A New Paradigm for Querying Blobs in Vehicular Networks

Ouri Wolfson
University of Illinois at Chicago

Bo Xu
Argonne National Laboratory and University of Illinois at Chicago

A novel paradigm and several derived strategies that query binary large objects in a wireless vehicular network reveal the answer throughput and communication overhead of various approaches.

A vehicular ad hoc network (Vanet) is a set of vehicles, or mobile peers, that communicate with each other via short-range wireless technologies such as WiFi and dedicated short range communications (DSRC). An attractive ITS application of Vanets is to share binary large objects (blobs) such as voice and video clips for situation awareness and urban monitoring.[1–4] For example, many taxicabs currently use cameras that capture continuous videos of the traffic ahead. Additionally, driver-recorded time- and location-stamped audio clips—such as those providing a brief 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 (such as allowing police to track wanted cars). In fact, we have conducted field experiments in which videos are automatically captured by dashboard-mounted smartphones and disseminated among vehicles in a peer-to-peer fashion. A two-second 116-Kbyte sample clip can be viewed at www.cs.uic.edu/~boxu/video_clips/video_13126071744_11-21-24.asf. The clip shows the traffic condition of a segment of Ashland Avenue in Chicago. As the clip indicates, two seconds are sufficient to see whether 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, and blobs are disseminated as responses to received queries.[1] (See the “Relevant Work in Wireless Vehicular Networks” sidebar for more details on previous research.) Vanets have also been studied as an augment to the cellular communication for cellular offloading purpose.[5–7] In this case, blob sources reside in a fixed network and must be disseminated to numerous 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 then share the blobs with the other peers via the short-range communication. We refer to Vanets in which the cellular communication is available as *hybrid vehicular networks*.

Here, we consider hybrid vehicular networks in which blobs are generated by the mobile peers (such as a vehicle generating a two- to three-second video of the surroundings 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. As in a typical Vanet, a peer does not initially know the network ID (that is, the cell phone number) of other peers in the network. However, a peer can communicate directly with other peers within its WiFi transmission range without knowing their network ID. (Although we use the term WiFi here for simplicity, our results apply to other short-range networking technologies, including DSRC). Furthermore, we do not require or assume a central server or any other form of directory or storage service in the fixed network.

Such an environment renders a broad spectrum of possible query-processing strategies along three design dimensions. First, the peer-to-peer communication can use WiFi alone or a hybrid of WiFi and cellular communication. Pure cellular communication is not an option because it requires knowledge of the receiver’s

1070-986X/14/$31.00 © 2014 IEEE

Published by the IEEE Computer Society

Relevant Work in Wireless Vehicular Networks

Researchers have demonstrated that, in a static environment, the combination of push and pull is superior to pure push and pure pull.[1] In our case, the dissemination of blobs/queries follows a geometric structure, such as line segments or trees. Such structure-based methods do not work in vehicular ad hoc networks (Vanets) because of mobility and disconnection factors.

Many methods have been proposed for resource discovery and data dissemination in mobile ad hoc networks (Manets) and Vanets.[2,3] In contrast to our work, such methods do not use cellular communication. However, they do provide an insight that helped our work: WiFi communication in WiFi-communication, Match, Communication (WiMaC) strategies can be improved by using cooperative caching and prioritization.

In their work, Thierry Delot and his colleagues proposed a protocol with which parking reports are exchanged in a vehicle-to-vehicle fashion. [4] They consider short reports giving the locations of available parking slots, whereas here we address the exchange of much larger messages containing blobs.

Meng Guo and his colleagues proposed V3,[5] a vehicle-to-vehicle live video streaming architecture. The architecture adopts a query-to-data paradigm for query processing and uses only WiFi communication. Thus, the query-processing method in V3 is similar to the (Q)-WiFi strategy that we discuss in the main article. However, the V3 work does not compare the query-to-data paradigm with other WiFi-only strategies or with WiFi-cellular strategies.

Other researchers have studied reliable multimedia delivery in Vanets.[6,7] Those works focused on how to encode the frames of a video clip communicated via WiFi so that the receiver can recover from packet losses. Those methods provide the insight that the reliability of multimedia transmission in the WiMaC system can be improved by encoding and error correction.

