiMaC. There is no second stage if the Q-producer receives a blob report from the WiFi dissemination.

5b (blob,Q)-cell: Blob and query reports are disseminated via WiFi in the first stage. When a blob report B and a matching query Q collocate at a peer Z, Z sends B to the Q-producer via the cellular channel.

6b (meta,Q)-cell: Metadata and query reports are disseminated via WiFi in the first stage. When a metadata sub-report B and matching query Q collocate at a peer Z, Z sends Q to the B-producer via the cellular channel. In response, the B-producer sends the blob report B and all the other matching blob reports that it has to the Q-producer, via the cellular channel.

7b (blob,meta,Q)-cell: This strategy is a combination of 5b (blob,Q)-cell and 6b (meta,Q)-cell.

4.2 Strategy Dominance Analysis

4.2.1 Definitions and Assumptions

Let a peer receive an answer blob report at time t. The *response-time* of the answer is the length of the time period since the answer is produced until t.

We say that a strategy X is *dominated* by another strategy Y if the following 4 conditions are satisfied for every blob report B:

(1) For every query that B answers, if the answer is received in Y, it is also received in X;

(2) For every query that B answers, its response-time in Y is no higher than that in X.

(3) The WiFi communication cost of B in Y is not higher than that in X.

(4) The cellular communication cost of B in Y is not higher than that in X.

Intuitively, if X is dominated by Y, then the performance and the efficiency of X are no better than those of Y and therefore X is not worth further studying. In this subsection we identify the strategies that are dominated.

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

In the dominance analysis, the communication cost (but not the delay) of query-reports and metadata sub-reports is ignored for WiFi communication. Similarly, the communication cost of these reports is ignored for cellular communication. This is because query-reports and metadata sub-reports are very short. However, the simulations take into account the communication cost of the query-reports and metadata sub-reports (see subsection 5.1).

We say that strategy X is *weakly dominated* by strategy Y if the above dominance relationship only satisfies conditions 1-3, i.e. the cellular communication cost of Y may be higher. Weak dominance is appropriate for unlimited data plans offered by some cellular service providers.

4.2.2 Dominated Strategies

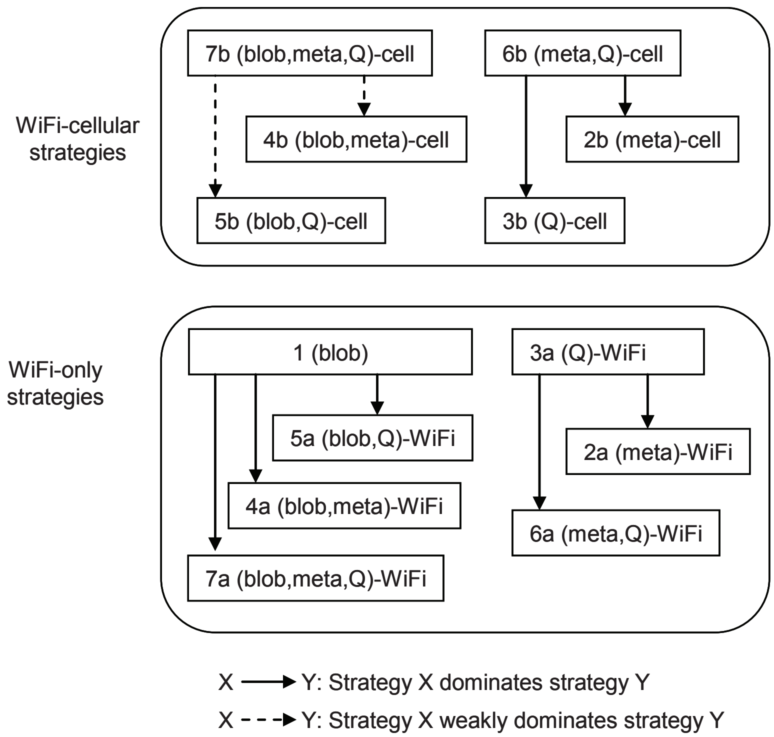


Figure 2. Dominance relationship among strategies.

