Complex & Intelligent Systems (2022) 8:2593–2622
https://doi.org/10.1007/s40747-021-00629-x

SURVEY AND STATE OF THE ART

Data collection protocols for VANETs: a survey

Maryam Gillani[1] · Hafiz Adnan Niaz[1] · Muhammad Umar Farooq[2] · Ata Ullah[3]

Received: 8 July 2021 / Accepted: 17 December 2021 / Published online: 10 January 2022

Abstract
We live in the era of Intelligent Transport Systems (ITS), which is an extension of Vehicular AdHoc Networks (VANETs). In VANETs, vehicles act as nodes connected with each other and sometimes with a public station. Vehicles continuously exchange and collect information to provide innovative transportation services; for example, traffic management, navigation, autonomous driving, and the generation of alerts. However, VANETs are extremely challenging for data collection, due to their high mobility and dynamic network topologies that cause frequent link disruptions and make path discovery difficult. In this survey, various state-of-the-art data collection protocols for VANETs are discussed, based on three broad categories, i.e., delay-tolerant, best-effort, and real-time protocols. A taxonomy is designed for data collection protocols for VANETs that is essential to add precision and ease of understandability. A detailed comparative analysis among various data collection protocols is provided to highlight their functionalities and features. Protocols are evaluated based on three parametric phases. First, protocols investigation based on six necessary parameters, including delivery and drop ratio, efficiency, and recovery strategy. Second, a 4-D functional framework is designed to fit most data collection protocols for quick classification and mobility model identification, thus eradicating the need to read extensive literature. In the last, in-depth categorical mapping is performed to deep dive for better and targeted interpretation. In addition, some open research challenges for ITS and VANETs are discussed to highlight research gaps. Our work can thus be employed as a quick guide for researchers to identify the technical relevance of data collection protocols of VANETs.

Keywords Best-effort protocols · Delay tolerant protocols · Data collection protocols · Intelligent transport systems · Real-time protocols · VANETs

Abbreviations

UVA	Unmanned aerial vehicles
UVAR	UAV-Assisted VANETs routing protocol
FCD	Floating car data collection
TEM	Two Exponents Model
SOSM	Simple Obstacle Shadowing Model
AddP	Adaptive data dissemination protocol
ABCCM	Adaptive beacon congestion control mechanism
PGB	Partitioning gradient based
DOT	Distributed optimized time
RIDE	Real-time traffic Information aware data collection solution,
DDGP	Distributed data gathering protocol
CSMA/CA	Carrier sense multiple access/collision avoidance
DCMPTB	Data collection mechanism for smart grids using public transportation buses
SDVN	Software Defined Vehicular Networks
DRDCDA	Data relationship degree-based clustering data aggregation
PGB	Preferred group based
ZRP	Zone Routing Protocol

✉ Hafiz Adnan Niaz
hafiz.niaz@ucdconnect.ie

Maryam Gillani
maryam.gillani@ucdconnect.ie

Muhammad Umar Farooq
ufarooq@ceme.nust.edu.pk

Ata Ullah
aullah@numl.edu.pk

[1] School of Computer Science, University College Dublin (UCD), Dublin, Ireland

[2] College of EME, National University of Sciences and Technology (NUST), Islamabad, Pakistan

[3] National University of Modern Languages, Islamabad, Pakistan

DGRP	Data gather based routing protocol
RTAD	Real-time adaptive dissemination system
ECDGP	Extended cluster-based data gathering protocol
BSP	Transfer Utility of Node's Buffer Scheduling Strategy
S-GyTAR	Secure-Greedy Traffic-Aware Routing Protocol
ADCS	Adaptive Data Collection Scheme
CAC	Call Admission Control
OLSR-V2	Optimized Link State Routing- Version 2
TCDGP	Token based clustering data gathering protocol
TD-SDMA	Token based-space division multiple access
ADOPEL	Adaptive data collection protocol using reinforcement learning
HyBR	Hybrid bio-inspired bee swarm routing protocol
SeDyA	Secure dynamic aggregation
SAS	Secure Data Aggregation Scheme
CS-DC	Compressive sensing based data collection
ALCA	Agent learning–based clustering algorithm
VeMAC	MAC protocol for VANETs
OBV	OFDMA based MAC protocol for VANETs
DiPRoPHET	Distance-based probability routing protocol using history of encounters and transitivity
CDGP	Clustering data gathering protocol
QoI-DG	Quality of Information- Data Gathering
DB-VDG	Delay-Bounded Vehicular Data Gathering
SSS	Strategy Selection Algorithm
Sp-Cl	Spring clustering
HTAR	Hybrid Traffic-Aware Routing
SMITE	Stochastic compressive data collection
CMGM	Dynamic Clustering-based Adaptive Mobile Gateway Management
PBRS	Probabilistic Bundle Relaying Schemes
LBCA	Lane Based Clustering Algorithm
CGP	Clustered Gathering protocol
GyTAR	Greedy Traffic-Aware Routing Protocol
HVR	History-based vector routing
CASCADE	Cluster-based accurate syntactic compression of aggregated data
M-DMAC	Modified-Distributed and Mobility-Adaptive Clustering
PBS	Partitioning-Based Scheduling
WHA	Whale Optimization Algorithm
CCA	Cooperative Collision Avoidance

Introduction

VANETs are composed of vehicles with dynamic connections and rapidly changing, abrupt movements [1]. Different vehicles are connected in an ad hoc manner, with vehicles joining or leaving the group without generating any pre-indication [2]. A massive increase in vehicles has imposed significant challenges to fast, reliable and secure data collection [3]. Recently, vehicular data collection has attracted greater interest, due to increasing developments in Intelligent Transport Systems (ITS) [4]. VANETs are highly dependent on data exchanges, such as receiving data for traffic monitoring, notification of accidents, weather alerts, all aimed at maintaining a secure transportation system for the world [5]. VANETs integrate wireless LANs, ad hoc and cellular networks and differ from other ad hoc networks due to the hybrid nature of the network architecture, high node mobility and dynamic application scenarios.

Data collection is critical because vehicles are equipped with various sensors for speed, location, temperature and pressure estimation, as well as audio and video streaming support that demand intelligent data collection [7, 8]. Vehicles can share the data using Vehicle to Vehicle (V2V) or Vehicle to Infrastructure (V2I) communications. Source and destination vehicles can also exchange data, as shown in Fig. 1. The data collection process begins by taking data from various incorporated sensors and then performing aggregation and compression to reduce communication cost and time, depending upon the protocol and adapted scheme [9]. Data collection protocols are designed by considering both predictable movement patterns and abrupt variation in speed. Data is collected from distinct vehicles by involving intermediaries, where the shortest path is selected by means of routing protocols [10].

ITS can process the data to control environmental pollution, traffic prediction, accident analysis and road congestion control [11]. Data collection schemes and protocols in VANETs perform their task through deployment in cell phones or embedded systems integrated into vehicles to monitor sensor readings, while regularly fulfilling data collection requirements. Collected data is transferred after pre-determined intervals or in real-time to the data centres via backbone networks for further analysis, processing and storage [12]. Vehicles may also communicate directly with the Roadside Unit (RSU), or share data via a base station, as illustrated in Fig. 1.

Three kinds of data collection protocols exist in VANETs, i.e., Real-Time (RT) data collection protocols, Delay-Tolerant (DTN) and Best-Effort (BE) protocols. Real-time data collection protocols in VANETs are time-sensitive, and data must be collected and transmitted within a tolerable time-delay. However, real-time protocols are extremely sensitive to packet delays and packet loss. Therefore, efficient data dissemination is one of the most crucial elements to propel the desired flow of VANETs services, particularly related to real-time specifications. These protocols are essential in medical emergencies, security agencies, or when amass-

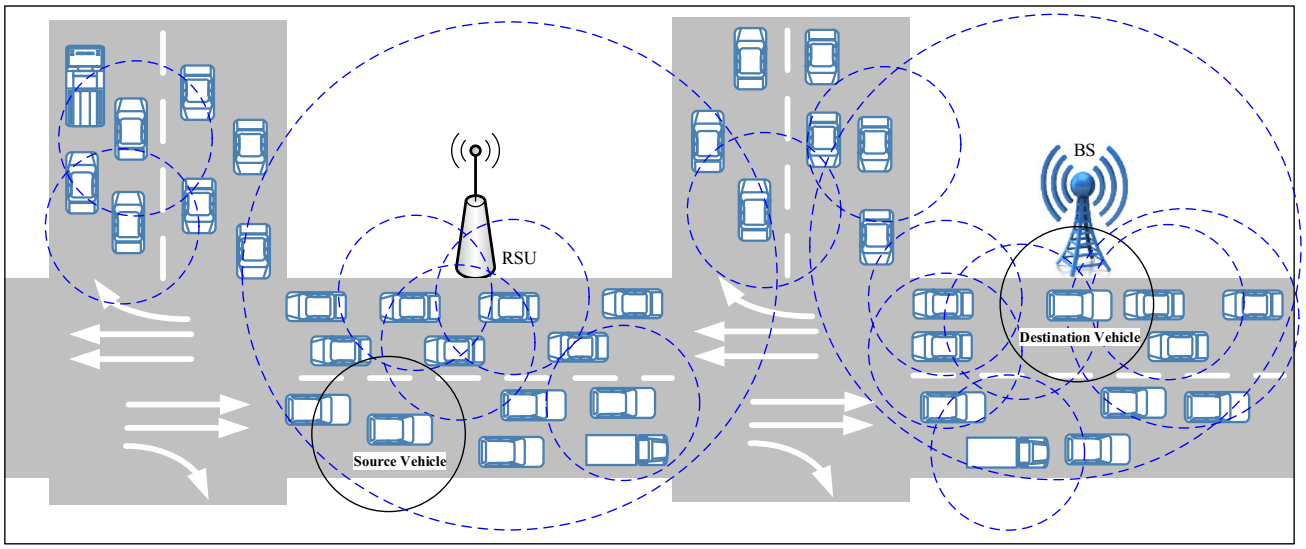

Fig. 1 VANETs architecture and data collection scenarios

ing sensitive military and defence-related information [13]. Real-time data collection is challenging due to the density of vehicles, dense road topologies, and high dependability requirements for a large set of intelligent applications, combined with cloud services [14]. Changes in traffic patterns, whether urban, rural or highway scenarios, also affect the functionality of real-time data collection protocols [15].

DTN-based data collection protocols in VANET can manage delays in receiving and sending data within relaxed and pre-defined thresholds [16]. In case of frequent network failures, DTN can achieve better performance than real-time scenarios [17, 18]. DTNs are well suited to highly mobile and terrestrial environments, assuring delivery of data by means of automatic store-and-forward mechanisms. Immediate data forwarding is also possible in a DTN when the required sources are available. However, the timeliness of data transmission is sometimes affected in DTNs, which is acceptable in specific applications in VANETs, such as in routine matters, weather prediction systems, audio/video streaming applications, underwater communication, and wildlife monitoring.

Best-effort protocols are designed to achieve the "best" possible attainable workload at a designated time with a probability of violation at run time [19]. These kinds of protocols do not provide guaranteed reliability and functional and offer no definite bounds on delivery time. Internet Protocol (IP) assists in a best-effort delivery system, trying to reduce data loss as much as possible, though data loss is inevitable in exceptional cases, such as network hardware failure. Different packets may take different routes throughout the network and be subjected to random delays. However, these kinds of protocols still work for scenarios where data timeliness and reliability are can afford to be somewhat compromised. Higher layer protocols are usually used to add reliability and cost-effectiveness to take full advantage of the network's capabilities.

Recent work over the past decade has considered new algorithms, protocol refinement and standardization, resulting in the IEEE 802.11p [20] and IEEE 1609 standards [21]. Due to the dynamic requirements of VANETs, an intelligent protocol for routing aware data collection is essential [22–24]. Routing involves the best path selection between source and destination. This affects timely data transmissions; optimizing data collection mechanisms alone does not guarantee timely delivery to the destination. Furthermore, routing can be expensive when two-way paths are set up among vehicles and the base station (BS) [25].

Optimized data collection is carried out using additional information, like vehicle location, direction and average road speeds, to find long-lived one-way data paths from data sources to the BS. In some cases, no data path exists between vehicles and the BS, and this can be resolved using the 'store, carry and forward' strategy, where vehicle data resides data at the link layer until delivered to the next hop towards the BS, as in a DTN approach.

Various state-of-the-art projects are based on data collections throughout the world [26]. These include The Car-2-Car Communication Consortium (C2C-CC) [27], which particularly focuses on improving road safety through the Cooperative Intelligent Transport System (C-ITS) [28]. Networks on Wheels (NOW) [29] adopts the same theme, while additionally considering security and Vehicle Infrastructure Integration (VII) [30–33]. Secure Vehicle Communication (SeVeCOM) [34, 35], Internet Intelligent Transport System Consortium [36, 37] and the Advanced Safety Vehicles Projects [38–40] are also among the highlighted projects.

The aim of this work is to explore, taxonomize, and discuss existing data acquisition techniques in VANETs. We are particularly interested in the communication and routing aspect of these techniques in DTN, best-effort, and real-time based VANET models. The motive is to provide an all-inclusive overview on data acquisition techniques to help users in making the right choice of the protocol. In literature, numerous works focus on summarizing the data acquisition techniques for vehicular ad hoc networks. In [41], the authors investigate topology-based, cluster-based, location-based and fog-based data collection techniques and highlight important issues that need to be addressed in data collection protocols for VANETs. In [42], prediction based protocols for vehicular ad hoc networks are summarized. The article highlights the results of traffic conditions, driving conditions and urban layout on the predictability of vehicle locations. A thorough investigative study and taxonomy of clustering protocols for VANETs is presented in [43]. The authors also provide a comparison of different parameters, including stability, density and convergence, for thorough understanding of VANET clustering algorithms. In [44], the authors provide a critical representation and taxonomy of named data networking-(NDN)-based data dissemination algorithms for VANETs. The authors provide a qualitative comparison on the basis of forwarding strategy, granularity, caching scheme and latency, etc.

However, one or more of the following limitations have been found in the existing works:

(1) The existing surveys are not comprehensive, and do not cover all types of data collection techniques;
(2) The existing surveys are not up to date and do not include recent works on data collection protocols;
(3) Application-specific analysis of data collection approaches is not provided.
(4) The existing surveys lack a standard classification/evaluation criteria, eventually ignoring application-specific analysis of data collection approaches.

In addition to addressing these limitations, this article makes the following contributions:

(1) It provides more profound knowledge and summarizes diverse sets of current and/or important data collection protocols, while covering two decades of advancements made in the area of data collection protocols. Specifically, the work includes DTNs, Best-effort, and real-time data collection protocols for VANETs.
(2) A Detailed taxonomy is provided for data collection protocols in VANETs with four-step hierarchal divisions. This taxonomy enables researchers to quickly get up to speed without digging into details and lengthy discussions (Fig. 2).
(3) An area-based critical analysis and comparison is performed by considering six fundamental parameters for relevant schemes.
(4) A 4-D functional framework is designed for explaining and analysing data collection protocols under standard application requirements. The proposed 4-D functional framework critically covers deep underlying factors to analyse various protocols presented in each category. It also accommodates almost every protocol to fit the VANETs category, due to its flexible and adaptable design.
(5) A detailed qualitative analysis is presented to guide readers on the merits and demerits of data collection protocols.
(6) Some open research challenges are covered with respect to recent literature with a detailed dendrogram that can act as a guide to potential research areas for researchers looking for research gaps.

To assist readers, a list of acronyms is provided in the appendix of this work.