References

1. X. Liu, Q. Huang, and Y. Zhang, "Balancing Push and Pull for Efficient Information Discovery in Large-Scale Sensor Networks," *IEEE Trans. Mobile Computing,* vol. 6, no. 3, 2007, pp. 241–251.
2. U. Lee et al., "Dissemination and Harvesting of Urban Data Using Vehicular Sensing Platforms," *IEEE Trans. Vehicular Technology,* vol. 58, no. 2, 2009, pp. 882–901.
3. Y. Zhang, J. Zhao, and G. Cao, "Roadcast: A Popularity Aware Content Sharing Scheme in Vanets," *Proc. 29th IEEE Int'l Conf. Distributed Computing Systems,* IEEE, 2009, pp. 223–230.
4. T. Delot et al., "A Cooperative Reservation Protocol for Parking Spaces in Vehicular Ad Hoc Networks," *Proc. 6th Int'l Conf. Mobile Technology, Application & Systems,* ACM, 2009, article no. 30.
5. M. Guo, M. Ammar, and E. Zegura, "V3: A Vehicle-To-Vehicle Live Video Streaming Architecture," *Proc. 3rd IEEE Int'l Conf. Pervasive Computing and Comm.,* IEEE, 2005, pp. 171–180.
6. J. Park et al., "Emergency Related Video Streaming in Vanet Using Network Coding," *Proc. 3rd Int'l Workshop Vehicular Ad Hoc Networks,* ACM, 2006, pp. 102–103.
7. N. Qadri et al., "Robust P2P Multimedia Exchange within a Vanet," *Wireless Personal Comm.,* vol. 63, no. 3, 2012, pp. 561–577.

network ID, whereas a query originator initially knows only the description of the requested blobs but not the network ID of the peers containing these blobs. Second, the query processing may adopt push or pull, or a combination of the two. Third, a blob can be described by a brief metadata tag (such as the time and location at which a multimedia clip was produced), and the match (yes or no) between a query and a blob can be determined solely based on the metadata. Because of size differences, the metadata and content of a given blob can be disseminated independently, and by different means (WiFi or cellular). Figure 1 shows the three design dimensions.

We introduce a novel paradigm of query-processing strategies based on these three design dimensions. The paradigm—called WiFi-communication, Match, Communication (WiMaC)—generates a list of 13 possible query-processing strategies. We then define the notion of dominance 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 conducted analysis based on this definition and determined that four of the 13 WiMac strategies dominate the others. Finally, we compare these four strategies using a simulation in the context of querying vehicle-captured multimedia traffic information.

The Model

Our model environment system consists of a set of mobile peers. This set of peers might

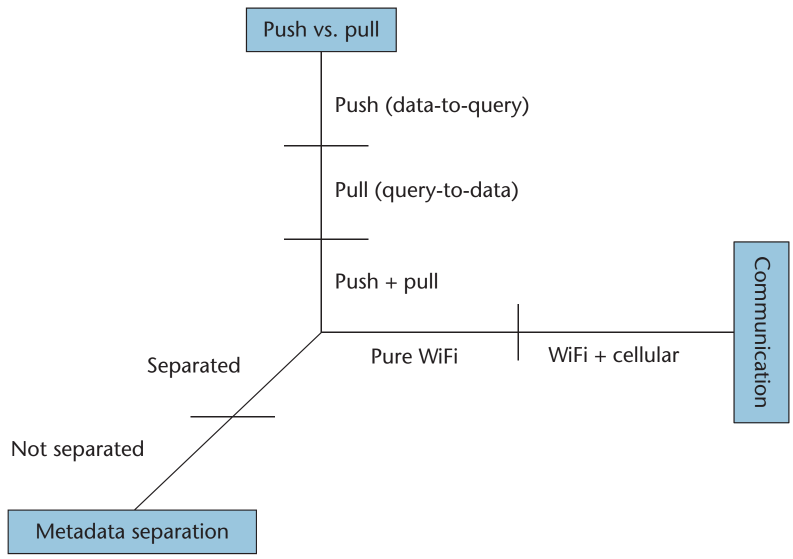


Figure 1. The design space for blob query processing in hybrid vehicular networks. The environment makes possible a broad spectrum of query-processing strategies related to communication, metadata, and push–pull approaches.

change over time. Each peer (such as a vehicle) is equipped with three capabilities:

- blob data production, such as video, voice, or multimedia clips;
- short-range wireless communication, such as WiFi; and
- infrastructure-based communication, such as 3G cellular.

Through the infrastructure, a peer can transmit messages to another peer using Multimedia Message Service (MMS) 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 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 a peer's network ID from personal information. However, a detailed privacy and security analysis is an orthogonal issue left for future work.

Reports and Reports Databases

Each peer periodically produces blob reports. Formally, a blob report B is a couple, <Meta(B), Blob(B)>, where Meta(B) and Blob(B) are the metadata and blob subreports, respectively. The metadata subreport contains attributes describing the blob such as the time when the blob was produced, the location at which it was produced, the network ID of the producing peer, and so on. Blob(B) is the blob itself, such as a music or video file.