The dominance relationships are summarized in Figure 2, and further analysis is provided in [13]. Observe that strategies 1 and 3a are incomparable because 3a disseminates only blobs that answer queries, whereas 1 disseminates all blobs thus its communication cost is higher; on the other hand, since 1 disseminates all blobs as

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

soon as they are produced, its response time is lower. Similarly, 7b and 6b are incomparable because the WiFi communication cost of 7b is higher, but its response time may be lower.

Observe that each dominated strategy is dominated by a strategy from the same group and thus is not worth further studying. Thus the next section focuses on the non-dominated strategies.

5. Comparison of Non-dominated Strategies by Simulations

In this section we compare by simulation the four non-dominated query processing strategies, namely: 1 (blob), 3a (Q)-WiFi, 7b (blob,meta,Q)-cell, and 6b (meta,Q)-cell (see Table 1). The comparisons are based on the application of delivering traffic multimedia clips among moving vehicles in order to warn drivers about traffic jams and dangers.

5.1 Multimedia Traffic Information Application

Based on experiments with a smart-phone video camera, the size of a blob report in the simulations is taken to be 65K bytes. Each query report specifies a *target-region* which indicates that the query producer is interested in receiving multimedia clips that started to be captured in this region.

Each blob report and each query report has a *lifetime* which is the length of the time period starting from the report producing time until the report expiration time. In the simulations all the reports have the same lifetime which is a system parameter. A report is dropped by the vehicles when its lifetime expires. A blob report R, or its metadata sub-report, *satisfies* a query report Q if (i) R is produced after the produce-time of Q; (ii) R.location falls within Q.target-region.

5.2 Simulation Environment

5.2.1 Mobility and Communication

The simulation area is taken to be a portion of the highway system in Chicago, the total length of which is 96km. We used SWANS++ [5] as the simulation tool which integrates vehicle mobility and WiFi communication. To deal with broadcast storms we used smart-flooding [1] and cooperative caching [11] for reports dissemination via WiFi. Smart-flooding reduces redundant rebroadcasts by regulating that only the receivers that are close to the boundary of the sender's transmission range rebroadcast. Cooperative caching

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

optimizes broadcast size and prioritizes broadcast content to maximize effective throughput. Since these are existing techniques we do not elaborate on them further.

In order to assure the robustness of the conclusions drawn from the SWANS++ communication model, we also conducted simulations with the WiFi communication component of SWANS++ replaced by a simple model we developed. We also conducted simulations with DSRC range and bandwidth, since DSRC is the designated inter-vehicle communication protocol. It turns out that the simple communication model and DSRC lead to the same conclusions as SWANS++ WiFi. In this paper we only present the results for SWANS++ WiFi due to space limitations. We augmented SWANS++ with cellular communication based on the typical parameters of 3G communication (see [14]). Table 2 lists all the simulation parameters.

5.2.2 Generation of Queries and Blob Reports

We consider range queries. The query target region is at distance 1600 meters ahead along the route of the query producer, and has a width of 500 meters. This means that the query producer is interested in traffic multimedia clips captured in the area lying between 1350 meters and 1850 meters ahead from its location at the query production time.

Every 10 seconds, each vehicle produces a blob report with a probability such that on average k blob reports are produced in the system per second. k is a system parameter and is called the *blobs supply*. Every 300 seconds, each participating vehicle produces a query with a probability called the *query ratio*.