The remaining part of the survey is organized as follows: "Delay tolerant network protocols" provides details of delay-tolerant network protocols, with related subsections (Random and probability, Flooding and variance, Geo-location, Movement vector), followed by detailed summary and analysis in table form. "Best-effort protocols" gives details of best-effort protocols, along with subsections for hierarchical protocols and flat data acquisition protocols. Hierarchical protocols are further categorized based on predictive, infrastructure, distributed, and zone-based clustering forms. In subsequent subsections, we also provide sub-classifications of Infrastructure based clustering while covering multi-hop, MAC-assisted (contention-based and contention-free) forms, followed by a detailed summary and supporting analysis in table form. "Real-time protocols" presents real-time protocols, with subsections related to cluster-based and position-based protocols, followed by a detailed summary and deep analysis in table form. "4-D Functional analysis" presents a multi-dimensional model, named 4D functional analysis, that covers further categorical analysis based on various parameters, both in table and text-based format. "Open research challenges" discusses open research challenges and "Concluding notes" concludes the survey.

Figure 2 illustrate the taxonomy of data collection protocols for VANETs. There are three main categories including delay tolerant, best-effort and real-time protocols. Delay tolerant protocols are sub-divided into random, flooding, geolocation and movement vector based protocols. Best-effort protocols for hierarchical category are sub-divided

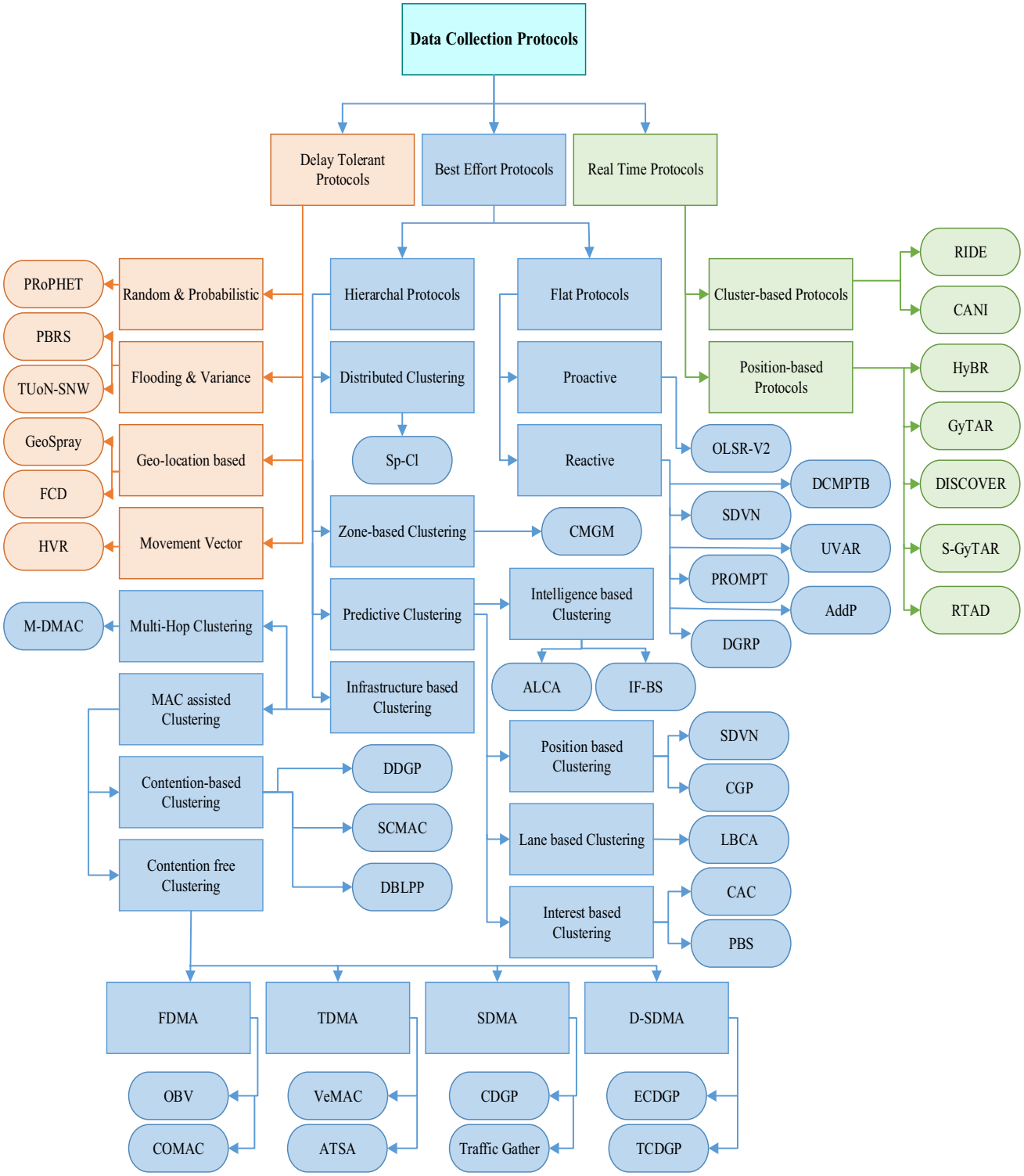

Fig. 2 Taxonomy of data collection protocols for VANETs

into predictive, infrastructure, distributed and zone-based clustering. Best-effort protocols for flat data collection protocols are sub-divided into reactive and proactive protocols. Real-time protocols are discussed and categorized based on cluster and position-based protocols. Aforementioned categories and subdivisions with associated examples are briefly discussed in their respective sections.

Delay tolerant network protocols

Networks with inconsistent connectivity and non-consecutive end-to-end paths between nodes are named Delay-Tolerant Networks (DTNs) [45]. DTNs follow a "store-carry-forward" mechanism to manage frequent connectivity disruption. For example, suppose a node receives data when there is no connection to transmit it: in that case, the message resides in a buffer until it recognises a chance to disseminate after establishing a connection. Now, the question arises, what are the best strategies for deciding about time quanta to hold the data and efficiently forward in a timely manner. For this purpose, dynamic and static network information can be used for decision making.

Dynamic network information includes area, collision and vehicular information, while the static network contains social connectivity details among nodes [45]. Both types of information play their role in seeking the best node and time to transmit messages. In VANETs, links are periodically disconnected to save energy or frequently fail due to high mobility. Ad hoc networks are decentralized types of networks that are infrastructure-independent and do not require any pre-existing infrastructure [46]. The lack of fixed infrastructure contributes to new research problems, such as network configuration, device discovery, topology maintenance, ad hoc addressing and self-routing [47]. DTNs are sub-divided into various categories, discussed below.

Random and probabilistic protocols

In random forwarding protocols, nodes forward data packets to other nodes that contact them in the first instance. Forwarding of packets starts with a random search for the destination. Sometimes, packets keep on moving within a specific set of nodes or may reach a dead end. Random forwarding is further divided into two categories, i.e., deterministic routing and stochastic DTN routing. In deterministic routing, knowledge of the current topology is extended future behaviour, and changes are predicted accordingly. Stochastic routing, on the other hand, is based on the unknown or random movement of nodes in which few or no future predictions can be made. Packet distribution in this scenario is achieved through random, hop-by-hop movements with a certain probability of getting to the destination; i.e., there is no guaranteed delivery.

In random probabilistic protocols, data is sent to the hosts in a specific order. This particular order is taken into account for hop counts and data delivery, based on previous encounters. Maurice et al. [48] proposed a Probabilistic Bundle Relaying Scheme (PBRS) for a two-hop vehicular DTN. This scheme probabilistically determines the carrying capacity of each vehicle according to the suitability of transmitted data packets. Data Bundles are given to the current vehicles only if doing so contributes to reducing the mean transit delay. PBRS is functional with minimal knowledge of the network.

Nidhi et al. [49] proposed a probabilistic relaying scheme (PRS) that is an extension of PBRS. It includes multi-copy vehicular DTNs [48]. In PRS, multiple copies are generated to increase the packet delivery ratio, while giving more benefits over multi-hop protocols. Transmitting multiple copies might increase data consumption but increases the packet delivery ratio. Anders et al. proposed a Probability Routing Protocol using History of Encounters and Transitivity (PRoPHET) [50]. This is a probabilistic routing protocol in which data from past encounters is utilized to optimize the packet delivery ratio. Better performance is attained by determining the next suitable hops for given packets. Distance-based PRoPHET (DiPRoPHET) [51], enhances (PRoPHET)'s protocol delivery ratio, as well as covering message delivery delay issues, by introducing it in a cross-layer process to retrieve the distance value from the lower layer and then use it in the upper layer.

Flooding and variance

Flooding is a strategy in which data packets keep replicating through a sufficient number of nodes until destination nodes receive them. Network knowledge can be utilized to select a subset of nodes to create a path and reach the destination. Each node is supposed to retain a copy or multiple copies of each message to transmit opportunistically. However, flooding causes network congestion and reduces the message delivery ratio. It also increases competition for network resources like bandwidth and storage.

Flooding can be classified into Single-Copy (SC) and Multi-Copy (MC) methods. In the single-copy form, a single data packet in the network is forwarded by various nodes, whereas in multi-copy, replicated data packets are forwarded through contact-based sharing [52]. Guizhu et al. [53] proposed a Transfer Utility of Node's Buffer Scheduling Strategy (BSP) to forward multiple copies dynamically. Amrita et al. proposed a Seasonality Aware Social (SAS) [54] forwarding technique that focuses on controlled forwarding through modelling contact history between node-pairs. It focuses on the weighted similarity index through repetitive contact patterns in real mobility traces via direct connections.

Geolocation-based protocols

Geolocation-based protocols are suitable when the source knows the coordinates of the destination node [6, 55].

Pierpaolo et al. [56] proposed a Floating Car Data (FCD) collection protocol for urban scenarios to provide connectivity through Dedicated Short Range Communication (DSRC) and cellular communication, such as Long-Term Evolution (LTE) offloading. Onboard LTE radio modules are consid-

ered as collecting vehicles. FCD is capable of collecting data directly from LTEs through individual LTE channels. In this scenario, delays are increased during dense cellular traffic. Moreover, Cooperative Awareness Messages (CAM) are periodically exchanged to report vehicle mobility and interaction scenarios.

Rui et al. [57] proposed a Gateway Location Awareness (GLA) scheme. This is a location-aware ranking classification that chooses vehicles with a higher tendency to forward information within a short span, while interpreting nodes according to moving patterns. GLA is combined with Aging Social Aware Ranking (ASAR) for improved performance. ASAR additionally allows the selection of a vehicle with more frequent connections, rather than selecting one with little or no connectivity with corresponding vehicles. This hybrid approach maximizes the data delivery with a lower data overhead.

Bilgin et al. proposed a Data Collection Mechanism for smart grids, using Public Transportation Buses (DCMPTB) [58] while using smart metres and smart grid communication systems. This protocol is designed to utilize I2V and V2V for data transfers from smart metres to public buses (I2V) and then moves from one bus stop to another bus stop through V2V. The source already knows the following coordinates of the destination node; i.e., buses know the next bus stop.

Vasco et al. originated the idea of GeoSpray [59], in which the hybrid approach is designed for single and multiple copy requirements. It follows asynchronous communication with the store-carry-forward mechanism. To exploit alternative paths, GeoSpray starts with a limited number of multiple copies and then switches towards a forwarding scheme to take the best possible advantage of all the vehicles in contact. In other words, GeoSpray uses two schemes in one design to gain maximum benefit from it.

Movement vector

Movement vectors specify the speed and direction of movement of a vehicle. They are shared by vehicles to update their current location. Position-based routing protocols use it to choose the shortest paths with low delays. It also helps to decide on path re-establishment and whether data packets should be replicated or not, according to link characteristics and vehicle mobility. Packets are replicated if a neighbour moves with high velocity and is close to leaving the group in a particular region. Packets are not replicated when vehicles are moving in the same direction towards the destination.

Hyunwoo et al. proposed History-based Vector Routing (HVR) [60] that allows each node to maintain the vector information of other encountered nodes, and then this information is shared to other nodes. While utilizing historic information, nodes start predicting the location of each packet's destination to achieve accurate forwarding. Zhaojun et al. proposed the Pass and Run protocol [61], which is specifically designed to protect the privacy of communication in DTNs. This protocol is tracking-resistant (i.e., does not allow tracking) and works through addressing the vehicle location and considering the driving patterns and history of vehicles to prevent misuse of information. Pass and Run uses greedy and random strategies to decide whether to submit the data packet to the RSU or to transfer it to the next vehicle.

A summary analysis of the aforementioned DTN protocols is given in Table 1, based on parameters like End-to-end delay, Average forwarded messages, Packet delivery ratio, Packet drop ratio, Recovery strategy, and Effect of traffic Density. End-to-end delay is the time taken by a data packet to traverse from a source to destination. It is calculated through $d = N*L/R$ (Packet of length L over N links with transmission rate R). 'Average forwarded messages' is defined as the number of messages forwarded at a given time to deliver a data packet. 'Packet delivery ratio' is the ratio of the number of packets initiated from the source and the number of packets received at the destination. 'Packet drop ratio' calculates the total number of data packets received at the destination divided by the number of data packets sent from the source. 'Recovery strategy' refers to the capability of the protocol to respond to unpredictable failure or collapse. Moreover, the recovery strategy illustrates if a 'plan B' exists to deal with accidental and unusual scenarios. The effect of traffic density on protocols' performance is another integral aspect in the analysis of performance and the success ratio. Traffic density indicates the number of vehicles present/interacting at a given time and location (road segment).

From Table 1, it can be seen that low end-to-end delay yields a high level of forwarded messages, better packet delivery ratio and low packet dropping ratio. Thus, it can be concluded that a high ratio of packet delivery indicates better performance. However, exceptions like Pass and Run [61], where the packet drop ratio is high, with medium packet delivery rate but high average forwarded messages indicate that different design constraints result in moderate results. Although the Pass and Run protocol tries to keep delay to a minimum by transmitting data packets to the nearest nodes, it does not assure the sender that the packet will be delivered to the destination in due course, because of unpredictable paths of vehicles. DTNs are thus better suited for weather prediction systems, underwater communication and wildlife monitoring, where some delay is tolerable and the value of the data and time are equally important; i.e., slight delays are fine as long as data is securely received in a cost-effective way.

Table 1 Summary of delay-tolerant network protocols

Protocol category	Protocol title	End-to-end delay	Average forwarded messages	Packet delivery ratio	Packet drop ratio	Recovery strategy	Effect of traffic density
Random and probability	PBRS	Low	High	Medium	Medium	No	High Traffic = High queueing delay
	PRoPHET	Medium	High	High	Medium	Yes	High Traffic = better performance
	DiPRoPHET	Low	Low	High	Low	No	Adaptable to traffic
Flooding and variance	BSP	Medium	Medium	High	Low	Yes	Adaptable to traffic
	SAS	Low	Low	Medium	Low	No	No influence of traffic density
Geo-location	FCD	Low	Medium	Medium	Low	Yes	High Traffic = better performance
	GeoSpray	Low	Medium	High	Medium	Yes	Adaptable to traffic
	DCMPTB	Low	Low	High	Low	Yes	High Traffic = High data transmission time
Movement vector	HVR	Low	High	High	Low	No	High Traffic = better performance
	Pass and run	Low	High	Medium	High	No	High Traffic = better performance

Best-effort protocols

Best-effort protocols try to achieve the "best" possible attainable workload per given time, with a probability of violation at run time [62]. These kinds of protocols seek to maximize application benefits by meeting most of the requirements. They do not claim complete reliability, which is supposed to be provided by the higher layer protocols, but to deliver packets towards a destination within designated time constraints.