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

A peer is the *producer* of the query and blob reports that it produces. Each query and each metadata subreport contains the network ID of the report's producer. Each query, Q, has an expiration time that indicates that query's processing stops when its expiration time is reached. The query expiration time is attached to Q when it is created and remains fixed during the query's dissemination. Similarly, each blob report has a fixed blob expiration time, beyond which it becomes invalid. In this sense, query processing in a Vanet is "best effort" because delivery of all the answers is not guaranteed. Each peer maintains a reports database that stores the metadata subreports, 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.[8]

WiMaC Query-Processing Strategies

Some or all of the reports that satisfy Q may reside on peers that are different than the query producer (Qp). Because the Qp does not typically have the network IDs of such peers and does not even know how many reports satisfy the query, all query-processing strategies start with a WiFi dissemination to neighboring peers. This dissemination may be of the query, the blob reports, the metadata subreports, or some combination. Blob reports and metadata subreports propagate differently by WiFi; the metadata report propagates faster and thus might meet 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 subreport is located at another peer. In this case, the blob must be transferred to the Qp through additional communication. This is the WiMaC paradigm, and all query-processing strategies we discuss here are special cases of WiMaC.

Strategy Design Space

Table 1 shows the structure of the design space. There are 13 WiMaC strategies, and each is denoted by a strategy number and strategy name. The strategy name is formed as follows. If there is not a second stage, then the strategy is named by the first stage—that is, (blob). If there is a second stage, then the strategy is named by the two stages connected by a "-". For example, the name *2b (meta)-cell* denotes a 2b strategy that disseminates metadata subreports in the first stage and uses cellular communication in the second.

There are seven WiFi-only strategies:

- **1 (blob).** In stage one, 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. This strategy has no second stage, and it corresponds to the push (data-to-query) paradigm.

- **2a (meta)-WiFi.** In stage one, metadata subreports are disseminated via WiFi. When a metadata subreport Meta(B) **//Are both metadata and blob reports "B"? Please clarify.//** reaches the producer of a matching query Q, the producer of Q (Qp) disseminates Q via WiFi. When the producer of B (Bp) receives Q, the Bp disseminates B via WiFi to reach the Qp and provide an answer to Q.

- **3a (Q)-WiFi.** In stage one, queries are disseminated via WiFi. When a query Q reaches the producer of a matching blob report B, the Bp disseminates B via WiFi to reach the Qp. 3a (Q)-WiFi corresponds to the pull (query-to-data) paradigm.

- **4a (blob,meta)-WiFi.** In stage one, metadata and blob reports are disseminated separately via WiFi. If the producer of a matching query Q receives a blob report B, then there is no second stage. If the Qp receives a metadata sub-report Meta(B), the Qp disseminates Q via WiFi. When a peer Z with a matching blob report B receives Q, Z disseminates B via WiFi to reach the Qp.
- **5a (blob,Q)-WiFi.** In stage one, blob and query reports are disseminated via WiFi. When a blob report B and a matching query Q collocate at a peer Z, Z disseminates B via WiFi to reach the Qp.

- **6a (meta,Q)-WiFi.** In stage one, **//rewording correct?//** metadata subreports and queries are disseminated via WiFi. When a metadata subreport Meta(B) and a matching Q collocate at a peer Z, Z disseminates Q via WiFi. When the Bp receives Q, the Bp disseminates the corresponding blob report B via WiFi to reach the Qp.

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

WiFi-Cellular Strategies. In the WiFi-cellular strategies, after a match is discovered, the answer B is communicated from a peer P to the Qp through the cellular channel. However, P first inquires via the cellular channel whether the producer has already received B

Table 1. Design space of the WiMaC paradigm.

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

Blob = blob report; meta = metadata sub-report; Q = query; cell = cellular. For the strategy names, 1 is (blob), 2a is (meta)-WiFi, 3b is (Q)-cell, and so on.

(from other peers); if so, the transmission of B is suppressed.

- **2b (meta)-cell.** In stage one, metadata subreports are disseminated via WiFi; when a **metadata subreport Meta(B)** reaches the producer of a matching query Q, the Qp sends Q to the Bp via the cellular channel. In response, the Bp sends B and all the other matching blob reports that it has to the Qp via the cellular channel.
- **3b (Q)-cell.** In stage one, queries are disseminated via WiFi. When a query Q reaches the producer of a matching blob report B, the Bp sends B to the Qp via the cellular channel.
- **4b (blob,meta)-cell.** This strategy is identical to (meta)-cell (2b), except that blob reports are also disseminated in WiMaC's first stage. There is no second stage if the Qp receives a blob report from the WiFi dissemination.
- **5b (blob,Q)-cell.** In stage one, blob and query reports are disseminated via WiFi. When a blob report B and a matching query Q collocate at a peer Z, Z sends B to the Qp via the cellular channel.
- **6b (meta,Q)-cell.** In stage one, metadata and query reports are disseminated via WiFi. When a **metadata subreport Meta(B)** and matching query Q collocate at a peer Z, Z sends Q to the Bp via the cellular channel. In response, the Bp sends the blob report B and all other matching blob reports it has to the Qp through the cellular channel.
- **7b (blob,meta,Q)-cell.** This strategy is a combination of 5b (blob,Q)-cell and 6b (meta,Q)-cell.