Table 2. Simulation parameters and their values

Parameter		**Values**
Total length of road segments, total simulated area.		96 km, 24×31 sq. km.
Traffic condition	Light-congestion: 4000 vehicles, 64km/hour average speed over time among all road segments	
	Heavy-congestion: 8000 vehicles, reduced speed-limit for 50% of road segments, such that average speed over time among all road segments is	

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

	25km/hour
Penetration ratio (i.e., the fraction of vehicles that generate multimedia clips and participate in the WiMac query processing)	1% ~ 50%
WiFi transmission range, data transmission rate	250 meters, 2 Mbps[3]
DSRC transmission range, data transmission rate	500 meters, 12 Mbps
Side-length, capacity of each cell	2.5 km, 30 users
Data transmission rate of cellular channel	384 Kbps
Mobility model	STRAW
WiFi communication model	SWANS++, simple
Query ratio	0.25, 0.5, 0.75, 1
blobs supply (reports per second)	4, 8, 12, 16, 20
Query distance, query width	1600 meters, 500 meters.
Query/clip report lifetime	60, 120, 180, 240, 300 seconds
Sizes of query, metadata, and multimedia-clip	40bytes, 28bytes, and 65Kbytes
Reports database size	6 Mbytes
Length of a simulation run	3600 simulated seconds

5.3 Performance Measures

For each strategy we evaluated the following performance measures:

Answer throughput: The answer throughput is the average number of distinct answers (i.e., matching blob reports) received for each query. An answer is counted towards the throughput only if both the answer and the query have not expired at the time when the answer is received.

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

Communication overhead: The average number of bytes per vehicle submitted to the MAC level during the simulation, by the WiFi channel and the cellular channel, respectively. In other words, this overhead is the amount of attempted communication, the amount communicated successfully is lower.

For each parameter configuration, a simulation run is conducted to obtain the corresponding performance measures. In this run, each query of each vehicle is a sample, thus each simulation run contains enough samples to provide statistical significance. For example, for the heavy congestion situation, when penetration ratio=5%, report lifetime=300 seconds, query ratio=0.25, and blob supply=4 (see Figure 3(a)), 1286 queries are generated. In this case, for strategy 1 (blob), the answer throughput is 3.75; the standard deviation for the number of distinct query-answers received is 4.14. Thus, the 95% confidence interval is 0.22, i.e. less than 6% of 3.75.

5.4 Simulation Results

5.4.1 Answer throughput

Figures 3(a) and 3(e) show the answer throughput as a function of the penetration ratio for the heavy-congestion scenario and the light congestion scenario respectively. Figure 3(b) shows the answer throughput as a function of the report lifetime. Figure 3(c) shows the answer throughput as a function of the query ratio. Figures 3(d) and 3(f) show the answer throughput as a function of the blobs supply. It can be seen that in all the figures, ranking of the strategies based on throughput is 6b (meta,Q)-cell > 7b (blob,meta,Q)-cell > 1 (blob) > 3a (Q)-WiFi. Furthermore, the answer throughput of 6b is higher in the light congestion scenario than in the heavy scenario whereas for 7b, 1, and 3a, the reverse is true.

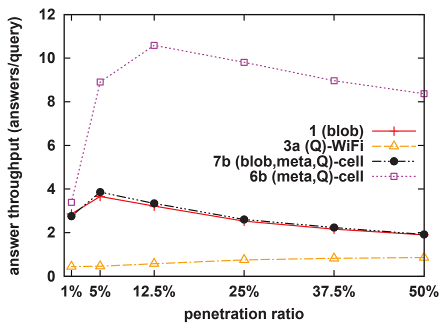
Best strategy. Strategy 6b (meta,Q)-cell is the clear winner. The advantage of 6b increases as the penetration ratio increases. In some cases, the answer throughput of 6b is seven times higher than those of the other strategies.

It is surprising that strategy 7b (blob,meta,Q)-cell is much worse than 6b. Compared to 6b which

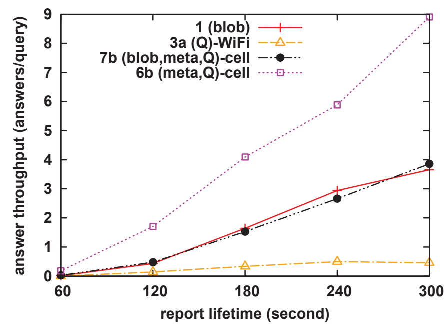
[3] 2Mpbs is the max transmission rate. i.e., without contention and collisions. Contention and collisions are accounted for by the simulation system, and they reduce the bandwidth according to the density of vehicles.