Hierarchical protocols

Hierarchical protocols are distributed at multiple levels of clustering, along with sub-groups [43, 63]. They tend to manage the assigned tasks individually. Data being forwarded to one group is not necessarily given to another group. These protocols are applicable for wide-area grouping of vehicles, where standalone sub-groups can also be set up along with a Cluster Head (CH). Each CH exchanges the information to member vehicles in a hierarchy.

Predictive clustering

Predictive clustering utilizes the recent geographic positions, specific interests and predicted future behaviour of vehicles to structure a cluster [64]. Clusters and cluster-head selection are based on predictable movements of vehicles [65], assigning priorities for controlled access of cluster formation. Although vehicles keep on changing their positions, they are somewhat detectable due to the routes of roads [66]. Saliha et al. proposed Fitness Clustering [67] based on rapid and real heuristics. It primarily targets data dissemination in emergency cases.

In Fitness Clustering, the original message is optimized to reduce the number of exchanged packets. It focuses on making stable clustering by considering parameters like transmission period, the degree of connectivity, relative velocity and the lifetime validity of the link. Islam et al. proposed a Prediction-Based Efficient Clustering Scheme (MPECS) [68] that uses a Voronoi Diagram to divide the area into distinct regions and then allow every vehicle to decide its own longevity and cost the cluster head in its cur-

rent area. This technique evaluates the vehicle's impact on clustering stability and cost, so that more extended cluster stability comes with minimal overhead and cost.

Intelligence-based clustering Intelligence-based clustering maintains the hierarchy to eradicate unbalanced cluster formation by utilizing machine learning and artificial intelligence [69, 70]. In these protocols, CH election is mainly performed after cluster formation [71]. Neeraj et al. proposed an Agent-Learning-based Clustering Algorithm (ALCA) [72] to eradicate the issues related to high density, random mobility and finding an exact route. Agents learn from the deployed environment, where neighbouring agents also collaborate for information sharing, and estimation of vehicles is maintained through clustering. The CH is elected through node density and the direction of vehicle mobility.

Reward or penalty functions are suggested based on various parameters, such as the agent's ID, action set, learning rate, and learning factor. Learning agents decide to increment or decrement the parametric values through these functions until maximum values are achieved. Manisha et al. proposed an intelligent forwarding–based stable and reliable data dissemination scheme (IF-BS) [73]. IF-BS works intelligently to let vehicles decide the next forwarding node by considering the stability of connecting edges and waiting for metrics. When the next node is assured from source to destination, less link disruption occurs, and more data delivery is assured.

Position based clustering In position-based Clustering, position coordinates of vehicles and CHs are the main consideration for clustering. Cluster structure depends on the vehicle's geographic positioning, and its CH is elected based on priorities associated with vehicle requirements [74]. Stable CHs are preferred, and their stability is evaluated through various factors, including a more extended trip of the vehicle, high speed and proximity to the base station (BS) [75]. These clusters can manage the rapid movement of vehicles and are considered a key clustering protocol for VANETs.

Ismail et al. [76] proposed a geographical Clustered Gathering Protocol (CGP) where the CH performs data collection, aggregation and dissemination. It then transfers this data to a sink or BS. CGP uses an opportunistic approach, with a store-and-forward mechanism that is used when the next road segment is empty, and the vehicle has to wait for the CH to come closer to it. CGP works on single-way communication and is applicable to single and straightforward road topologies. However, it is not easy to manage for more extensive and complex regions. Cluster management overhead is also a constraint associated with CGP.

Position-based routing can be managed efficiently by utilizing the RSUs, BSs and smart vehicles with internet access to directly access the servers to reduce the communication overheads associated with V2V messaging. Shahab et al. proposed Probabilistic Direction Aware Cooperative Collision Avoidance (P-DACCA) [77], which estimates the probability of a collision based the expected state of nodes. Through this calculation, an early warning is generated when the probability exceeds a predefined threshold. This factor avoids upcoming threats of collisions and reduces the number of collisions, but also reduces communication overhead, as well as giving low latency.

Zhenzehn et al. proposed the Software-Defined Vehicular Network (SDVN) [78], the first-ever algorithm that successfully utilized cooperative cellular and ad hoc network accesses for extensive data collection. After SDVN, multiple protocols have been designed and suggested that follow the same mechanism. SDVN possesses high monetary cost, but it fulfils the data delivery ratio through a cellular predictive process. It takes a predictive decision based on real-time network status, rather than empirical knowledge.

Rakesh et al. proposed the Data Relationship Degree-based Clustering Data Aggregation (DRDCDA) [79], which is based on universal delegate sensing vehicle selection. This kind of selection is used as a unique factor to calculate the vehicle data and then measure a correlation with data from neighbouring vehicles to perform local cluster formation.

Lane based clustering Lane based clustering protocols estimate the road lanes, based on the traffic flow of vehicles. In some schemes, it is assumed that each vehicle knows its lane, while other schemes often consider a virtual lane to assume lane-based connectivity. The CH is elected based on levels given in a lane in relation to other lane nodes. For example, a vehicle with a high lane level will be elected as CH, and so on.

Mohammad et al. [80] proposed a Lane-Based Clustering Algorithm (LBCA) as a stable clustering method where each vehicle identifies its lane using a lane detection system. Lane analysis is performed using weight-based metrics for right lane, left lane and no turn. It shows a longer CH lifetime than the Lowest-ID, Highest-Degree and Utility Function algorithms.

Interest-Based clustering Interest-based clustering utilizes the area of interest for specified intentions of getting results. It is a "concern centred" technique that is suitable when data collection and delivery is desired for a specified area of particular interest. It is primarily applicable in emergency scenarios [81]. For example, from the complete road topology data, it extracts data where an accident or emergency occurs. Thus, it achieves high efficiency with low communication costs.

Yaoyao et al. [82] discussed a Partitioning-Based Scheduling (PBS) algorithm that utilizes mobile devices as mechanical information carriers in partitioned networks. PBS stores the partitioning of nodes and cluster formations in a KD-Tree, where powerful nodes maintain records [83].

Tarek presented a hybrid, dynamically allocated resource policy for quality of service and fair data packet scheduling via a Call Admission Control (CAC) scheme. The CAC scheme dynamically utilizes vehicles and vehicle density transmission powers to provide the desired throughput for real-time communication. The performance of vehicular communication is enhanced when all stations are considered greedy to transfer packets. For contention-based channel access, a back-off mechanism is used to provide fairness among hostile 802.11p users of the same access category [84].

Ghada et al. proposed All-Member-Interests-based Merging (AMIM) [85], in which decisions are taken for the benefit of all members of clusters, rather than only considering the CH. AMIM considers vehicle speed, position and direction, focussing on Link Expiration Time (LET) and Signal-To-Noise Ratio (SNR). AMMI works together with Double-Head Cluster (DHC), where two functioning CHs are selected to overcome frequent re-clustering. When a Cluster member loses its connection with its CH, the alternative CH catches it to carry on processing data delivery in the area of interest.

Infrastructure-based clustering

Infrastructure-based clustering is about cluster formation based on a partially defined infrastructure of clusters for the communication process [86]. Clustering focuses on a low degree of velocity and high node connectivity for cluster leadership [65]. CH is elected based on a vehicle's relative velocity [87]. This scheme also plays a vital role in preventing cluster re-ordering when two CHs come within range simultaneously.

Multi-hop clustering Cluster formation is accomplished through multiple hop distances, where every node is considered a maximum of K-hops away from at least one CH [88]. It is significant in extending cluster sizes and reducing the number of cluster heads. Two primary factors are the number of K-Hop neighbours and identification IDs. Grzegorz presented a Modified Mobility-Adaptive Clustering (M-DMAC) [89] for high-mobility nodes, where a vehicle with the highest neighbouring ratio is selected as CH. M-DMAC is a modified version of DMAC in which a generic clustering is used, which is not reliable for dealing with a changed mobility pattern. M-DMAC focuses on avoidance of re-clustering by incrementing the stability of clusters through estimating connection time for moving nodes, i.e., freshness checking.

MAC-assisted clustering MAC-assisted clustering uses link-layer information for cluster formation. The topological insights available to the link-layer help in selecting CHs with minimum relative speeds and distances to all neighbours. For instance, a point coordinator (access point) in the IEEE 802.11 family can be considered as a CH. Similarly, parameters like neighbour count, mean squared deviation in received signal strengths, and relative distances to the registered RSUs can be used in a fully distributed MAC. However, MAC-assisted clustering increases link-level traffic, which also raises the probability of collisions in contention-based MAC protocols. Collisions trigger retransmissions at application and/or transport layers. As a result, lower transmission efficiency is achieved. This issue is addressed through the reduction of channel contention for timely and reliable message delivery [87]. In addition, these protocols are less affected by variation in vehicle speeds [88]. CSMA/CA-based protocols are inherently designed to wrestle collisions. In CSMA/CA, every associated terminal should be able to detect the transmission of all other terminals [91]. However, not all packets transmitted from different terminals can be sensed, due to the hidden node problem, mobility, and various other infrastructure-based obstacles. This factor negatively affects CSMA's performance. Increased collisions result in extended delays, rescheduling of transmissions, data loss, and wastage of resources. To better utilize CSMA performance on VANETs protocols, various modifications have been proposed that are discussed thoroughly in this section.

Contention-based protocols

Contention-based protocols are flexible and responsive to the dynamic nature of networks with appropriate intermittent and short message sharing. These protocols are reliable for sharing safety messages [90]. They allow multiple users to utilize the same channel without predefined coordination. Bouziane et al. discussed a Distributed Data Gathering Protocol (DDGP) [91] that uses vehicles and mobile collectors for data collection. DDGP enables vehicles to access the channel in a distributed way, based on their location information. The efficiency of the protocols is increased by removing expired and redundant data. A Col packet is sent that contains the length of the collection area, data packet type, acknowledgement packet, and announcement packet. Data collection is accomplished through these parameters, along with segmentation and clustering [92]. Two segments, i.e., Collection Segments (CS) and Silence Segments (SS), increase efficiency. However, these segments cannot deal with hurdles or blockages in road topology, such as road blockage in accidents.

Contention-free protocols

Contention-free protocols require centralized scheduling by proper allocation of resources, i.e., time slots, channels and positioning of nodes, to avoid collision [90]. These protocols consider time-synchronization of nodes that is not applicable for large-scale VANETs. It results in slow responses to distributed networks. For example, multiple

access techniques, Time Division Multiple Access (TDMA), result in low throughput with collision-free medium access. Multiple Access Techniques become problematic in low traffic loads due to idle slots [93].

TDMA-based systems encounter issues in the synchronization of nodes due to rapid topology change. Secondly, changing time slots in a decentralized scenario is also a tedious task to perform. Hassan et al. proposed a TDMA-based protocol, VeMAC [94], that exclusively targets hidden terminal problems through single- and multi-hop broadcast services at the control channel level. VeMAC eliminates transmission collisions through excessive node mobility, which avoids collisions by allocating disjointed sets of time slots to nodes in opposite directions compared to roadside units. This characteristic makes VeMAC favourable for attaining higher throughput at the control channel.

In Frequency Division Multiple Access (FDMA), a collision-free medium-access technique handles communication with different radio channels [95]. However, it increases the cost of sensor nodes, because each channel is assigned to only one user at a time. Therefore, Alessandro et al. [96] proposed an orthogonal FDMA-based Obvious (OBV) protocol that uses carrier frequency and available bandwidth through control channels. OBV is divided into segments, or frames, with each frame separated by a contention period (CP) and a contention-free period (CFP). The CP is retrieved through a contention-based algorithm and utilized to exchange resources to be used in the CFP for data transmission in the specified resources.

In Spatial Division Multiple Access (SDMA), a geographical zone is partitioned into multiple divisions and then mapped to the respective channel. It allows scheduled access time slots based on a vehicle's location on the road topology. Although it does not maximize the usage of available bandwidth, it is a widely accepted technique for stable resource allocation for vehicular mobility [97]. However, SDMA has the least applicability in multi-hop message delivery. In SDMA-based protocols, a time slot is given that is also called the allowed time for a vehicle to transmit data.

Bouziane et al. proposed a Clustered Data Gathering Protocol (CDGP) [98] that minimizes the extent of collisions in a highly dense network and enhances the robustness and reliability of data collection. Its clustering technique is based on a hybrid architecture, a data collection phase through Dynamic SDMA (D-SDMA) and a retransmission mechanism to handle faulty messages. CDGP comprises three main tasks, i.e., propagation of collection messages through a Road Side Unit (RSU) for initiation of the data collection process; formation of clusters, along with CH election, in each collection segment; and a data collection phase. The CH in each segment allocates a time slot to each block of road containing a vehicle. In the propagation of collection messages phase, the RSU starts the collection process by dispatching a beacon packet (Col) that contains the RSU position (RSU-POS), length (A), direction (DC), data type (DT) and validation time (VT).

In the cluster formation process, clusters and CH are formed within the validation time. Second, if the CH allocates a time slot to an empty block, then the whole duration of the slot is wasted, thus increasing the waiting time. Another protocol based on CDGP is Extended Cluster-Based Data Aggregation (ECDGP) [99]. ECDGP is proposed to be applicable on DTN, as well as Real-time Scenarios, with additional features of supporting multiple data types and aggregation of data before delivering them to the initiator. Moreover, it offers flexible data collection through aggregation and segmentation. The retransmission mechanism is developed to ensure reliability.

Wang et al. proposed TrafficGather [100], adapting the same SDMA concept. TrafficGather divides roads into road blocks with separate clusters. This protocol allows each vehicle to transmit traffic information at a designated time slot. The only drawback of this protocol is that TrafficGather is limited in that a large number of time slots are lost when they are allocated to empty cells, especially in a sparse network. Furthermore, the use of a flooding strategy during the last phase may cause the 'broadcast storm' problem.

In Dynamic Spatial Division Multiple Access (D-SDMA), reallocation and retransmission functions for erroneous data packets are further included [91]. Bouziane et al. [101] presented a Token-based Clustered Data Gathering Protocol (TCDGP) one year after the previous presented CDGP; TCDGP is meant to overcome gaps in the earlier protocol.

TCDGP is slightly different from CDGP in its functionalities. It inherits all the characteristics of CDGP and adds one, i.e., token-based dynamic SDMA (TD-SDMA). In TD-SDMA, each CH periodically sends a token packet with two fields: a Block_num of the packet intended for sending and an Ack field (a single bit) that is used as receipt of data to retransmit the data in case of error. TCDGP allows reservation of a time slot only for vehicles having data to send, thus resolving a slot wastage problem. On the other hand, the TCDGP protocol has more message overhead, due to transmitting a token packet to each block segment to assign time slots.

Distributed clustering

Distributed Clustering Protocols are designed to control CH allocation in a distributed environment [102]. In distributed clustering, the number of links increases with the number of channels, while making inter-cluster communication more effective. However, it shows reduced effectiveness in realistic vehicular speed scenarios, due to high transmission overhead and low vehicle density situations. Distributed clustering is also known as decentralized clustering, because of its more insufficient cluster connection time and cluster stability.

Inter cluster links initially increase due to overlapping regions of vehicles, but keep on declining and disconnecting because of rapid vehicle movement. Transmission efficiency becomes affected by the rapid vehicle movement and makes it more decentralized.

Leandros et al. [103] proposed a distributed Spring Clustering (Sp-Cl) scheme for stable clustering. It focuses on making fewer clusters as compared to other lower ID clustering schemes. Cluster stability is measured through cluster configuration against vehicle mobility. Nodes keep on joining and leaving the clusters with Sp-Cl dealing with vehicle transitions among clusters by reducing re-clustering.