Strategy Dominance Analysis

Given the number of possible WiMaC strategies, the question now is how to choose which one(s) should be used in real applications. For real applications, the performance aspects in terms of answer throughput, response time, and communication cost are usually the primary concerns. In the following analysis, we show that, under reasonable assumptions, some of the WiMacC strategies are better than others in all these aspects and in this sense are "dominating" strategies.

Definitions and Assumptions. Let a peer receive an answer blob report at time *t*. The answer's response time is the length of the time period because the answer is produced until *t*. A strategy, X, is dominated by another strategy, Y, if the following four conditions are satisfied for every blob report (B):

- For every query that B answers, if the answer is received in Y, it is also received in X.
- For every query that B answers, its response time t in Y is no higher than that in X.
- The WiFi communication cost of B in Y is not higher than that in X.
- The cellular communication cost of B in Y is not higher than that in X.

Intuitively, if X is dominated by Y, then X's performance and efficiency are no better than those of Y. In this case, X is not worth further study. In the following, we identify the strategies that are dominated.

In the dominance analysis, the communication cost (but not the delay) of query reports and metadata subreports is ignored for WiFi communication. Similarly, the communication cost of these reports is ignored for cellular communication. This is because query reports and metadata subreports are short. However, the simulations account for the communication cost of the query reports and metadata subreports (as we describe later).

We say that strategy X is weakly dominated by strategy Y if the dominance relationship satisfies only conditions 1–3—that is, Y's cellular communication cost might be higher. Weak dominance is appropriate for unlimited data plans offered by some cellular service providers.

Dominated Strategies. The dominance relationships are summarized in Figure 2, and further analysis is provided elsewhere.[9] As Figure 2 shows, strategies 1 and 3a are incomparable because 3a disseminates only blobs that answer queries, whereas 1 disseminates all blobs and thus has higher communication costs. On the other hand, because 1 disseminates all blobs as 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 might be lower.

As Figure 2 shows, each dominated strategy is dominated by a strategy from the same group and thus is not worth further study.

Using Simulations to Compare Nondominated Strategies

We now use simulations to compare the four nondominated query-processing strategies: 1 (blob); 3a (Q)-WiFi; 7b (blob,meta,Q)-cell; and 6b (meta,Q)-cell (see Table 1). The comparisons are based on delivery of traffic multimedia clips among moving vehicles to warn drivers about traffic jams and dangers.

Multimedia Traffic Information Application

Based on experiments with a smartphone video camera, we use 65 Kbytes as the size of a blob report in the simulations. Each query report specifies a target region that indicates that the Qp 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. The vehicles drop a report when its lifetime expires. A blob report B, or its metadata subreport, satisfies a query report Q if B is produced after the produce time of Q, or B.location falls within Q.target-region.

Simulation Environment

We developed a simulation testbed to simulate movement of vehicles, intervehicle communication, and generation of query and blob reports.

Mobility and Communication. For the simulation area, we chose a portion of the highway system in Chicago, the total length of which is 96 km. For the simulation tool, we used SWANS++ (http://www.aqualab.cs.northwestern.edu/projects/143-swans-extensions-to-the-scalable-wireless-ad-hoc-network-simulator), which integrates vehicle mobility and WiFi communication. To deal with broadcast storms, we used smart flooding[10] and cooperative caching[8] for reports dissemination via WiFi. Smart flooding reduces redundant rebroadcasts by regulating that only receivers close to the boundary of the sender's transmission range rebroadcast. Cooperative caching optimizes broadcast size and prioritizes broadcast content to maximize effective throughput. Because these are existing techniques, we will not elaborate on them further.

