Proceedings of the 12th International Conference on Information Technology (ICIT-2009), Dec 21-24, 2009, Bhubaneswar, India

An Energy Efficient Query Processing Protocol in Wireless Sensor Network

T Panigrahi, P M Pradhan and G Panda
Dept. of ECE, National Institute of Technology, Rourkela-769008, India
tpanigrahi80@gmail.com

***Abstract*—Since sensor networks can be thought of as a distributed database system, several architectures proposed to interface the application to the sensor network through querying protocol. However sensor networks are so massively distributed, so careful consideration should be put into the efficient organization of data and the execution of queries. Here we consider the problem of information discovery in a densely deployed wireless Sensor Network (WSN) where the initiator of search is unaware of target information. Here we discuss a new type of protocol known as Coverage Based Increasing Ray Search which is an energy efficient and scalable search protocol. The basic principle of this protocol is to route the search packet along a set of trajectories called rays that maximises the likelihood of discovering the target information by consuming least number of transmission. The rays are organised such that if the search packet travels along all these rays, then the entire terrain area will be covered by its transmissions.**

***Index Terms*—Wireless sensor network, energy efficiency, search, querying protocol.**

I. INTRODUCTION

Efficient design and real time implementation of wireless sensor networks has become a hot area of research in recent years. It is due to the vast potential of sensor networks to enable applications that connect the physical world to the virtual world. By networking large numbers of tiny sensor nodes using wireless links, it is possible to obtain data about physical phenomena that was difficult or impossible to obtain in more conventional ways by using traditional wired networks. In the coming years, as advances in VLSI and MEMS technology allow the cost of manufacturing sensor nodes to continue to drop, increasing deployments of wireless sensor networks are expected, with the networks eventually growing to large numbers of nodes (e.g., thousands). Potential applications for such large-scale wireless sensor networks exist in a variety of elds, including medical monitoring, environmental monitoring, surveillance, home security, military operations, and industrial machine monitoring [1].

It should be noted that sensor networks do share some commonalities with general ad hoc networks. Thus, protocol design for sensor networks must account for the properties of ad hoc networks, including the lifetime constraints imposed by the limited energy supplies of the nodes in the network, unreliable communication due to the wireless medium and need for self configuration, requiring little or no human intervention. However, several unique features exist in wireless sensor networks that do not exist in general ad hoc networks. These features present new challenges and require modication of designs for traditional ad hoc networks.

One of the most important aspects of WSN is the nature of the data sink(s). Sink nodes are storage points for most of the data emerging from environmental sensing of sensor nodes. In Paper [2], the author describes how data is gathered and catogorized in WSN which are as follows:

- PUSH or CONTINUOUS COLLECTION: Sensor nodes periodically sense environment and send data to the sink node.
- PULL or QUERYING: Sensor nodes sense environment and store the information locally. On need basis, the sink node queries for the required information.
- PUSH-PULL: This paradigm involves both PUSH and PULL. Sensor nodes push the sensed events to different sensor nodes in the network in a predetermined way that is used by the search initiator for finding the target information.

The use of PUSH, PULL, or PUSH-PULL approach depends on various factors of WSN such as, types of application, available memory, and energy efficiency. As sensor nodes are battery powered, energy is a premium resource in most cases. PUSH approach is efficient when continuous sensing is required and PULL approach is efficient for low frequency data gathering. In PULL paradigm, WSN can be considered as a distributed database and on need basis, the sink node sends queries for data collection. Some of the factors that inuence the usage of PUSH-PULL approaches are the rate of occurrence of events, the query rate, the type of events sensed, and available memory resources on sensor nodes. If the query rate is low and the rate of occurrence of events is high or event type is audio or video then it is clearly not feasible to store them in multiple sensor nodes as they may consume the memory completely.