Oliveria et al. proposed Adaptive Data Dissemination Protocol (AddP) [104], a multi-hop broadcasting protocol that deals explicitly with the high-density area, following distributed clustering. In AddP, CH is selected based on vehicles' position and velocity, thus making relay selection dependent upon density and distance factors. It dynamically adjusts the periodicity of beacon messages, while reducing the communication overhead.

A modified version of AddP is later proposed as Optimal Adaptive data dissemination Protocol OAddP [105], which deals with different traffic flow, utilizing prediction-based decision-making schemes to generate clusters and disseminate data. Selo et al. proposed a distributed approach based on Coalitional Game Clustering (CGC) [106] that allows every vehicle to make a distributed cluster with other vehicles based on coalition value. This value is based on connection lifetime and speed difference among vehicles. In CGC, distributed clustering only requires values like link quality and speed of neighbouring vehicles. This approach helps in achieving a high SNR with balanced distributed clusters.

Zone-based clustering

Zone-based clustering works based on the highest residual energy in each zone specified for providing location and detecting objects for real-time reporting [107]. It is concerned with the formation of clusters in different zones. These zones can be formed based on different parameters, i.e., N-Hop neighbours, interest group, energy, etc. Data acquisition schemes may vary within zones, outside zones and among different zones. Clusters are formed based on zone interests. In zone-based clustering, the clustering process is relatively controlled, and this ultimately makes the CH election and packet delivery ratio effective [108].

Abderrahim et al. [109] proposed a Clustering-based, Multi-metric adaptive mobile Gateway Management Mechanism (CMGM) for a VANET-3G integrated network architecture that further uses the concept of mobile gateways. CMGM works for clustering gateway candidates, in which the CH acts as a gateway to interface VANET with the 3G environment.

Brendha et al. proposed a Zone-Based Cluster Head for Routing (CZCHR) [110]. Forwarding collector packets are generated from one road end to another, based on various parameters like buffer queue, length and link lifetime. Despite sending forwarding collector packets (FCP) in all nodes in the zone, it follows the traffic-aware technique to send FCP when the nodes in the zone leave from the current road to the intersection. This strategy makes it lightweight, less crowded and energy efficient.

Flat data collection protocols

All nodes connected to a network are treated as equally operational on a flat topology in flat data collection protocols. A sink is supposed to receive data from sources through multi-hop paths [111]. Flat protocols are also known as homogeneous protocols because all nodes have the same capabilities. Flat data acquisition protocols can be implemented in small networks because small networks give better results with a flat topology [112]. Reactive routing protocols maintain only active routes, unlike conventional routing protocols. Routes are maintained for the nodes that are currently being used for sending data packets.

Nodes are supposed to take path information from packet headers or their internally maintained routing tables. In larger and mobile networks, reactive routing is suitable, especially in VANETs [113]. Anjana et al. [114] proposed a Data Gathering based Routing Protocol (DGRP). Without maintaining a routing table, it allows the adoption of changes to opt for the best possible choice by considering Quality of Service parameters. All routing paths are created through source nodes present in the network. After collecting all the information, source nodes send a query to destinations through the network system. Complexity and overhead requirements for a distributed location database service can be considered as DGRP constraints.

Omar et al. [115] proposed an intelligent Unmanned Aerial Vehicle Assisted Routing protocol (UVAR) for urban VANETs. UVAR improvises data routing and vehicle connectivity through an aerial UAV, while targeting the ground only when the network is poorly dense. Furthermore, UVAR targets forwarding data packets through aerial vehicles to the ground reactively to outperform conventional V2V communications. Thus, it is functional in both environments, i.e., in the ground for improving data delivery efficiency and in the sky for transmission of data packets using reactive routing.

UVAR is remarkable in re-establishing communication links, along with re-linking disconnected road segments. However, UAVs, despite being efficient, add extra costs for batteries, fuel, and maintenance. This protocol compliments UAV-UAV communication and UAV-to-Ground-Vehicles to retain diverse information about the connectivity status [116]. This additional cost makes it difficult to incorporate UVAR

as a better solution. Moreover, UVAR does not utilize GPS information and trajectory calculation during route discovery and data forwarding.

Boangoat et al. designed a protocol named PROMPT [117], which is a cross-layer, position-based, delay-aware communication protocol. PROMPT works on positions independently of vehicle movement and relies on vehicle monitor information exchange statistics that help in selecting the most suitable paths. In contrast, proactive protocols are functional on shortest-path algorithms; these are table-based protocols and keep all the required information of connected nodes in tables. Tables are usually shared with neighbouring nodes, so that every change must be also updated to other nodes [118]. In VANETs, proactive protocols are not suitable, due to rapid changes in positioning. Moreover, consumption of more bandwidth and large table sizes of information make proactive protocols inappropriate for VANETs. Wassim et al. proposed a proactive protocol named Adaptive Data Collection Scheme (ADCS) using 3G/LTE [119]. ADCS being a proactive protocol, ended up giving a high packet loss ratio.

A summary of Best-Effort Protocols is given in Table 2. This summary illustrates that best-effort protocols are better suited for high traffic density due to their adaptability to traffic conditions [76–79]. Second, unlike DTN protocols [48, 51, 60, 61] the average levels of forwarded messages in best-effort are either low or medium for the majority of protocols [72, 73, 79, 83, 84, 94, 96, 98–101, 104, 105, 110, 114, 117, 120] but still give a better packet delivery ratio: this is another positive factor of best-effort protocols. A better packet delivery ratio, along with a lower packet drop ratio, distinguishes them from DTNs. Furthermore, best-effort protocols are significantly higher performers in terms of flexibility in adaptability to traffic density. However, a recovery strategy is less well followed in best-effort protocols due to the nature of their operations; i.e., they are neither time-bound nor guarantee reliability. Best-effort protocols are thus considered suitable when the value of cost and space are more important than timely and reliable data collection and delivery, as, for example, in retrieving data about weekly accident cases on a specific road for future traffic analysis [121].

Real-time protocols

In real-time protocols, the value of data decreases rapidly with time and limits the tolerable delay [122]. The value of data is of prime concern because of VANETs' high dependence on current data to make decisions, to improve user safety, traffic flows and to assist auto driving. Data delivered after the designated time is least helpful in rapidly changing traffic conditions. Outdated or delayed data does not offer the desired real-time traffic monitoring which is essential aspect of the smart traffic navigation services currently used by millions of drivers. Thus, real-time protocols are significant and better suited for analysing current and live traffic conditions. In this case, primitive sensor readings are not required to be stored at the device, reducing storage overhead. However, data being sent to a network can be relatively large and, thus, the communication overhead increases. In addition to time constraints, the diversity of areas for information extraction and dynamic path selection are also critical factors for evaluation. We have categorized the protocols as follows.

Cluster-based protocols

Clustering involves the grouping of nodes according to density, velocity, position, and geography. Due to frequent mobility in VANET, clustering algorithms perform dynamic restructuring of connectivity patterns among neighbouring vehicles. It works on a few clusters to have more control over its structure without exceeding a communication overhead. These protocols work on a virtual backbone infrastructure to efficiently deliver and collect data in VANETs [11, 123].

Zongjian et al. proposed a Real-time traffic-Information-aware Data Extraction (RIDE) [122] scheme for satisfying data collection time constraints. It treats data collection as a schedule optimization problem and proves it to be an NP-complete problem. It is a real-time traffic adaptive data collection protocol that considers the criticality of time to minimize the data transmission overhead [122]. RIDE is based on a Dynamic Programming (DP)-based solution designed to manage small-scale data collection where only a small number of vehicles are involved. DP gives an optimal solution in which one problem can be divided into subproblems, whose solution can be memorized to substitute for future values instead of recalculation. RIDE further uses a Genetic Algorithm (GA) for substantial road segments and complex situations. GA works by choosing a random neighbour of the source in the next layer. After that, it assigns data forwarding counts randomly according to the constraints. The iteration keeps on repeating until it reaches the BS, thus achieving real-time data collection [122].

Clustering Adaptation Near Intersection (CANI) [85], uses Online Sequential Extreme Learning Machine (OS-ELM) to let vehicles continuously learn and update in real-time to predict behaviour and adapt clustering near intersections to accomplish data collection.

Position-based protocols

In position-based protocols, each vehicle keeps track of its neighbouring vehicles by periodically sending beacon messages [124]. It maintains dynamically updatable data storage at the sink, which causes communication overhead. The sender traces the position of the destination by utilizing

Table 2 Summary of best-effort protocols

Protocol category	Protocol name	End-to-end delay	Average forwarded messages	Packet delivery ratio	Packet drop ratio	Recovery strategy	Effect of traffic density
Intelligence based	ALCA	Low	Medium	High	Low	No	High Traffic, high connectivity preservation ratio
	IF-BS	Low	Low	High	Low	Yes	High Traffic = High Latency, High PDR
Position based	SDVN	Low	High	High	Medium	Yes	Adaptable to traffic density
	P-DACCA	Low	Low	High	Low	Yes	Adaptable to traffic density
	DRDCDA	Low	Medium	High	Medium	No	Adaptable to traffic density
Lane based	LBCA	Medium	Medium	High	Low	No	Medium traffic density = better performance
Interest based	CAC	Medium	Medium	High	Low	No	Adaptable to traffic density
	PBS	Medium	Low	High	Low	No	Adaptable to traffic density
Multi-hop	M-DMAC	Medium	High	Medium	Low	No	High traffic = longer CH connection
Mac assisted	DDGP	Low	Medium	High	Low	No	Adaptable to traffic density
Mac assisted (contention free)	OBV	Low	Medium	High	Low	Yes	Adaptable to traffic density
	VeMAC	Low	Medium	Medium	Medium	Yes	High traffic = High data collisions
	CDGP	Medium	High	Medium	Medium	No	High traffic = % of retransmission increases
	TrafficGather	Medium	Low	High	Low	No	High traffic = mass transmission redundant data
	TCDGP	Medium	Medium	High	Medium	No	High traffic = High efficiency
	ECDGP	Low	Medium	High	Low	No	Adaptable to traffic density
Distributed clustering	Sp-Cl	Low	High	High	Medium	Yes	Adaptable to traffic density
	CGC	Low	High	High	Low	Yes	Adaptable to traffic density
	AddP	Low	Low	High	Medium	No	Adaptable to traffic density
	OAddp	Low	High	High	Low	No	High density = Better performance
Zone based clustering	CMGM	Low	Medium	High	Medium	Yes	Adaptable to traffic density
	CZCHR	Medium	Medium	High	Medium	No	Adaptable to traffic density
Flat data collection protocols	SDVN	Low	High	High	Medium	Yes	Adaptable to traffic density
	PROMPT	Low	Medium	High	Low	No	Adaptable to traffic density
	DGRP	Medium	High	Medium	Medium	No	High density, complexity, routing overhead
	DCMPTB	Low	Low	High	Low	No	Better traffic density, smart metres

coordinates. Salim et al. has explored a Hybrid Bee Swarm Routing (HyBR) protocol based on a continuous learning paradigm for maximum data packet delivery with minimum delay. It combines two routing methods, i.e., topology and geography-based routing, where the former works through sending beacon messages where all nodes are informed of their neighbours and activated links. Each node possesses its routing table, which contains various routes toward the desired destination. In the latter case, a fitness function is given in which optimal route discovery is made by selecting mutation operators, parents and crossovers, according to the geographical information of the vehicles [125]. The main problems in HyBR are route poisoning and outdated information in the routing table.

Ion et al. proposed a protocol named DISCOVER [123] that collects the data in a large city area using a single network structure, i.e., a multi-hop made up of vehicles only. DISCOVER is distributed and adapted for different traffic densities and traffic conditions in a real-time manner. There are two designated waves, i.e., a forward wave and a reverse wave. The forward wave is meant for dissemination, and the reverse wave works for data collection. In this protocol, FCD works for periodic delivery of vehicular data via the RSU.

Tarek et al. proposed the Secure-Greedy Traffic-Aware Routing protocol (S-GyTAR) [126]. Real-time traffic evaluation is performed to identify malicious nodes and thus stop them from forwarding data. This protocol continuously monitors traffic for secure data communication in real-time scenarios. In S-GyTAR, CH evaluates the trustworthiness of cluster members through Reputation Model (RM). This protocol is a modified version of GyTAR (a position-based protocol) [127], in which data is sent through the network, intersection by intersection, until it reaches its final destination. Although a part of this protocol relies on clustering, it is best suited in position-based protocol due to its ability to position suspicious vehicles and then stop them from functioning ahead.

Julio et al. presented a Real-Time Adaptive Dissemination (RTAD) [128] that allows each vehicle to automatically adopt the best-suited dissemination scheme for specific situations. RTAD utilizes parameters like vehicle density and topological characteristics. As a result, more vehicles are informed through fewer messages, thus mitigating broadcasting storms.

A summary of real-time protocols is presented in Table 3. In real-time protocols, the levels of average forwarded messages are relatively high, which ensures guaranteed packet delivery with a lower packet drop ratio. One of the critical aspects of real-time protocols is the presence of recovery strategies in their design that makes them preferable when dealing with any accidental situation when it is important to deliver data within a designated time. Real-time data collection protocols add considerable cost for the functioning of the network, but facilitate updated, timely, and reliable data transfers. We can see in Table 3 that end-to-end delay is low for all protocols, with high levels of forwarded messages. However, exceptions like [122, 125] exist that give high packet delivery ratios with low average forwarded messages and low packet drop ratios. Real-time protocols are now considered through hybrid and intelligent operational domains that facilitate low forwarded messages with a better delivery ratio. [122, 125] are designed based on artificial intelligence approaches that shift the paradigm gradually from high-cost protocols to cost-effective protocols.

Another possibility of real-time protocols that can be seen from Table 3 is their adaptability to deal appropriately with different traffic patterns. However, real-time protocols can be compromised occasionally by the packet dropping ratio [123, 126, 127] and high number of forwarded messages [85, 123, 126–128], and their high cost, but these protocols cannot compromise over data delivery. Thus, the packet delivery ratio cannot be low in real-time protocols to ensure guaranteed data delivery within the specified time. These protocols are ideally considered for defence services and emergency applications, where the value of data is more critical then cost.

4-D Functional analysis

A functional framework is designed to analyse data collection protocols based on four characteristics: node mobility, protocol class, protocol type and data acquisition initiation, as illustrated in Fig. 3. We analyse the protocols according to these four characteristics, as explored in Table 4. Node mobility identifies whether a protocol supports static and mobile sinks or sensors. It has been observed that most of the protocols follow the pattern of mobile sensor, static sink (MS) or mobile sensor, mobile sink (MM). Only PBS [83] and SpCl [103] follow the static sensors, mobile sink (SM) option, where the mobile data collector takes the data from sensing points that collect the data from vehicles. No one adopted the static sensors, static sink approach, as vehicles are in motion, together with sensors. The next characteristic, titled "protocol class", is used to identify the category of the protocol, such as flat protocol (FP), hierarchical or cluster-based (CB) and semi-structured (SST), which are hybrid.