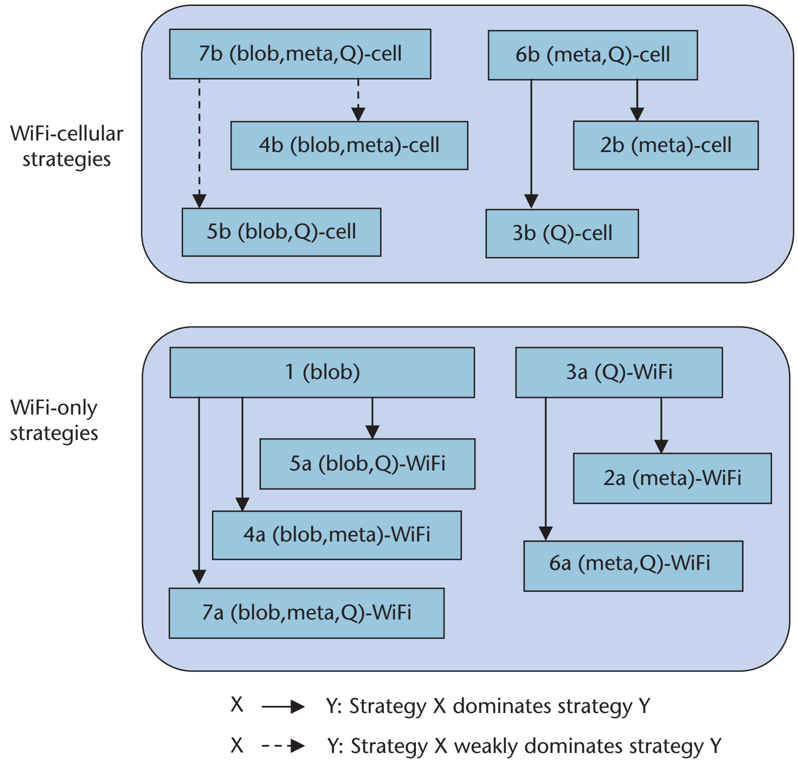


Figure 2. Dominance relationship among strategies. Because strategy 3a disseminates only blobs that answer queries, it is incompatible with strategy 1, which disseminates all blobs and thus has higher communication costs.

To assure the robustness of the conclusions drawn from the SWANS++ communication model, we also conducted simulations with the SWANS++ WiFi communication component replaced with a simple model we developed. In addition, we conducted simulations with DSRC range and bandwidth because DSRC is the designated intervehicle communication protocol. As it turns out, the simple communication model and DSRC lead to the same conclusions as SWANS++ WiFi. Here, we present only the results for SWANS++ WiFi due to space limitations. We augmented SWANS++ with cellular communication based on the typical parameters of 3G communication.[11] Table 2 lists all the simulation parameters.

Query and blob report generation. We consider range queries with a target region of 1,600 meters ahead along the Qp's route, with a width of 500 meters. Thus, the Qp is interested in traffic multimedia clips captured in the area lying between 1,350 meters and 1,850 meters ahead of its location at the query production time.

*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: 4,000 vehicles, 64 km/hour average speed over time among all road segments
	Heavy-congestion: 8,000 vehicles, reduced speed-limit for 50% of road segments, such that average speed over time among all road segments is 25 km/hour
Penetration ratio (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
DSRC transmission range/data transmission rate	500 meters/12 Mbps
Side-length/cell capacity	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	1,600 meters/500 meters
Query/clip report lifetime	60, 120, 180, 240, 300 seconds
Sizes of query, metadata, and multimedia clip	40 bytes, 28 bytes, and 65 Kbytes
Reports database size	6 Mbytes
Length of a simulation run	3,600 simulated seconds

* The maximum transmission rate is 2 Mpbs, without contention and collisions, which are accounted for by the simulation system and reduce the bandwidth according to the density of vehicles.

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 called the *blobs supply.*) Every 300 seconds, each participating vehicle produces a query with a probability called the *query ratio*.

Performance Measures

For each strategy, we evaluated two performance measures:

- *Answer throughput*. The answer throughput is the average number of distinct answers (that is, matching blob reports) received for each query. An answer is counted toward the throughput only if both the answer and query have not expired at the time when the answer is received.
- *Communication overhead*. The average number of bytes per vehicle submitted to the MAC level during the simulation, by the WiFi channel and cellular channel, respectively. In other words, this overhead is the amount of attempted communication; the amount communicated successfully is lower.

For each parameter configuration, we conduct a simulation run to obtain the corresponding performance measures. In this run, each query of each vehicle is a sample, so each simulation run contains enough samples to provide statistical significance. For example, for the heavy-congestion situation (Figure 3a), the penetration ratio is 5 percent, the report lifetime is 300 seconds, the query ratio is 0.25, the blob supply is four, and 1,286 queries are generated. In this case, for strategy 1 (blob), the answer throughput is 3.75, and the standard deviation for the number of distinct query answers received is 4.14. Thus, the 95 percent confidence interval is 0.22—that is, less than 6 percent of 3.75.

Simulation Results: Answer Throughput

Figures 3a and 3e show the answer throughput as a function of the penetration ratio for the heavy- and light-congestion scenarios, respectively. Figure 3b shows the answer throughput as a function of the report lifetime. Figure 3c