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

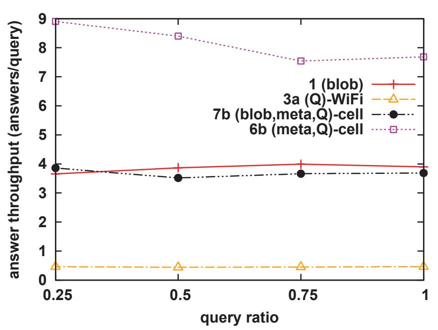
disseminates only metadata and query reports in the first WiMaC stage, strategy 7b disseminates blob reports as well in the first WiMaC stage, thus vehicles have a chance to receive answers from the WiFi dissemination directly. The poor performance of 7b is probably due to the fact that the WiFi dissemination of multimedia reports occupies a lot of WiFi bandwidth, which creates contention and collisions in the dissemination of metadata sub-reports and query reports. This interference significantly slows down the discovery of matches. Indeed, consider Figure 4(a) which shows the WiFi communication overhead of the four strategies. The WiFi communication overhead of 7b is much higher than that of 6b.



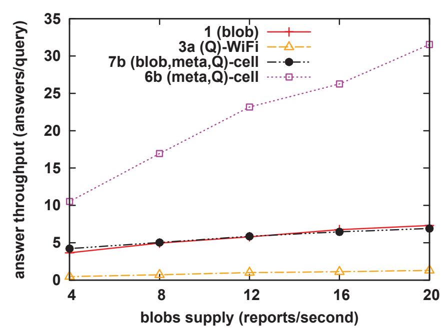
(a) Answer throughput versus penetration ratio, heavy congestion, report lifetime = 300 seconds, query ratio = 0.25, blobs supply = 4 per second.



(b) Answer throughput versus report lifetime, heavy congestion, penetration ratio = 0.05, query ratio = 0.25, blobs supply = 4 per second.



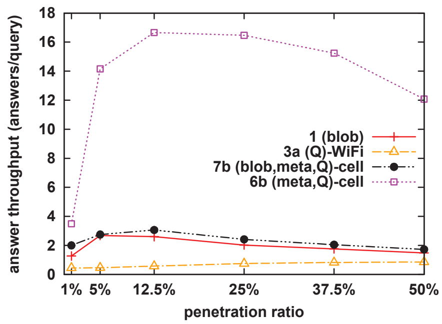
(c) Answer throughput versus query ratio, heavy congestion, penetration ratio = 0.05, report lifetime = 300 seconds, blobs supply = 4 per second.



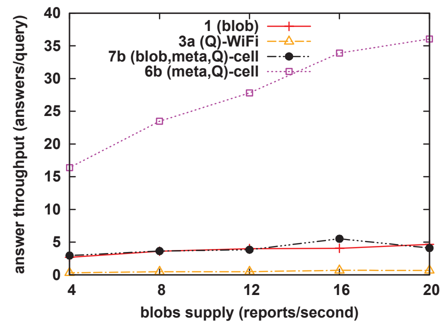
(d) Answer throughput versus blobs supply, heavy congestion, penetration ratio = 0.05, report lifetime = 300 seconds, query ratio = 0.25.

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.



(e) Answer throughput versus penetration ratio, light congestion, report lifetime = 300 seconds, query ratio = 0.25, blobs supply = 4 per second.



(f) Answer throughput versus blobs supply, light congestion, penetration ratio = 0.05, report lifetime = 300 seconds, query ratio = 0.25.

Figure 3. Comparison in terms of the answer throughput performance, heavy-congestion, SWANS++.