The next characteristic, "Protocol Classification", identifies the suitability of the protocol for Real-Time (RT), Delay Tolerant (DT) or Best-Effort (BE) scenarios as shown separately in each category. The "Data Initiator" for data acquisition initiation can detect when the sink node pulls data as "Sink Pull (PS)" or data is pushed by the source node as "Sensor Push (SP)". Most protocols follow the SP approach, which is adopted to continuously push the sensing data to central repositories. However, it may cause a bottleneck, and

Table 3 Summary of real-time protocols

Protocol category	Protocol name	End-to-end delay	Average forwarded messages	Packet delivery ratio	Packet drop ratio	Recovery strategy	Effect of traffic density
Real-time clusters	RIDE	Low	Low	High	Low	Yes	Adaptable to traffic density
	CANI	Low	High	High	Low	No	High density, better clustering
Real-time position based	HyBR	Low	Low	High	Low	Yes	Adaptable to traffic density
	GyTAR	Low	High	Medium	Medium	Yes	Adaptable to traffic density
	DISCOVER	Low	High	High	Medium	Yes	Adaptable to traffic density
	S-GyTAR	Low	High	High	Medium	Yes	High density = increased overhead
	RTAD	Low	High	High	Low	Yes	Adaptable to traffic density

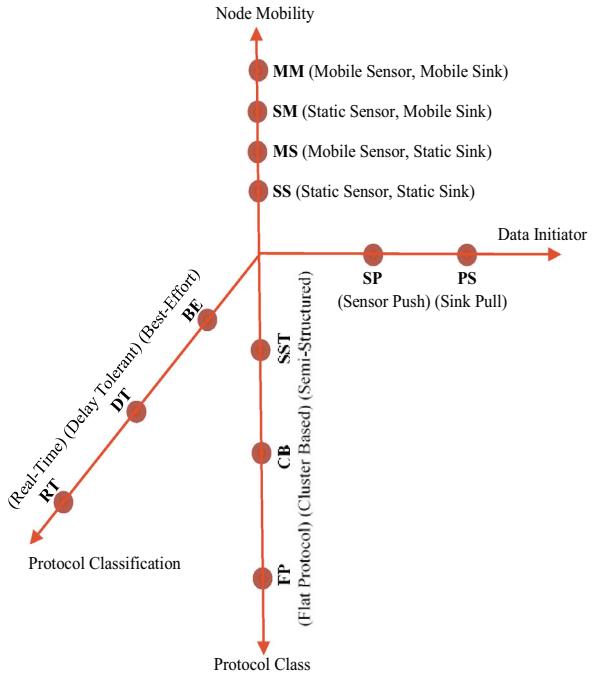

Fig. 3 4-D Functional Framework

data may be lost during transmission in case of congestion. PS-based approaches are used when the query forwarded to all sensors is weak, resulting in loss of critical data.

Table 5 explores a comparison of data collection protocols that are categorized as DTN, BE or RT. It explores the system model, proposed technique, and related metrics. It shows how a protocol is good in privacy protection, but not energy-efficient, and similar combinations of advantages and limitations. We also consider efficiency, latency, motion estimation, area of application, environment, and mechanism. It has been observed that the majority of protocols are designed and tested to be functional in an urban environment. This gives irregular patterns, with a high density of vehicles, along with a high ratio of traffic jams, accidents, and greater chances of being attacked by malicious nodes. This is why the urban environment gives a more challenging environment to newly developed protocols, and thus most researchers choose it for beneficial results and analysis.

The protocol's system model or deployment environment is based on the urban (U) or Highway (H) scenario. In the next column, we identify the mechanism adopted in the protocol. Next, the clustering and cross-layer support is identified as Yes or No. The source of delivery is mostly Vehicle (V), but other options are Transport Buses (TB), CANI and UAV in [58, 78] and [115], respectively. It can be noticed that protocols involved in multi-tasking and which cover multiple attributes in addition to data collection are likely to be less energy efficient [49, 56, 59, 115]. This is because more energy is consumed if a protocol is dealing with multiple attributes.

Another critical aspect of the protocols designed in VANETS is that not every protocol developed for them uses vehicles as the source for data collection. For example, unmanned aerial vehicles [115] and public transport [58] are relatively new trends to consider for data collection in VANETs. Next, we explore whether the protocols consider privacy protection and what type of routing approaches are used, selecting from Reactive (R), proactive (PR) or hybrid (HY). In the following column, we note that most of the protocols adopted the AWS mobility model, and a number of simulation tools are also presented. Moreover, the protocols

Table 4 **4**-D Functional mapping on data collection protocols

Protocol category	Protocol name	Node mobility	Protocol classification	Protocol class	Data initiator
Best-effort protocols					
Intelligence based	ALCA	MM	BE	CB	PS
	IF-BS	MM	BE	FP	SP
Position based	SDVN	MS	BE	CB	SP
	P-DACCA	MM	BE	CB	SP
	DRDCDA	MS	BE	CB	SP
Lane based	LBCA	MM	BE	CB	SP
Interest based	CAC	MS	BE	CB	PS
	PBS	SM	BE	CB	PS
M-Hop	M-DMAC	MM	BE	CB	SP
MAC (contention based)	DDGP	MS	BE	CB	SP
Mac Assisted (contention free)	OBV	MM	BE	FP	SP
	VeMAC	MS	BE	CB	SP
	CDGP	MS	BE	CB	SP
	TrafficGather	MM	BE	CB	SP
	TCDGP	MS	BE	CB	SP
	ECDGP	MS	BE	CB	PS
Distributed clustering	Sp-Cl	SM	BE	CB	SP
	CGC	MM	BE	CB	SP
	AddP	MM	BE	CB	SP
	OAddP	MM	BE	CB	SP
Zone based clustering	CMGM	MS	BE	CB	SP
	CZCHR	MM	BE	CB	SP
Flat data acquisition protocols	SDVN	MM	BE	FP	SP
	PROMPT	MM	BE	FP	SP
	DGRP	MS	BE	FP	PS
	DCMPTB	MS	BE	FP	SP
Real-time protocols					
Cluster based	RIDE	MS	RT	CB	PS
	CANI	MM	RT	CB	SP
Position based	HYBR	MM	RT	SST	SP
	GyTAR	MM	RT	SST	SP
	DISCOVER	MM	RT	SST	SP
	S-GyTAR	MM	RT	CB	SP
	RTAD	MM	RT	SST	SP
Delay tolerant networks protocols					
Random and probability	PBRS	MM	DT	SST	SP
	PRoPHET	MS	DT	SST	SP
	PRS	MM	DT	SST	SP
	DiPRoPHET	MS	DT	SST	SP
Flooding and variance	BSP	MM	DT	SST	SP
	SAS	MS	DT	SST	SP
Geo-location	FCD	MM	DT	SST	SP
	GeoSpray	MS	DT	SST	PS

Table 4 continued

Protocol category	Protocol name	Node mobility	Protocol classification	Protocol class	Data initiator
	GLA	MM	DT	SST	SP
	DCMPTB	MS	DT	CB	PS
Movement vector	HVR	MM	DT	SST	SP
	Pass and run	MS	DT	SST	SP

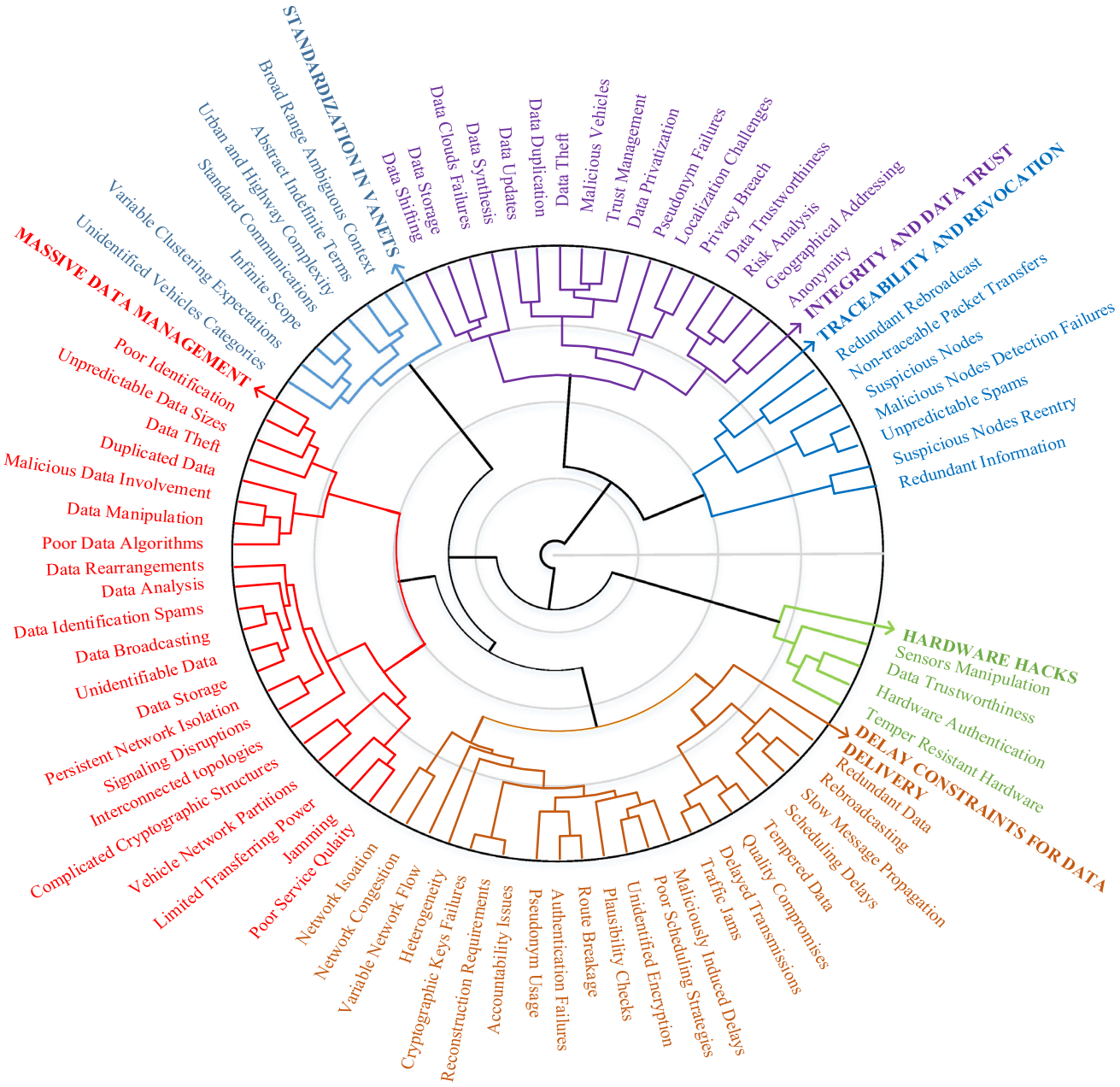

Fig. 4 Dendrogram of open research challenges for VANETs

are also analysed to present position verification, latency, and motion estimation support. These matrices are essential for designing a dependable, optimised, adaptive and scalable protocol for data collection in VANETs (Fig. 4).

Table 5 Categorical comparison of data collection protocols

Protocol title	Environment	Mechanism	Clustering	Cross layer	Source of delivery	Energy efficiency	Privacy protection	Routing approach	Simulation tool	Protocols standardization	Layered perspective	Position verification	Latency	Motion estimation
Delay tolerant network protocols														
SAS [54]	U	Katz	✘	✓	V	✓	✘	PR	ONE simulator	A	CL	✓	✓	✓
PRS [49]	U	GBRS	✘	✘	V	✘	✘	R	Java simulator	A	NL	✓	✓	✓
GLA [57]	U	ASAR	✘	✓	V	✓	✓	HY	Emulator	I	NL	✓	✓	✓
Pass and Run [61]	U	GCF, R	✘	✘	V	✘	✓	HY	SUMO	A	CL	✓	✓	✓
FCD [56]	U	–	✓	✓	V	✘	✘	HY	OMNET++, SUMO	I	CL	✓	✓	✓
DCMPTB [58]	U,H	–	✘	✓	TB	✓	✘	R	NS-2	A	CL	✓	✘	✓
BSP [53]	–	Spray, Wait	✘	✓	–	–	✘	–	ONE simulator	A	CL	✓	✘	✓
DiPRoPHET [51]	–	–	✘	✓	–	✓	✘	–	NS-2	A	CL	✘	–	✓
PBRS [48]	H	GBRS	✘	–	V	✓	–	–	Java simulator	A	NL	✓	✓	✓
GeoSpray [59]	U,R	–	✘	–	–	✘	✘	HY	VTDNsim	I	AL	✘	✓	✓
HVR [60]	–	–	✘	–	–	✘	–	–	NS-2	A	NL	✘	✓	✓
DB-VDG [129]	U	SSS	✘	✓	V	✓	✘	–	NS-3.6	I	CL	✓	✓	✓
ADOPEL [130]	H	–	✘	✓	V	✓	✘	R	MATLAB	A	CL	✓	✓	✓
Best-effort protocols														
BE AMIM [131]	U,H	DHC	✓	–	V	✓	✘	R	SUMO	I	LL	✓	✘	✓
P-DACCA [120]	U	CCA	✓	✘	V	✓	✘	R	NS-2	A	LL	✓	✓	✓

Table 5 continued

Protocol title	Environment	Mechanism	Clustering	Cross layer	Source of delivery	Energy efficiency	Privacy protection	Routing approach	Simulation tool	Protocols standardization	Layered perspective	Position verification	Latency	Motion estimation
IF-BS [73]	U	GF	✘	✘	V	✓	✘	R	NS-2.35, SUMO	A	LL	✓	V	✓
OAddP [105]	U,H	WHA	✓	✘	V	✓	✘	R	NS-2, SUMO	A	LL	✓	✓	✓
CZCHR [110]	U,H	–	✓	✘	V	✓	✘	R	NS-2	A	PL	✓	✘	✓
CGC [106]	U,H	CGT	✓	✘	V	✓	✘	R	SUMO, Matlab	A	PL	✓	✓	✓
MPECS [68]	U	–	✓	✘	V	✓	✘	HY	SUMO, NS-2	A	LL	✓	✓	✓
Fitness Clustering [123]	–	–	✓	✘	V	✓	✘	HY	–		–	✓	✘	✓
UVAR [115]	U	–	✘	✓	UAV	✘	✘	HY	NS-2	A	CL	✓	✓	✓
AddP [104]	U,H	ABCCM	✘	✓	V	✓	✓	HY	OMNET + + , SUMO	I	CL	✓	✓	✓
DDGP [91]	U,H	CSMA/CA	✓	✓	V	✓	✘	PR	OMNET + + , SUMO	A	CL	✓	✓	✓
SDVN (Preco) [78]	U	–	✓	✓	CANI	✓	✓	PR	SUMO	I	CL	✓	✓	✓
DRDCDA [79]	H	–	✓	✓	V	✓	✘	HY	NS-2	A	CL	✓	✓	✓
ZRP [132]	U,H	–	✓	✘	–	✓	–	HY	NS-2, Vanet Mobisim	A	LL	✓	✓	✓
DGRP [114]	U,H	–	✘	–	V	✓	✓	R	OPNET	A	PL	✓	✓	✓
ECDGP [99]	U,H	DSDMA	✓	–	V	✓	✘	HY	testbed in C + +	A	CL	✓	✓	✓
ADCS [119]	–	–	✘	–	V	✓	✘	PR	NS-3, SUMO	A	LL	✓	✓	✓