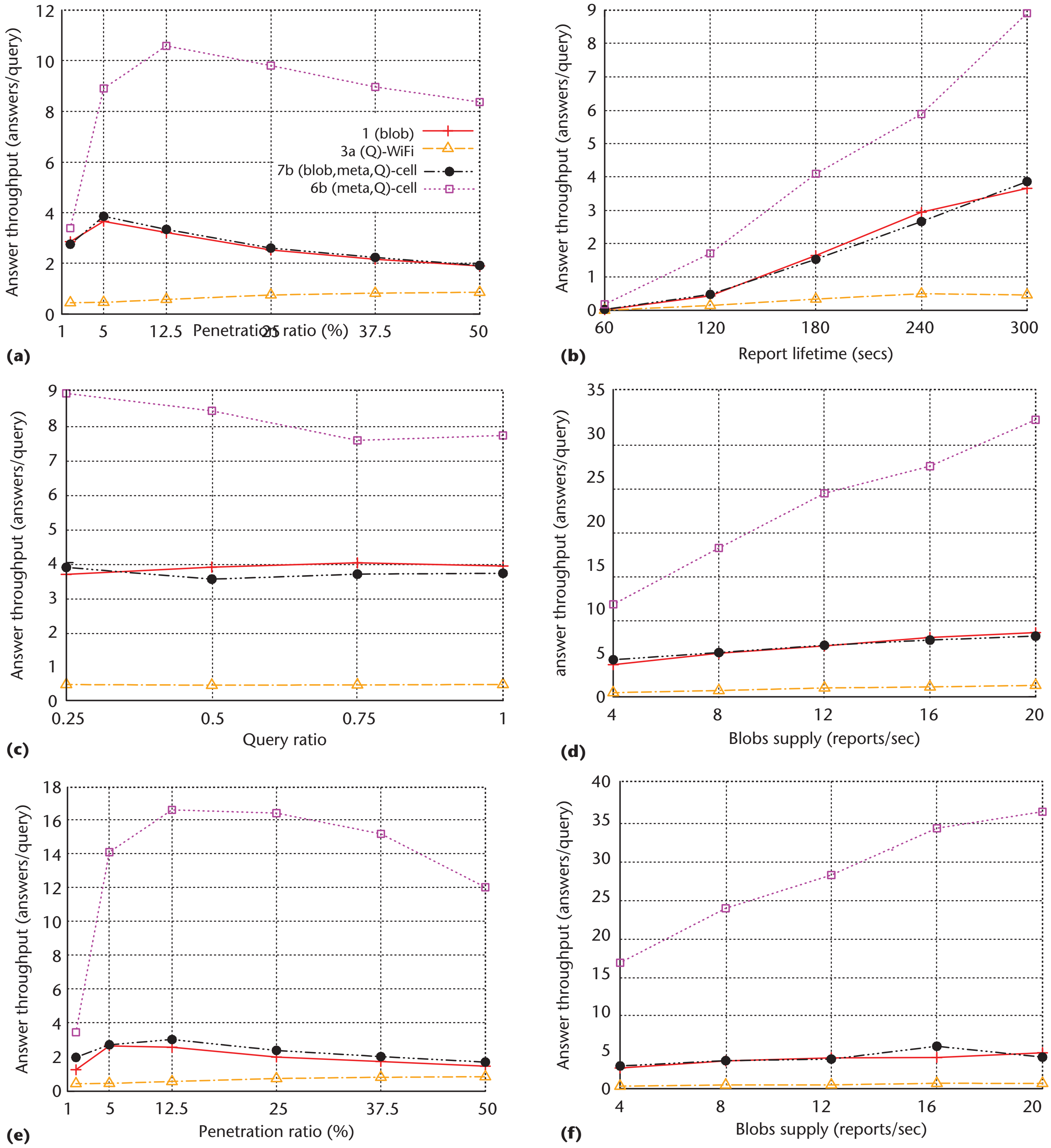


Figure 3. Comparison of answer throughput performance, heavy-congestion using SWANS++. (a) Answer throughput versus penetration ratio, heavy congestion. The report lifetime is 300 seconds, the query ratio is 0.25, and the blobs supply is four per second. (b) Answer throughput versus report lifetime, heavy congestion. The penetration ratio is 0.05, the query ratio is 0.25, and the blobs supply is four per second. (c) Answer throughput versus query ratio, heavy congestion. The penetration ratio is 0.05, the report lifetime is 300 seconds, and the blobs supply is four per second. (d) Answer throughput versus blobs supply, heavy congestion. The penetration ratio is 0.05, the report lifetime = 300 seconds, and the query ratio = 0.25. (e) Answer throughput versus penetration ratio, light congestion. The report lifetime is 300 seconds, the query ratio is 0.25, and the blobs supply is four per second. (f) Answer throughput versus blobs supply, light congestion. The penetration ratio is 0.05, the report lifetime is 300 seconds, and the query ratio is 0.25.

shows the answer throughput as a function of the query ratio. Figures 3d and 3f show the answer throughput as a function of the blobs supply. 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.

Strategy options. 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 with 6b—which disseminates only metadata and query reports in the first WiMaC stage—strategy 7b also disseminates blob reports in the first stage, and thus vehicles can 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 4a, which shows the WiFi communication overhead of the four strategies. The WiFi communication overhead of 7b is much higher than that of 6b.

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

WiMaC Feasibility. Even with 1 percent penetration ratio and light congestion, the answer throughputs of 1, 6b, and 7b are at least one. This fact is surprising because when the penetration ratio is 1 percent, the inter-participating-vehicle distance is approximately 2,400 meters, which is much higher than the WiFi transmission range of 250 meters. With this density, the network is highly disconnected. In other words, most of the time, a vehicle does not have any neighbors in its transmission range. Yet, on average, each query receives at least one answer. This is due to the store-and-forward, or *cooperative caching*, mechanism, which enables WiFi dissemination even when the network is highly disconnected.