Comparison of WiFi-only strategies. 1 (blob) is better than 3a (Q)-WiFi. 1 and 3a represent two paradigms of query processing, i.e., 1 represents push (data-to-query) and 3a represents pull (query-to-data). The simulation results show that push is better than pull for the considered environment. Intuitively, the pull strategy requires a round-trip dissemination in order for a query originator to receive an answer: the query has to travel from the query originator to the answer producer and then the answer has to travel back from the answer producer to the query originator. If either way does not go through or experiences a long delay, then the answer does not reach the query originator within the lifetime; and this scenario is likely in a highly mobile environment.

Feasibility of WiMaC. Even in the 1% penetration ratio and light congestion, the answer throughputs of 1, 6b, and 7b are at least 1. This fact is surprising, because when the penetration ratio is 1%, the inter-participating-vehicle distance is approximately 2400 meters, i.e. much higher than the WiFi transmission range of 250 meters. With this density, the network is highly disconnected. In other words, it is expected that most of the time a vehicle does not have any neighbors within its transmission range. Yet on average each query receives at least one answer. This is due to the store-and-forward (namely cooperative caching) mechanism, which enables WiFi dissemination even when the network is highly disconnected.

Impact of the penetration ratio (Figure 3(a)). When the penetration ratio increases, the answer

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

throughputs of 1, 7b, and 6b initially increase, and then decrease. Intuitively, when the penetration ratio increases, two effects are generated:

(i) The WiFi network becomes more connected, which pulls the answer throughput up;

(ii) The contention and collisions increase for the WiFi network, which pulls the answer throughput down.

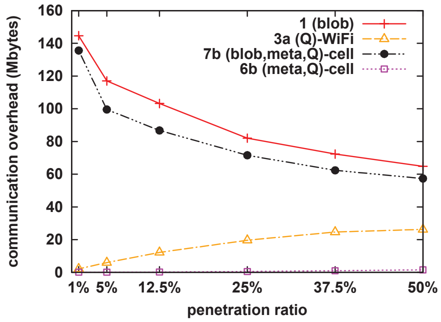
The answer throughput curves shown in Figure 3(a) are the result of the interplay of these two effects. For 6b, another factor contributes to the drop of the answer throughput as explained below. In 6b, answers are delivered only by cellular communication. The number of answers that can be delivered is limited by the capacity of the cellular channel. When the penetration ratio increases, there are more queries to answer. On the other hand, cellular communication is one-to-one; it does not scale well to the number of receivers. When the penetration ratio is high, the capacity of the cellular channel is exceeded and only a fraction of discovered queries can be answered.

Impact of the report lifetime (Figure 3(b)). For all the strategies, the throughput increases with the report lifetime.

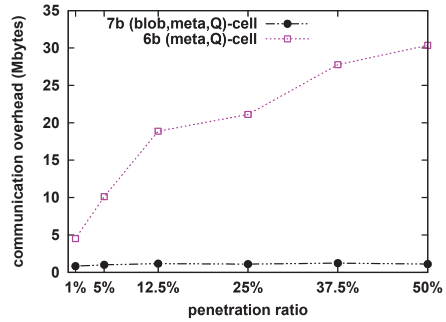
Impact of the query ratio (Figure 3(c)). The throughputs of 1 (blob) and 3a (Q)-WiFi change little change with the query ratio. This is because these strategies use WiFi broadcasting to disseminate blob reports, and each broadcast satisfies multiple queries. The throughput of 6b (meta,Q)-cell decreases with the query ratio. The reason is again, the cellular communication does not scale well to the increase of the number of queries.

Impact of the blobs supply (Figure 3(d)). For all the strategies, the throughput increases with the blobs supply because there are more answers available.