Table 5 continued

Protocol title	Environment	Mechanism	Clustering	Cross layer	Source of delivery	Energy efficiency	Privacy protection	Routing approach	Simulation tool	Protocols standardization	Layered perspective	Position verification	Latency	Motion estimation
CAC [84]	H	–	–	✓	V	✓	✓	–	NS-2	A	CL	✓	✓	✓
OLSR-V2 [133]	U,H	–	✘	–	–	–	–	PR	–	A	LL	–	–	✓
TCDGP [101]	H	TD-SDMA	✓	✓	V	✓	✘	PR	Java based simulator	A	CL	✓	✓	✓
SeDyA [134]	H	PCSA	✓	✘	V	✓	✓	HY	JiST/SWANS	A	PL	✓	✓	✓
CS-DC [135]	H	–	✓	✓	V	✓	✓	–	NS-2, MOVE	A	CL	✓	✓	✓
ALCA [72]	U,H	AC	✓	–	V	✓	✘	–	MobiSim	A	CL	✓	–	✓
VeMAC [94]	U,H	TDMA	✘	✓	V	✓	–	–	Matlab	A	CL	✓	✓	✓
OBV [96]	U,H	–	✘	✓	V	✓	✘	–	VISSIM, SHINE	A	CL	✓	✓	✓
CDGP [98]	H	D-SDMA	✓	–	V	✓	✘	HY	Java based simulator	A	CL	✓	✘	✓
QoI-DG [136]	–	ODAR	✘	_	V	✓	✘	HY	SUMO	I	CL	✓	✘	✓
Sp-Cl [103]	U.H	–	✓	–	V	–	–	–	Customized	A	CL	✓	–	✓
SMITE [137]	–	–	✓	–	V	✓	✘	–	NS-2	A	CL	✓	✓	✓
CMGM [109]	U,H	–	✓	✓	–	✓	–	HY	NS-2	A	CL	–	✓	✓
PROMPT [117]	U	–	✘	✓	V	✓	✘	–	NS-2	A	CL	✓	✓	✓
LBCA [80]	U,H	–	✓	–	V	✓	✘	–	NS-3	I	CL	✓	✘	✓
CGP [76]	U	CSMA	✓	✓	V	✓	✘	HY	Qualnet 4.5	A	CL	✓	✘	✓
TrafficGather [100]	–	SDMA	✓	–	V	✘	✘	–	Qualnet	A	CL	✘	✘	✓

Table 5 continued

Protocol title	Environment	Mechanism	Clustering	Cross layer	Source of delivery	Energy efficiency	Privacy protection	Routing approach	Simulation tool	Protocols standardization	Layered perspective	Position verification	Latency	Motion estimation
CASCADE [138]	H	–	✓	✓	V	✓	✓	HY	ASH	A	CL	✓	✓	✓
M-DMAC [89]	U	–	✓	–	V	–	–	–	JiST/SWANS++, Vanet MobiSim	A	CL	✓	–	✓
PBS [83]	U	–	✓	–	–	✓	–	–	C++	A	CL	–	✓	✓
MobEyes [139]	U	–	✘	–	V	✓	✓	PR	NS-2	I	NL	✓	✓	✓
Real-time protocols														
RT GyTAR [127]	U	GCF	✘	✘	V	✓	✘	HY	Qualnet	A	PL	✓	✓	✓
RIDE [122]	U,H	–	✓	✘	V	✓	✘	PR	NS-3, SUMO	A	CL	✓	✓	✓
DISCOVER [140]	U	–	✓	✓	V	✓	✘	HY	OMNET++, SUMO	I	CL	✓	✓	✓
RTAD [128]	U	CA	✓	✓	V	✘	✘	HY	NS-2, SUMO	I	CL	✓	✘	✓
S-GyTAR [126]	U	–	✓	–	V	✓	✓	R	NS-3	I	NL	✓	✘	✓
HyBR [125]	U,H	–	✘	–	V	✓	✓	R	NS-2	I	NL	✓	✘	✓
CANI [85]	U	OS-ELM	✓	✘	V	✓	✘	R	SUMO, TRACI, NS-3.26	I	NL	✓	✘	✓
HTAR [141]	–	–	✓	–	V	✓	✘	HY	NS-2, TraNS	A	CL	✓	–	✓

U Urban, *CANI* Cellular and Ad-hoc Network Interfaces, *GCF* Greedy carry-and-forward, *DSDMA* Dynamic Space Division Multiple Access, *PCSA* Probabilistic Counting with Stochastic Averaging, *GBRS* Greedy Bundle Relaying Scheme, *ODAR* optimization of data aggregation and routing, *SSS* Strategy Selection Algorithm, *WHA* Whale Optimization Algorithm, *CCA* Cooperative Collision Avoidance, *ASH* Application-aware SWANS (Scalable Wireless Ad hoc Network Simulation) with Highway mobility, *UAV* Unmanned Aerial Vehicle, *CSMA/CA* Carrier Sense Multiple Access/Collision Avoidance, *CGT* Coalition Game Theory, *OS-ELM* Online Sequential Extreme Learning Machine, *I* Industry, *A* Academia, *NL* Network Layer, *H* Highway,✓ Yes, ✘ No, *H* History, *PR* Proactive, *HY* Hybrid, *V* Vehicles, *AC* Agent Control, *TB* Transportation Buses, *RWS* Real world scenario, *GF* Greedy Forwarding, *FMM* Freeway Mobility Model, *DHC* Double-Head Clustering Algorithm, *CL* Cross Layer, *AL* Application Layer, *PL* Physical Layer, *LL* Link Layer

In Table 5, parameters like Energy Efficiency, Privacy Protection and Routing Approach are amendable to improve protocols performance. Energy Efficiency as key component of data collection protocols is relevant to protocol categories as well. For example, DTN protocols are not energy efficient mostly. This is due to the fact that while waiting for the data to be delivered, vehicles keep on transferring the data to other vehicles and sometimes retain it to themselves while using massive memory units. That's why, energy efficiency largely compromised over DTN. On the other hand, Best-Effort protocols are mostly energy efficient in comparison with DTN and Real-Time protocols, because of their nature of collecting data only when it is feasible without replicating and transferring it unnecessarily.

Another critical aspect is, Real-Time protocols are also energy efficient due to their time bound characteristics. They generate high targeted average forwarded messages to reduce regenerations and avoid keeping data for longer period of time. Privacy protection is a matter of overall protocol priorities. To save protocol resources and completion time, researchers avoid privacy protection. That's why, irrespective of category only 12 protocols out of almost 60 protocols has incorporated privacy protection factor. Routing approaches are mainly Hybrid, Reactive, and Proactive. The choice of approach is also based on protocols need to interact with other vehicles and BS. For example, in Table 5, Hybrid, Proactive, and Reactive are randomly distributed among different categories.

Layered perspective deals with identification of protocols on the basis of cross layer, network layer, physical layer, application layer, and link layer. Majority of protocols are cross-layer protocols because cross-layer design allows protocol to share and exchange network information among different layers. This quality ensure the best route selection by considering energy consumption as well as other performance requirements. Network layer protocols are better in handling the routing and sending the data between different networks. Link layer protocols are operational only on local network segment (link). Application layer protocols are mostly shared communication protocols that defines how application processes among clients and servers.

As per above mentioned discussion, we can interpret the factor like energy efficiency, routing approach and privacy protection can be amended, but change in one factor may result in change of protocol category as well. A protocol can be converted to energy efficient, but while increasing efficiency we may limit the delivery time. Hence, changing DTN or Best Effort to Real-time protocol. Similarly, to add privacy protection we may need to compromise on time and efficiency and a Best effort may switched to DTN resultantly. A slight change in one parameter influence others and create changes in protocols classifications.

Protocols standardization indicate whether the protocols have been standardized in industry or just proposed by academia. 74% protocols are academically researched and proposed and 26% protocols are practically implemented in various industrial scenarios and actual cities. Academically proposed protocols are well-supported and tested through virtual environments created by SUMO and Mobisim. Protocols that are actually implemented as a data collection solution for real environment of cities/counties are the ones that are industrially standardized. As per stat, there are lesser amount of protocols that are standardized in industry i.e. 26% and more protocols are merely proposed in academia.

It is likely that energy and time efficiency is not swiftly being implemented among actual cities. This trend is expected to be changed and future may bring more industrially supported protocols as interest in smart cities and IoV based projects are considerably rising. An important factor related to standardization is lack of standardizing bodies. There are few standardization bodies i.e. Internet Engineering Task Force (IETF) that work for internet routing protocol standardization criteria and document internet standards for routing criteria. There are no prominent VANETs standardization bodies as such that particularly deals with Intelligent Transport systems, VANETs and particularly data collection protocols for VANETs. Protocols that are implemented in cities and counties are approved through IETF, IEEE and related umbrella standardization bodies.

Open research challenges

VANETs have drawn remarkable interest in both industrial and academic sectors due to their potential applications and services. The boom in self-driving cars and other prediction-based traffic services has significantly increased the demand for improvement in VANETs. However, locating vehicles' positioning, maintaining, and interpreting an exact view of the entire network, a high number of nodes, rapidly changing node mobility, swift topological changes, and frequent network disconnections add potential challenges in the area of VANETs.

Designing energy-efficient and cost-effective communication approaches for data collection is a dire need at this time [142, 143]. VANETs are challenging for data communication because of frequent node disruptions, high node density, and limited infrastructure availability to cope with the change [144, 145]. Furthermore, the intervention of suspicious and malicious vehicles affects normal vehicle operations in terms of data collection and makes it a critical task [146]. Data is collected from all kinds of vehicles together, raising authenticity concerns and adding doubt in analysing it. Therefore, different protocols are being designed nowadays to meet the

challenging data communication needs in the best possible manner [147].

When a protocol is developed, researchers focus on single or multiple objectives. Some of them focus on making lightweight protocols, and others target energy efficiency to save considerable space. It is nearly impossible to incorporate all the desirable factors in one protocol without compromising any of the factors of storage, time or efficiency. In other words, there are always trade-offs.

Other challenges include geographical mapping and addressing, risk management and trust analysis, data confirmation and authenticity evaluation, inter-vehicular and intra-vehicular communication baselines, reliability checks, data prioritization, addressing and monitoring issues, privacy and anonymity, real-time changes and protocol competency to deal with real-time changes and so on. The challenges are infinite and exploit different parameters, making VANETs vulnerable to various attacks. A few of the latest and most crucial research challenges are discussed here.

Hardware hacks

VANETs are directly and indirectly dependent on hardware, as vehicles use hardware-based sensors for speed, temperature, location and various other devices for monitoring and security purposes. Manipulation of sensors by hardware hackers might alter actual and real-time data to deliver fabricated information to alter planned routes and affect vehicles' navigation for effective data communication. Attackers find it easy to exploit hardware along the path of the data to the destination; it passes through various nodes assisted by hardware, thus creating chances to over-write it. Hardware authentication is challenging in VANETs because of the millions of hardware devices involved at any given time. Data trustworthiness becomes complicated when an infinite number of devices communicate concurrently to effect the desired data collection. Hardware hacks for safety–critical systems are even more critical to handle [148]. Lack of tamper resistance (intentional causing of malfunctions by users) in sensors and devices allows exploitation of physical access to system and vehicular products. Although there are many data protection techniques for VANETs, energy-efficient hardware protection from hacks is still an area to be resolved [23].

Delay constraints for data delivery

Due to rapidly changing traffic conditions, data needs to be delivered within a specific amount of time; updated data is most credible and valuable for different real-time applications. Security-based data is especially important to receive within a certain deadline due to its high sensitivity, costs, and the risk factors associated with it. The time factor is incorporated with private data to give the least possible time for attackers to exploit data authenticity [149]. The contamination of redundant data, along with rebroadcasting, affects data delivery and leads to unforeseeable delays. Delay is purposefully added through slow message propagation to limit or pause the next data initiation phase, which results in scheduling delays. With redundant, rebroadcasted and outdated data, come quality compromises and delayed transmissions.

Maliciously induced delays significantly damage traffic routes, causing jams and poor monitoring. Plausibility checks also occasionally add delays; in other words, while checking quality reasonability, delay happens. Unidentifiable encryption, route breakage, accountability issues, encryption failures and authentication failures are other critical causes of delays in data delivery. Heterogeneous networks collaborating to accomplish data communication is also another leading aspect of extreme network congestion and network isolation that requires reconstruction of network scheduling strategies.

When the desired information takes a more extended period than expected or exceeds the allowed threshold, this indicates the theft of data by attackers before it could reach the required destination [147]. This problem arises in data communication because of rapidly changing traffic conditions and other factors, such as link disruption or road jams, resulting in delays in data delivery, thus giving more time to attackers for data theft. The delaying factors mentioned above are areas that need considerable attention to fully explore the desired solution and ensure timely data delivery [144].

Massive data management

A considerable amount of work has been done for data communication in nearly all possible scenarios, but the management of the enormous data blocks coming from a massive number of vehicles that are sending data at high rates per second is still challenging. Such massive data management becomes even more complicated with poor identification of data sources. Unpredictable data sizes, duplicated data, malicious data involvement, and manipulated data also make data management a complex task. Duplicated and manipulated data considerably increase data sizes, adding to the complexity of data handling. Data analysis sometime requires data rearrangements upon locating, for instance, identification spams, broadcasting failures, and unidentifiable data.

Data management with signalling disruptions, interconnected topologies, and complicated cryptographic structures hinders the smooth functioning of data algorithms. The exploitation of the data already saved is subject to saving data in such a way as to avoid data theft, protect saved data blocks, and allow changing and overwriting data without creating opportunities for manipulation by potential attackers. There is a need for algorithms to deal with a massive amount of data storage units. These concerns are the least answered

questions and still give researchers space to investigate these areas to drive potential solutions [150].

Standardization in VANETs

When the term 'vehicle' is mentioned in this context, it does not only indicate a normal car used for personal tasks. The word 'vehicle' is a broader term that includes public buses, trains, electric scooters, motorbikes, six-wheelers, loaders, taxis, and ambulances—some of the vehicles better suited for urban scenarios and some for highways. There are even some categories of vehicles that are not yet identified; unidentified vehicle categories with broader and ambiguous contexts make it complicated to properly define the functionalities and scope of VANETs. For example, electric two-wheelers are not allowed on some motorways and big highways. A taxi might give less link disruption than a privately owned car that spends less time on the road. What are the standards for all vehicles to develop a data communication protocol for variable clustering requirements [77]?. The word 'vehicle' itself is a generic and broad term, and standardization of these small things can significantly impact the design of protocols and algorithms.

Integrity and data trust

VANETs are highly dependent upon V2V and V2I communication. At this time, there is a critical need to establish protocols to resolve data integrity issues in both V2I and V2V communication [150]. When data is transferred hop by hop (vehicle to vehicle), data must be received from source to destination without any alteration [151]. Malicious vehicles try to drop or fabricate received data and then send manipulated data without anyone even knowing about suspicious activities [152].

An efficient and effective detection scheme to deal with such V2V frauds is needed to handle fabrication and modification in a real-time manner, or preferably before it happens [153]. Data trust can be exploited during data cloud failures, especially during storage, synthesis, updates and movement. Data duplication also challenges integrity, along with various other trust management concerns, e.g., localization challenges, privacy breaches, or anonymity with geographical addressing.

Traceability and revocation

Extensive and remarkable work is being carried out to catch malicious and suspicious nodes during data communication, dissemination, and routing processes [154]. However, the least focus is given to permanently blocking malicious nodes from contacting the network again. In other words, protocols work to highlight malicious nodes, but do they keep data of those malicious nodes to permanently block their access to the same network again? Every time a malicious node is detected and rejected, it has to pass through the same process repeatedly to get access. Eventually, the protocol will be detecting it frequently to stop its possible attempt. Researchers can maintain a database of malicious nodes to save time, energy, and effort. Moreover, non-traceable packet transfer with redundant information and re-entry of suspicious nodes makes this area of VANETs yet to be explored and challenging at the same time.