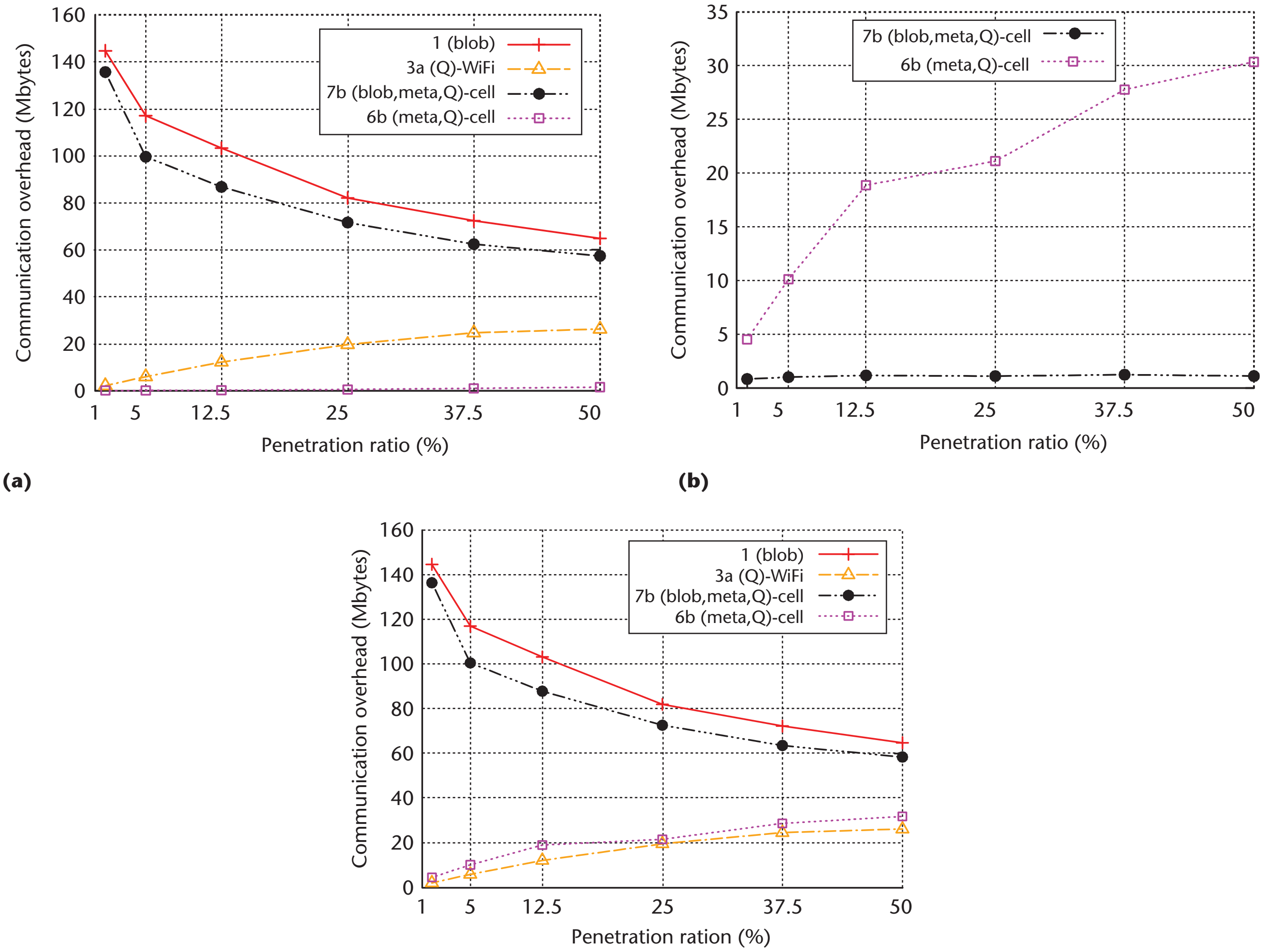
Penetration Ratio Impact (Figure 3a). When the penetration ratio increases, the answer throughputs of 1, 7b, and 6b initially increase and then decrease. Intuitively, when the penetration ratio increases, two effects are generated:

- The WiFi network becomes more connected, which pulls the answer throughput up.
- The contention and collisions increase for the WiFi network, which pulls the answer throughput down.

The answer throughput curves shown in Figure 3a result from the interplay of these two effects. For 6b, another factor contributes to the drop of the answer throughput drop. In 6b, answers are delivered only by cellular communication. The number of answers that can be delivered is thus limited by the cellular channel's capacity. 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 cellular channel's capacity is exceeded and only a fraction of discovered queries can be answered.