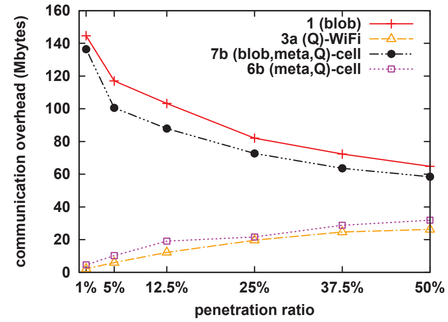
Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE



(a) WiFi communication overhead versus penetration ratio, heavy congestion, report lifetime = 300 seconds, query ratio = 0.25, blobs supply = 4 per second.



(b) Cellular communication overhead versus penetration ratio, heavy congestion, report lifetime = 300 seconds, query ratio = 0.25, blobs supply = 4 per second.



(c) Total communication overhead versus penetration ratio, heavy congestion, report lifetime = 300 seconds, query ratio = 0.25, blobs supply = 4 per second.

Figure 4. Comparison in terms of the communication overhead performance, heavy-congestion, SWANS++.

5.4.2 Communication Overhead

Figures 4(a,b) show the WiFi/cellular communication overhead as a function of the penetration ratio. The throughput for the same configuration is presented in Figure 3(a). It can be seen that the winning strategy in terms of throughput, 6b (meta,Q)-cell, has the lowest WiFi communication overhead, because only metadata reports and query reports (which are short) are disseminated via WiFi. 6b has the highest cellular communication overhead. The total communication overhead of 6b (including WiFi and cellular) is only slightly higher than that of 3a (Q)-WiFi and is much lower than those of 1 (blob) and 7b (blob,meta,Q)-cell. On the other hand, the throughput of 6b is seven times higher than those of the other strategies (see Figure 3(a)). This fact suggests that 6b is very efficient on bandwidth consumption.

6. Conclusion

In this paper we developed and compared strategies for querying blob data in vehicular networks, where cellular communication is also available. We first introduced the WiMac paradigm and derived a list of 13 possible query processing strategies. We determined that 4 of the 13 WiMac strategies dominate the others. Then we compared the four non-dominated strategies by simulations in a vehicular environment. The simulations revealed that 6b has by far a higher throughput than the other three (up to 7-fold). Furthermore, the

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

communication cost of strategy 6b is also lower than that of the others. Intuitively, strategy 6b operates as follows. It separates metadata dissemination from its blob report, it combines push of metadata and pull by queries, and uses the cellular infrastructure to communicate blobs.

7. REFERENCES

[1] T. Osafune, L. Lin, and M. Lenardi. Multi-hop vehicular broadcast (MHVB). *6th International Conference on ITS Telecommunications*, pp. 757-760, Chengdu, China, June 21-23, 2006.

[2] J. Park, U. Lee, S. Y. Oh, M. Gerla, and D. S. Lun. Emergency Related Video Streaming in VANET using Network Coding. *VANET*, 2006.

[3] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi. Dissemination and Harvesting of Urban Data Using Vehicular Sensing Platforms. *IEEE Transactions on Vehicular Technology*, (58)2:882-901, 2009.

[4] J. Zhao, Y. Zhang, and G. Cao. Data pouring and buffering on the road: A new data dissemination paradigm for vehicular ad hoc networks. IEEE Transactions on Vehicular Technology, (56)6:3266-3277, 2007.

[5] http://www.aqualab.cs.northwestern.edu/projects/swans++/

[6] D. Choffnes and F. Bustamante. An Integrated Mobility and Traffic Model for Vehicular Wireless Networks. *2nd ACM international workshop on Vehicular ad hoc networks*, pp. 69-78, Cologne, Germany, September 2, 2005.

[7] X. Liu, Q. Huang, and Y. Zhang. Balancing Push and Pull for Efficient Information Discovery in Large-Scale Sensor Networks. IEEE Trans. on Mobile Computing, 6(3):241-251, 2007.