Concluding notes

VANETs have been critically challenged recently, due to their extensive applicability in ITS, Internet of Vehicles IoV and growing interest in smart cities. Data collection has been a widely studied aspect of VANETs for secure and smooth communication flows. Data collection protocols are paramount for ITS and IoV, regarding efficiency, efficacy, time and cost-effectiveness. This paper encapsulated a detailed overview of data collection protocols based on three primary categories of VANETs, i.e., DTN, BE, and RT. We examined these protocols with a structured taxonomy design to provide broader insight into categories, sub-categories and relevant supporting examples suitable for data collection in VANETs. Each technique is thoroughly interpreted and investigated, based on various evaluative parameters, such as End-to-End Delay, Packet Delivery Ratio, Packing Drop Ratio, Average Forwarded Messages, Recovery Strategy, since these parameters affect the success of a VANETs.

Later, we supported our comparative analysis via a 4-D functional framework comprised of four integral data collection areas, i.e., Node Mobility, Data Initiator, Protocol Class and Protocol Classification. Our proposed 4-D functional framework can categorize any VANET-based protocol without requiring researchers to pass through extended literature readings. Finally, a comparison table of data collection protocols with different evaluative parameters is provided to assist users in determining the better choice based on suitability and credibility. Due to the diversity of data collection protocols, we have mapped them in various general categories for feasibility and ease of understandability. This diversity of data collection protocols demands selection criteria based on mechanism adapted, network layers, routing approach, latency, privacy approach, motion estimation and sources of delivery used. In this regard, we have performed an exhaustive categorical comparison to highlight the advantages and disadvantages of data collection schemes under different metrics, highlighting varying network characteristics.

Moreover, we have included the simulation tools used in the selected schemes to guide researchers as to the cred-

ibility of other research experimentation. This parameter enables us to analyse the worth and actual capabilities of DTN, RT and BE protocols for data collection schemes associated with VANETs. A three-step analysis (Parametric, 4-D functional and Categorical) allows readers to instantly identify a protocol's advantages, disadvantages, operational benefits, constraints, and other valuable features necessary to fully understand its domain. Furthermore, we have opted to mark and highlight every protocol in a tabular format to avoid lengthy literature reads for researchers. Current open research challenges with brief dendrogram is presented in the last section to broadly specify the barriers and application gaps of VANETs. These challenges can facilitate researchers of this area to proposed solution in response to any of the gap identified. Covered open research challenges deduced after studying more than sixty data collection protocols to draw researchers' attention to unexplored and underserved areas to bridge various gaps can let us enjoy the full benefits and services of VANETs.

Funding These research was supported by University College Dublin.

References

1. Fotros M, Rezazadeh J, Sianaki OA (2020) A survey on vanets routing protocols for iot intelligent transportation systems. Workshops of the International Conference on Advanced Information Networking and Applications. Springer, pp 1097–1115
2. Kaur R, Ramachandran RK, Doss R, Pan L (2021) The importance of selecting clustering parameters in VANETs: a survey. Comput Sci Rev 40:100392
3. Abdulshaheed HR, Yaseen ZT, Salman AM, Al-Barazanchi I (2020) A survey on the use of WiMAX and Wi-Fi on vehicular ad-hoc networks (VANETs). IOP Conf Ser 870(1):012122
4. Fotros M, Rezazadeh J, Ameri Sianaki O (2020) A survey on VANETs routing protocols for IoT intelligent transportation systems. In: Barolli L, Amato F, Moscato F, Enokido T, Takizawa M (eds) Web, artificial intelligence and network applications. Springer International Publishing, Cham, pp 1097–1115
5. Manivannan D, Moni SS, Zeadally S (2020) Secure authentication and privacy-preserving techniques in Vehicular Ad-hoc NETworks (VANETs),". Veh Commun 25:100247
6. Khan S, Sharma I, Aslam M, Khan MZ, Khan S (2021) Security challenges of location privacy in VANETs and state-of-the art solutions: a survey. Future Internet 13(4):96
7. Quyoom A, Mir AA, Sarwar A (2020) Security attacks and challenges of VANETs: a literature survey. J Multimed Inf Syst 7(1):45–54
8. Jiang X, Yu FR, Song T, Leung VC (2021) Resource allocation of video streaming over vehicular networks: a survey, some research issues and challenges. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2021.3065209
9. Pavithra T and Nagabhushana B (2020) A survey on security in VANETs. In: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, pp 881–889
10. Sheikh MS, Liang J, Wang W (2020) Security and privacy in vehicular ad hoc network and vehicle cloud computing: a survey. Wireless Commun Mob Comput. https://doi.org/10.1155/2020/5129620
11. Bhoi SK and Khilar PM (2013) Vehicular communication: a survey. IET Netw 3(3):204–217. https://standards.ieee.org/develop/wg/1609_WG.html
12. Soumya S, Ponnapalli VS (2020) A survey—vanets and protocols. In: ICDSMLA 2019. Springer, pp 413–419
13. Islam A, Ranjan S, Rawat AP, Maity S (2021) A Comprehensive Survey on Attacks and Security Protocols for VANETs. Innov Comput Sci Eng. https://doi.org/10.1007/978-981-33-4543-0_62
14. Hemalatha R (2021) A survey: security challenges of vanet and their current solution. Turkish J Comput Math Educ (TURCOMAT) 12(2):1239–1244
15. Aljabry IA and Al-Suhail GA (2021) A survey on network simulators for vehicular ad-hoc networks (VANETS). Int J Comput Appl 975:8887
16. Gonçalves Filho J, Patel A, Batista BLA, Júnior JC (2016) A systematic technical survey of DTN and VDTN routing protocols. Comput Stand Interface 48:139–159
17. Madni MAA, Iranmanesh S, Raad R (2020) DTN and Non-DTN routing protocols for inter-cubesat communications: a comprehensive survey. Electronics 9(3):482
18. Das SR, Sinha K, Mukherjee N, Sinha BP (2021) Delay and disruption tolerant networks: a brief survey. Intell Cloud Comput. https://doi.org/10.1007/978-981-15-5971-6_32
19. Ahmad SA and Shcherbakov M (2018) A survey on routing protocols in vehicular adhoc networks. In: 2018 9th international conference on information, intelligence, systems and applications (IISA), IEEE, pp 1–8
20. IEEE. IEEE Standards Association. https://standards.ieee.org/findstds/standard/802.11p-2010.html
21. IEEE. IEEE Standards Association. https://standards.ieee.org/develop/wg/1609_WG.html
22. Dua A, Kumar N, Bawa S (2014) A systematic review on routing protocols for vehicular ad hoc networks. Veh Commun 1(1):33–52
23. Ali I, Hassan A, Li F (2019) Authentication and privacy schemes for vehicular ad hoc networks (VANETs): a survey. Veh Commun 16:45–61
24. Cheng N et al (2018) Big data driven vehicular networks. IEEE Netw 32(6):160–167
25. Abraham A, Koshy R (2021) A survey on VANETs routing protocols in urban scenarios. Second international conference on networks and advances in computational technologies. Springer, pp 217–229
26. Hamdi MM, Audah L, Rashid SA, Mohammed AH, Alani S and Mustafa AS (2020) A review of applications, characteristics and challenges in vehicular ad hoc networks (VANETs). In: 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), IEEE, pp 1–7
27. Baldessari R et al (2007) Car-2-car communication consortium-manifesto vol. 1.1, pp 1–94, https://elib.dlr.de/48380/1/C2C-CC_manifesto_v1.1.pdf

28. Mai T, Jiang R, Chung E (2016) A cooperative intelligent transport systems (C-ITS)-based lane-changing advisory for weaving sections. J Adv Transp 50(5):752–768
29. Festag A et al (2008) 'NoW–network on wheels': project objectives, technology and achievements. In: Proceedings of 5rd International Workshop on Intelligent Transportation (WIT). Hamburg, Germany, pp 211–216
30. Farradyne P (2005) Vehicle infrastructure integration (VII). VII Architecture and Functional Requirement Document, vol. 1
31. Ma Y, Chowdhury M, Sadek A, Jeihani M (2009) Real-time highway traffic condition assessment framework using vehicle–infrastructure integration (VII) with artificial intelligence (AI). IEEE Trans Intell Transp Syst 10(4):615–627
32. Ma Y, Chowdhury M, Sadek A, Jeihani M (2012) Integrated traffic and communication performance evaluation of an intelligent vehicle infrastructure integration (VII) system for online travel-time prediction. IEEE Trans Intell Transp Syst 13(3):1369–1382
33. Misener JA and Shladover SE (2006) PATH investigations in vehicle-roadside cooperation and safety: a foundation for safety and vehicle-infrastructure integration research. In: 2006 IEEE Intelligent Transportation Systems Conference, IEEE, pp 9–16
34. Leinmüller T et al (2006) Sevecom-secure vehicle communication. In: IST Mobile and Wireless Communication Summit, no. POST_TALK.
35. Papadimitratos P et al (2008) Secure vehicular communication systems: design and architecture. IEEE Commun Mag 46(11):100–109
36. Meneguette RI, De Grande R, Loureiro A (2018) Intelligent transport system in smart cities. Springer, Berlin
37. Sedar R, Kalalas C, Vázquez-Gallego F and Alonso-Zarate J (2021) Intelligent transport system as an example of a wireless IoT system. In: Wireless Networks and Industrial IoT. Springer, pp 243–262
38. Armengaud E et al (2019) European innovation for next generation electrified vehicles and components. In: 2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE), IEEE, pp 1–6.
39. Mallozzi P, Pelliccione P, Knauss A, Berger C, Mohammadiha N (2019) Autonomous vehicles: state of the art, future trends, and challenges. Autom Syst Softw Eng. https://doi.org/10.1007/978-3-030-12157-0_16
40. Aksjonov H, Beglerovic H, Hartmann M, Jugade S, Vaseur C (2019) On driver–vehicle–environment integration for multi-actuated ground vehicles safety advancement. IEEE ICCVE 2019:7
41. Pourghebleh B, Jafari Navimipour N (2019) Towards efficient data collection mechanisms in the vehicular ad hoc networks. Int J Commun Syst 32(5):e3893
42. Abdel-Halim IT, Fahmy HMA (2018) Prediction-based protocols for vehicular ad hoc networks: survey and taxonomy. Comput Netw 130:34–50
43. Senouci O, Harous S, Aliouat Z (2020) Survey on vehicular ad hoc networks clustering algorithms: Overview, taxonomy, challenges, and open research issues. Int J Commun Syst 33(11):e4402
44. Al-Omaisi H, Sundararajan EA, Alsaqour R, Abdullah NF and Abdelhaq M (2021) A survey of data dissemination schemes in vehicular named data networking. Veh Commun 30:100353
45. Wei K, Liang X, Xu K (2013) A survey of social-aware routing protocols in delay tolerant networks: applications, taxonomy and design-related issues. IEEE Commun Surv Tutor 16(1):556–578
46. Joseph M and Scott C. Mobile Ad-hoc Networks (manet), IETF. https://www.ietf.org/proceedings/55/177.htm
47. Qiu T, Chen N, Li K, Qiao D, Fu Z (2017) Heterogeneous ad hoc networks: architectures, advances and challenges. Ad Hoc Netw 55:143–152
48. Khabbaz MJ, Fawaz WF, Assi CM (2011) Probabilistic bundle relaying schemes in two-hop vehicular delay tolerant networks. IEEE Commun Lett 15(3):281–283
49. Sonkar N, Pandey S, Kumar S (2019) Probabilistic bundle relaying scheme in a multi-copy vehicular delay tolerant network. Int J Veh Inf Commun Syst 4(1):43–54
50. Lindgren A, Doria A, Schelén O (2003) Probabilistic routing in intermittently connected networks. ACM SIGMOBILE Mob Comput Commun Rev 7(3):19–20
51. Sok P, Tan S and Kim K (2013) PRoPHET routing protocol based on neighbor node distance using a community mobility model in delay tolerant networks. In: 2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing, IEEE, pp 1233–1240
52. Benamar N, Singh KD, Benamar M, El Ouadghiri D, Bonnin J-M (2014) Routing protocols in vehicular delay tolerant networks: a comprehensive survey. Comput Commun 48:141–158
53. Wang G, Shao M, Li R, Ma Y and Wang B (2015) Spray and wait routing algorithm based on transfer utility of node in DTN. In: 2015 IEEE International Conference on Progress in Informatics and Computing (PIC), IEEE, pp 428–432
54. Paul AB, Akhil G, Biswas S, Nandi S, Sett N (2020) SAS: seasonality aware social-based forwarder selection in delay tolerant networks. International conference on innovations for community services. Springer, pp 245–265
55. Mujahid MA, Bakar KA, Darwish TS, Zuhra FT (2021) Cluster-based location service schemes in VANETs: current state, challenges and future directions. Telecommun Syst 76(3):471–489
56. Salvo P, Turcanu I, Cuomo F, Baiocchi A, Rubin I (2017) Heterogeneous cellular and DSRC networking for Floating Car Data collection in urban areas. Veh Commun 8:21–34
57. Oliveira R, Luís M, Sargento S (2019) On the performance of social-based and location-aware forwarding strategies in urban vehicular networks. Ad Hoc Netw 93:101925
58. Bilgin BE, Baktir S, Gungor VC (2016) A novel data collection mechanism for smart grids using public transportation buses. Comput Stand Interfaces 48:19–29
59. Soares VN, Rodrigues JJ, Farahmand F (2014) GeoSpray: a geographic routing protocol for vehicular delay-tolerant networks. Inf Fus 15:102–113
60. Kang H and Kim D (2009) HVR: history-based vector routing for delay tolerant networks. In: 2009 Proceedings of 18th International Conference on Computer Communications and Networks, IEEE, pp 1–6
61. Lu Z, Gao M, Liu Z, Qu G, Dunbar C (2019) Pass and run: a privacy preserving delay tolerant network communication protocol for cyber vehicles. IEEE Des Test 36(6):56–62
62. Ramaswamy L and Ravindran B (2002) A best-effort communication protocol for real-time broadcast networks. In: Proceedings International Conference on Parallel Processing, IEEE, pp 519–526
63. Taguchi K, Enokido T and Takizawa M (2003) Hierarchical protocol for broadcast-type group communication. In: 2003 International Conference on Parallel Processing Workshops, 2003. Proceedings, IEEE, pp 21–28
64. Bali RS, Kumar N, Rodrigues JJ (2014) Clustering in vehicular ad hoc networks: taxonomy, challenges and solutions. Veh Commun 1(3):134–152
65. Tal I and Muntean G-M (2021) Clustering and 5G-enabled smart cities: a survey of clustering schemes in VANETs. In: Research Anthology on Developing and Optimizing 5G Networks and the Impact on Society: IGI Global, pp 1012–1050
66. Wang Z, Liu L, Zhou M, Ansari N (2008) A position-based clustering technique for ad hoc intervehicle communication. IEEE Trans Syst Man Cybernet Part C (Appl Rev) 38(2):201–208