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

Impact of the query ratio (Figure 3c). The query ratio has little impact on the throughputs of 1 (blob) and 3a (Q)-WiFi. This is because these strategies use WiFi broadcasting to disseminate blob reports, and each broadcast satisfies multiple queries. The throughput of 6b



(c)

Figure 4. Comparison of communication overhead performance, heavy-congestion, using SWANS++. (a) WiFi communication overhead versus penetration ratio, heavy congestion. The report lifetime is 300 seconds, the query ratio is 0.25, and the blobs supply is four per second. (b) Cellular communication overhead versus penetration ratio, heavy congestion. The report lifetime is 300 seconds, the query ratio is 0.25, and the blobs supply is four per second. (c) Total communication overhead versus penetration ratio, heavy congestion. The report lifetime is 300 seconds, the query ratio is 0.25, and the blobs supply is four per second.

(meta,Q)-cell decreases with the query ratio. The reason, again, is that the cellular communication does not scale well to the increase in queries.

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

Simulation Results: Communication Overhead. Figures 4a and 4b show the WiFi/cellular communication overhead as a function of the penetration ratio; Figure 3a shows the throughput for the same configuration. As these figures show, 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. Configuration 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 that 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 3a). This fact suggests that 6b is efficient on bandwidth consumption.

Conclusion

Our simulations revealed that 6b has by far a higher throughput—up to seven times higher—than the other three. The 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. In the future, we plan to extend the study of P2P query processing paradigms in mixed communication networks—that is, networks with various combinations of capabilities such as short-range/anonymous, broadcast, and cellular. **MM**

Acknowledgments

This research was supported in part by the US Department of Transportation National University Rail Center (NURAIL); Illinois Department of Transportation (METSI); and National Science Foundation grants IIS-1213013, CCF-1216096, DGE-0549489, and IIP-1315169.

References

1. M. Guo, M. Ammar, and E. Zegura, "V3: A Vehicle-To-Vehicle Live Video Streaming Architecture," *Proc. 3rd IEEE Int'l Conf. Pervasive Computing and Comm.,* IEEE, 2005, pp. 171–180.
2. Y. Zhang, J. Zhao, and G. Cao, "Roadcast: A Popularity Aware Content Sharing Scheme in Vanets," *Proc. 29th IEEE Int'l Conf. Distributed Computing Systems,* IEEE, 2009, pp. 223–230.
3. U. Lee et al., "Dissemination and Harvesting of Urban Data Using Vehicular Sensing Platforms," *IEEE Trans. Vehicular Technology,* vol. 58, no. 2, 2009, pp. 882–901.
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 Trans. Vehicular Technology,* vol. 56, no. 6, 2007, pp. 3266–3277.
5. B. Han et al., "Cellular Traffic Offloading Through Opportunistic Communications: A Case Study," *Proc. 5th Workshop on Challenged Networks,* ACM, 2010.
6. S. Al-Chikhani, L. Al-Kanj, and Z. Dawy, "Video Distribution over Wireless Networks with Mobile-to-Mobile Cooperation," *Proc. Int'l Conf. Advances in Computational Tools for Engineering Applications,* IEEE, 2009, pp. 519–522.
7. M. Leung and S.-H. Chan, "Broadcast-Based Peer-to-Peer Collaborative Video Streaming Among Mobiles," *IEEE Trans. Broadcasting,* vol. 53, no. 1, 2007, pp. 350–361.
8. T. Zhong et al., "Trafficinfo: An Algorithm for Vanet Dissemination of Real-Time Traffic Information," *Proc. 15th World Congress on Intelligent Transportation Systems,* IEEE, 2008, pp. 1056–1061.
9. B. Xu, O. Wolfson, and J. Lin, "Multimedia Data in Hybrid Vehicular Networks," *Proc. 8th Int'l Conf. Advances in Mobile Computing and Multimedia* (MOMM), ACM, 2010, pp. 109–116.
10. T. Osafune, L. Lin, and M. Lenardi, "Multi-hop Vehicular Broadcast (MHVB)," *Proc. 6th Int'l Conf. ITS Telecomm.,* IEEE, 2006, pp. 757–760.
11. P.A. Hosein, "Capacity of Packetized Voice Services over Time-Shared Wireless Packet Data Channels," *Proc. 24th Joint Conf. IEEE Computer and Comm. Societies,* IEEE, 2005, pp. 2032–2043.

Ouri Wolfson is 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. His research interests include database systems, distributed systems, and mobile/pervasive computing. Wolfson has a PhD in computer science from Courant Institute of Mathematical Sciences at New York University. Contact him at wolfson@cs.uic.edu.

Bo Xu is 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. His research interests include mobile data management and computational transportation science. Xu has a PhD in computer science from the University of Illinois at Chicago. Contact him at boxu08@gmail.com.