[8] Y. Zhang, J. Zhao, and G. Cao. Roadcast: A Popularity Aware Content Sharing Scheme in VANETs. *29th IEEE International Conference on Distributed Computing Systems*, pp. 223-230, Montreal, Québec, Canada, June 22-26, 2009

[9] B. Han, P. Hui, M. Marathe, G. Pei, A. Srinivasan, A. Vullikanti. Cellular Traffic Offloading through Opportunistic Communications: A Case Study. *5th Workshop on Challenged Networks*, pp. 31-38, Beijing, China, September 25, 2010.

[10] S. Al-Chikhani, L. Al-Kanj, and Zaher Dawy. Video Distribution over Wireless Networks with Mobile-to-Mobile Cooperation. *International Conference on Advances in Computational Tools for Engineering Applications*, pp. 519-522, Notre Dame University, Lebanon, July 15-17, 2009.

[11] T. Zhong, B. Xu, P. Szczurek, and O. Wolfson. Trafficinfo: An Algorithm for VANET Dissemination of Real-Time Traffic Information. *15th World Congress on Intelligent Transportation Systems, New York, NY, Nov. 2008.*

[12] M. Leung and S.-H. Chan. Broadcast-Based Peer-to-Peer Collaborative Video Streaming Among Mobiles. IEEE Trans. on Broadcasting, (53)1: 350-361, 2007.

[13] B. Xu, O.Wolfson, J. Lin. Multimedia Data in Hybrid Vehicular Networks. 8th International Conference on Advances in Mobile Computing and Multimedia (MOMM), pp. 109-116, Paris, France, Nov. 2010.

[14] P. A. Hosein. Capacity of packetized voice services over time-shared wireless packet data channels. *24th Annual Joint Conference of the IEEE Computer and Communications Societies*, pp. 2032-2043, Miami, Florida, March 13-17, 2005.

[15] http://www.cs.uic.edu/~boxu/video_clips/video_13126071744_11-21-24.asf

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE

[16] M. Guo, M. Ammar, E. Zegura. V3: A vehicle-to-vehicle live video streaming architecture. Third *IEEE International Conference on Pervasive Computing and Communications*, pp. 171-180, Kauai, Hawaii, March 8-12, 2005.

[17] N. Qadri, M. Fleury, B. Rofoee, M. Altaf, M. Ghanbari. Robust P2P Multimedia Exchange within a VANET. *Wireless Personal Communications*, 63(3):561-577, 2012.

[18] T. Delot, N. Cenerario, S. Ilarri, and S. Lecomte. A cooperative reservation protocol for parking spaces in vehicular ad hoc networks. *6th International Conference on Mobile Technology, Application & Systems*, Nice, France, September 10-13, 2009.

Short Author Bios:

Ouri Wolfson is currently the Richard and Loan Hill Professor of Computer Science at the University of Illinois at Chicago, and an Affiliate Professor in the Department of Computer Science at the University of Illinois at Urbana Champaign. He received his B.A. degree in mathematics, and his Ph.D. degree in computer science from Courant Institute of Mathematical Sciences, New York University. His main research interests are in database systems, distributed systems, and mobile/pervasive computing.

Bo Xu is currently a Postdoctoral Appointee in the Transportation Research and Analysis Computing Center at Argonne National Laboratory, and a researcher in the Department of Computer Science at the University of Illinois at Chicago. He received a Ph.D. degree with the Department of Computer Science, University of Illinois at Chicago. His main research interests are in mobile data management and computational transportation science.

Complete contact information:

Ouri Wolfson, Department of Computer Science (M/C 152), 851 S. Morgan St., University of Illinois, Chicago, IL 60607-7053, (312) 996-6770 (voice), (312) 413-0024 (fax), wolfson@cs.uic.edu, Web: http://www.cs.uic.edu/~wolfson/

Bo Xu, Argonne TRACC Service Desk, 9700 South Cass Avenue, Building 222, Argonne, IL 60439, Phone: 630-252-5200, Fax: 630-252-5394

Digital Object Indentifier 10.1109/MMUL.2013.32 1070-986X/$26.00 2013 IEEE