67. Benkerdagh S, Duvallet C (2019) Cluster-based emergency message dissemination strategy for VANET using V2V communication. Int J Commun Syst 32(5):e3897
68. Abdel-Halim IT, Fahmy HMA, Bahaa-El Din AM (2019) Mobility prediction-based efficient clustering scheme for connected and automated vehicles in VANETs. Comput Netw 150:217–233
69. Nazib RA, Moh S (2021) Reinforcement learning-based routing protocols for vehicular ad hoc networks: a comparative survey. IEEE Access 9:27552–27587
70. Gillani M, Niaz HA, Tayyab M (2021) Role of machine learning in WSN and VANETs. Int J Elect Comput Eng Res 1(1):15–20
71. Ari AAA, Yenke BO, Labraoui N, Damakoa I, Gueroui A (2016) A power efficient cluster-based routing algorithm for wireless sensor networks: honeybees swarm intelligence based approach. J Netw Comput Appl 69:77–97
72. Kumar N, Chilamkurti N, Park JH (2013) ALCA: agent learning–based clustering algorithm in vehicular ad hoc networks. Pers Ubiquit Comput 17(8):1683–1692
73. Chahal M, Harit S (2019) A stable and reliable data dissemination scheme based on intelligent forwarding in VANETs. Int J Commun Syst 32(3):e3869
74. Goel N, Sharma G and Dhyani I (2016) A study of position based VANET routing protocols. In: 2016 international conference on computing, communication and automation (ICCCA), IEEE, pp 655–660
75. Kumar S, Verma AK (2015) Position based routing protocols in VANET: a survey. Wireless Pers Commun 83(4):2747–2772
76. Salhi I, Cherif MO and Senouci S-M (2009) A new architecture for data collection in vehicular networks. In: 2009 IEEE International Conference on Communications, IEEE, pp 1–6
77. Ullah A, Yao X, Shaheen S, Ning H (2019) Advances in position based routing towards ITS enabled FoG-oriented VANET—a survey. IEEE Trans Intell Transp Syst 21(2):828–840
78. Jiao Z, Ding H, Dang M, Tian R and Zhang B (2016) Predictive big data collection in vehicular networks: a software defined networking based approach. In: 2016 IEEE Global Communications Conference (GLOBECOM), IEEE, pp 1–6
79. Kumar R, Dave M (2016) Data relationship degree–based clustering data aggregation for VANET. Int J Electron 103(3):485–503
80. Mohammad SA and Michele CW (2010) Using traffic flow for cluster formation in vehicular ad-hoc networks. In: IEEE local computer network conference, IEEE, pp 631–636
81. Sucasas V, Radwan A, Marques H, Rodriguez J, Vahid S, Tafazolli R (2016) A survey on clustering techniques for cooperative wireless networks. Ad Hoc Netw 47:53–81
82. Gu Y, Bozdag D, Ekici E, Ozguner F and Lee C-G (2005) Partitioning based mobile element scheduling in wireless sensor networks. In: 2005 Second Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2005. IEEE SECON 2005, Citeseer, pp 386–395
83. Gu Y, Bozdağ D, Brewer RW, Ekici E (2006) Data harvesting with mobile elements in wireless sensor networks. Comput Netw 50(17):3449–3465
84. Bejaoui T (2014) QoS-oriented high dynamic resource allocation in vehicular communication networks. Sci World J. https://doi.org/10.1155/2014/718698
85. Alsuhli GH, Khattab A, Fahmy YA and Massoud Y (2019) Enhanced urban clustering in VANETs using online machine learning. In: 2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES), IEEE, pp 1–6
86. Alsabah MKJ, Trabelsi H and Jerbi W (2021) Survey on clustering in VANET networks. In: 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD), IEEE, pp 493–502
87. Singh JP, Bali RS (2015) A hybrid backbone based clustering algorithm for vehicular ad-hoc networks. Procedia Comput Sci 46:1005–1013
88. Guizani B, Ayeb B, and Koukam A (2015) A stable k-hop clustering algorithm for routing in mobile ad hoc networks. In: 2015 International Wireless Communications and Mobile Computing Conference (IWCMC), IEEE, pp 659–664
89. Wolny G (2008) Modified DMAC clustering algorithm for VANETs. In: 2008 Third International Conference on Systems and Networks Communications, 26–31 Oct. 2008, pp 268–273, https://doi.org/10.1109/ICSNC.2008.28
90. Lee J, Jeong J, Oh T, Jun J and Son SH (2016) DCMAC: data-oriented cluster-based media access control protocol for vehicular networks. In: 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), IEEE, pp 258–261
91. Brik B, Lagraa N, Lakas A, Cheddad A (2016) DDGP: Distributed data gathering protocol for vehicular networks. Veh Commun 4:15–29
92. Ren M, Zhang J, Khoukhi L, Labiod H and Vèque V (2021) A review of clustering algorithms in VANETs. Ann Telecommun 1–23
93. Almalag MS, Olariu S and Weigle MC (2012) Tdma cluster-based mac for vanets (tc-mac). In: 2012 IEEE international symposium on a world of wireless, mobile and multimedia networks (WoWMoM), IEEE, pp 1–6
94. Omar HA, Zhuang W, Li L (2012) VeMAC: a TDMA-based MAC protocol for reliable broadcast in VANETs. IEEE Trans Mob Comput 12(9):1724–1736
95. Demirkol I, Ersoy C, Alagoz F (2006) MAC protocols for wireless sensor networks: a survey. IEEE Commun Mag 44(4):115–121
96. Bazzi A, Zanella A, Masini BM (2014) An OFDMA-based MAC protocol for next-generation VANETs. IEEE Trans Veh Technol 64(9):4088–4100
97. Tomar RS and Verma S (2012) Enhanced SDMA for VANET communication. In: 2012 26th International Conference on Advanced Information Networking and Applications Workshops, IEEE, pp 688–693
98. Brik B, Lagraa N, Yagoubi MB and Lakas A (2012) An efficient and robust clustered data gathering protocol (CDGP) for vehicular networks. In: Proceedings of the second ACM international symposium on Design and analysis of intelligent vehicular networks and applications, pp 69–74
99. Brik B, Lagraa N, Lakas A, Cherroun H, Cheddad A (2016) ECDGP: extended cluster-based data gathering protocol for vehicular networks. Wirel Commun Mob Comput 16(10):1238–1255
100. Chang W-R, Lin H-T and Chen B-X (2008) Trafficgather: an efficient and scalable data collection protocol for vehicular ad hoc networks. In: 2008 5th IEEE Consumer Communications and Networking Conference, IEEE, pp 365–369
101. Brik B, Lagraa N, Cherroun H and Lakas A (2013) Token-based clustered data gathering protocol (TCDGP) in vehicular networks. In: 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC), IEEE, pp 1070–1074
102. Zahedi ZM, Akbari R, Shokouhifar M, Safaei F, Jalali A (2016) Swarm intelligence based fuzzy routing protocol for clustered wireless sensor networks. Expert Syst Appl 55:313–328
103. Maglaras LA and Katsaros D (2012) Distributed clustering in vehicular networks. In: 2012 IEEE 8th international conference on wireless and mobile computing, networking and communications (WiMob), IEEE, pp 593–599
104. Oliveira R, Montez C, Boukerche A, Wangham MS (2017) Reliable data dissemination protocol for VANET traffic safety applications. Ad Hoc Netw 63:30–44
105. Dwivedy B, Bhola AK and Yadav S (2019) Cluster based multi hop data dissemination protocol in V2V networks using

whale optimization technique. In: 2019 International Conference on Automation, Computational and Technology Management (ICACTM), IEEE, pp 228–231
106. Sulistyo S, Alam S, Adrian R (2019) Coalitional game theoretical approach for VANET clustering to improve SNR. J Comput Netw Commun. https://doi.org/10.1155/2019/4573619
107. Gangwar PK, Singh Y and Mohindru V (2014) An energy efficient zone-based clustering approach for target detection in wireless sensor networks. In: International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), IEEE, pp 1–7
108. Prabhu SB, Balakumar N (2016) Enhanced zone-based clustering method for energy efficient wireless sensor network. Int J Innov Res Electron Commun 3(4):17–22
109. Benslimane A, Taleb T, Sivaraj R (2011) Dynamic clustering-based adaptive mobile gateway management in integrated VANET—3G heterogeneous wireless networks. IEEE J Sel Areas Commun 29(3):559–570
110. Brendha R, Prakash VSJ (2019) Geographical zone-based cluster head for routing in urban vehicular network. In: Information and Communication Technology for Intelligent Systems. Springer, Berlin, pp 149–160
111. Biradar RV, Patil V, Sawant S, Mudholkar R (2009) Classification and comparison of routing protocols in wireless sensor networks. Special Issue Ubiquitous Comput Secur Syst 4(2):704–711
112. Arce P, Guerri JC, Pajares A, Lázaro O (2008) Performance evaluation of video streaming over ad hoc networks using flat and hierarchical routing protocols. Mob Netw Appl 13(3):324–336
113. Di Francesco M, Das SK, Anastasi G (2011) Data collection in wireless sensor networks with mobile elements: a survey. ACM Trans Sens Netw (TOSN) 8(1):1–31
114. Alhan A and Chawla M (2015) Analysis of encryption Dgrp-data gather routing protocol based on Opnet in VANETs. In: 2015 International Conference on Computational Intelligence and Communication Networks (CICN), IEEE, pp 1046–1051
115. Oubbati OS, Lakas A, Zhou F, Güneş M, Lagraa N, Yagoubi MB (2017) Intelligent UAV-assisted routing protocol for urban VANETs. Comput Commun 107:93–111
116. Krishna MMM. A survey UAV-assisted VANET routing protocol
117. Jarupan B, Ekici E (2010) PROMPT: a cross-layer position-based communication protocol for delay-aware vehicular access networks. Ad Hoc Netw 8(5):489–505
118. Hartenstein H, Laberteaux L (2008) A tutorial survey on vehicular ad hoc networks. IEEE Commun Mag 46(6):164–171
119. Drira W, Puthal D and Filali F (2014) ADCS: an adaptive data collection scheme in vehicular networks using 3G/LTE. In: 2014 International Conference on Connected Vehicles and Expo (ICCVE), IEEE, pp 753–758
120. Haider S, Abbas G, Abbas ZH, Boudjit S, Halim Z (2020) P-DACCA: A probabilistic direction-aware cooperative collision avoidance scheme for VANETs. Futur Gener Comput Syst 103:1–17
121. Zhang L, Jin B (2013) Dubhe: A reliable and low-latency data dissemination mechanism for VANETs. Int J Distrib Sens Netw 9(12):581821
122. He Z, Zhang D (2017) Cost-efficient traffic-aware data collection protocol in VANET. Ad Hoc Netw 55:28–39
123. AlMheiri SM and AlQamzi HS (2015) MANETs and VANETs clustering algorithms: a survey. In: 2015 IEEE 8th GCC Conference & Exhibition, IEEE, pp 1–6
124. Teymoori F, Nabizadeh H and Teymoori F (2013) A new approach in position-based routing protocol using learning automata for VANETs in city scenario. arXiv preprint http://arxiv.org/abs/1308.0099
125. Bitam S, Mellouk A, Zeadally S (2013) HyBR: A hybrid bio-inspired bee swarm routing protocol for safety applications in vehicular ad hoc networks (VANETs). J Syst Architect 59(10):953–967
126. Bouali T, Aglzim E-H and Senouci S-M (2014) A secure intersection-based routing protocol for data collection in urban vehicular networks. In: 2014 IEEE Global Communications Conference, IEEE, pp 82–87
127. Jerbi M, Senouci S-M, Rasheed T, Ghamri-Doudane Y (2009) Towards efficient geographic routing in urban vehicular networks. IEEE Trans Veh Technol 58(9):5048–5059
128. Sanguesa JA et al (2015) RTAD: A real-time adaptive dissemination system for VANETs. Comput Commun 60:53–70
129. Palazzi CE, Pezzoni F, Ruiz PM (2012) Delay-bounded data gathering in urban vehicular sensor networks. Pervasive Mob Comput 8(2):180–193
130. Soua A and Afifi H (2013) Adaptive data collection protocol using reinforcement learning for VANETs. In: 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC), IEEE, pp 1040–1045
131. Alsuhli GH, Khattab A and Fahmy YA (2019) Double-head clustering for resilient VANETs. Wireless Commun Mob Comput 2019:1–17
132. Setiabudi A, Pratiwi AA, Perdana D and Sari FR (2016) Performance comparison of GPSR and ZRP routing protocols in VANET environment. In: 2016 IEEE region 10 symposium (TENSYMP), IEEE, pp 42–47
133. Clausen T, Dearlove C, Jacquet P and Herberg U (2014) The optimized link state routing protocol version 2
134. van der Heijden RW, Dietzel S and Kargl F (2013) SeDyA: secure dynamic aggregation in VANETs. In: Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks. pp 131–142
135. Liu C, Chigan C and Gao C (2013) Compressive sensing based data collection in VANETs. In: 2013 IEEE Wireless Communications and Networking Conference (WCNC), IEEE, pp 1756–1761
136. Feng C, Zhang R, Jiang S, Li Z (2012) QoI-based data gathering and routing guidance in VANETs. International Conference on Web-Age Information Management. Springer, pp 87–98
137. Guo L, Beyah R and Li Y (2011) SMITE: a stochastic compressive data collection protocol for mobile wireless sensor networks. In: 2011 Proceedings IEEE INFOCOM, IEEE, pp 1611–1619
138. Ibrahim K and Weigle MC (2008) CASCADE: cluster-based accurate syntactic compression of aggregated data in VANETs. In: 2008 IEEE Globecom Workshops, IEEE, pp 1–10
139. Lee U, Magistretti E, Gerla M, Bellavista P, Corradi A (2008) Dissemination and harvesting of urban data using vehicular sensing platforms. IEEE Trans Veh Technol 58(2):882–901
140. Turcanu I, Salvo P, Baiocchi A, Cuomo F (2016) An integrated vanet-based data dissemination and collection protocol for complex urban scenarios. Ad Hoc Netw 52:28–38
141. Lee J-W, Lo C-C, Tang S-P, Horng M-F and Kuo Y-H (2011) A hybrid traffic geographic routing with cooperative traffic information collection scheme in VANET. In: 13th International Conference on Advanced Communication Technology (ICACT2011), IEEE, pp 1496–1501
142. Arif M, Wang G, Bhuiyan MZA, Wang T, Chen J (2019) A survey on security attacks in VANETs: Communication, applications and challenges. Veh Commun 19:100179
143. Sharma S, Kaul A (2021) VANETs cloud: architecture, applications, challenges, and issues. Arch Comput Methods Eng 28:2081–2102
144. Malhi AK, Batra S, Pannu HS (2020) Security of vehicular ad-hoc networks: a comprehensive survey. Comput Secur 89:101664
145. Yeferny T and Hamad S (2021) Vehicular ad-hoc networks: architecture, applications and challenges. arXiv preprint http://arxiv.org/abs/2101.04539

146. Hussain R, Lee J, Zeadally S (2020) Trust in VANET: A survey of current solutions and future research opportunities. IEEE Trans Intell Transp Syst 22(5):2553–2571
147. Ghosal A, Conti M (2020) Security issues and challenges in V2X: a survey. Comput Netw 169:107093
148. Gillani M, Ullah A and Niaz HA (2018) Survey of requirement management techniques for safety critical systems. In: 2018 12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 2018: IEEE, pp 1–5
149. Gayathri M, Gomathy C (2021) A deep survey on types of cyber attacks in VANET. J Crit Rev 8(01):1029–1039
150. Obaidat M, Khodjaeva M, Holst J, Zid MB (2020) "Security and privacy challenges vehicular ad hoc networks. In: Connected Vehicles in the Internet of Things. Springer, Berlin, pp 223–251
151. Gillani M, Ullah A and Niaz HA (2018) Trust management schemes for secure routing in VANETs—a survey. In: 2018 12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 2018: IEEE, pp 1–6
152. Mundhe P, Verma S, Venkatesan S (2021) A comprehensive survey on authentication and privacy-preserving schemes in VANETs. Comput Sci Rev 41:100411
153. Sun R, Huang Y, Zhu L (2021) Communication by credence: trust communication in vehicular Ad Hoc networks. Mob Netw Appl. https://doi.org/10.1007/s11036-020-01695-0
154. Lu Z, Qu G, Liu Z (2018) A survey on recent advances in vehicular network security, trust, and privacy. IEEE Trans Intell Transp Syst 20(2):760–776