Querying in WSNs is an active research area and there are many proposals for reducing the overhead of search cost based on the query type. Tiny AGgregation (TAG) Service is a generic aggregation service for wireless sensor networks that minimizes the amount of messages transmitted during the execution of a query [3]. In contrast to standard database query execution techniques, in which all data is gathered by a central processor where the query is executed, TAG allows the query to be executed in a distributed fashion, greatly reducing the overall amount of traffic transmitted on the network. Similarly TinyDB is a processing engine that runs

Acquisitions Query Processing (ACQP) [4], providing an easy-to-use generic interface to the network through an enhanced SQL-like interface and enabling the execution of queries to be optimized at several levels. In [5] the author classifies the type of WSN queries as (a) Continuous vs one-shot queries, (b)Aggregate vs non-aggregate queries, (c) Complex vs simple queries, and (d) Queries for replicated data vs queries for unique data. There are lot of proposals is available in literature for the above listed queries.

In this paper, we focus on PULL [6] and UNSTRUCTURED [7] WSNs where, the sink node sends simple and one-shot [5] queries for unique data. In UNSTRUCTURED WSNs, the search initiator has no clue about the location of target information. In the existing protocols the cost of search increases if the sensor node density increase. This limits the scalability of the protocols especially for densed network. Here we discuss increase ray search which is energy efficient query resolution protocols application to simple one-shot queries for unique data in UNSTRUCTURED WSNs.

For UNSTRUCTURED WSNs search proceeds blindly for tracking the target information. The following are the most widely used techniques for searching in UNSTRUCTURED WSNs: Expanding Ring Search (ERS) [8], [9], Random Walk Search [10], [11], and Variants of Gossip Search [12]. ERS is a prominent technique used in multihop network. It avoids network-wide broadcast by searching for the target information with increasing order of TTL (Time-To-Live) values. TTL limits the number of hops to be searched from the source node. If search fails continuously up to TTL Threshold hops, ERS initiates network-wide broadcast. The main disadvantage of this protocol is that it resembles ooding in the worst case.

In Random walk search, when a node has to forward the search packet, it randomly selects one of its neighbours and forwards the search packet to the selected neighbour. The basic idea here is the random wandering in the network in search of the target information until TTL (number of hops) is expired or the target information is found. The main disadvantage of Random walk is that the probability of nding the nearest replica of the target information is low and due to this, the Data Transfer Cost will be very high especially in the case of continuous queries.

In Gossip search, the source node broadcasts the search packet and all receivers either forward it with a probability p or drop it with a probability $1 \quad p$. In some cases, gossip dies early without reaching reasonable number of nodes even for higher values of p which increases the non-determinism of Gossip search. For this reason, in [12], the authors propose GOSSIP (p, k) where k is the number of hops for which the search packet has to be transmitted with probability 1 *i.e.*, for rst k hops the search packet is always forwarded after which it is forwarded with a probability p. The main disadvantage of Gossip search is that of sending message to most of the sensor nodes even when the target information is located close to the source node.

Increasing ray search operate by dividing the terrain into very narrow rectangular section called *rays*. Each ray is characterized by a source and destination point where the source node is the sink node which send the query and the destination node is on the circumference of the circular terrain. The ray searches in decreasing order of the unexplored area where the unexplored area is defined as the area not covered by any of the earlier search ray. The query packet starts from the sink node and travel to the destination to cover entire area of the ray. When the target node receives the query packet, a response packet is sent back to the sink node. For a fixed terrain the number of transmissions required to cover the entire terrain area is constant and because of this it is independent of node density for a given terrain size.

II. Assumptions

the following assumptions are taken for designing the protocol and also for mathematical analysis.

- The terrain is considered to be circular. The sink node is static and is placed at the centre of the circular terrain.
- Sensor nodes are stationary and are deployed uniformly in the terrain. It may not applicable in real time scenario.
- Sensor nodes are aware of their own location. The assignment of location to sensor nodes is part of the initial setup of WSN.
- We consider PULL and UNSTRUCTURED WSN where the sink node sends simple, one-shot queries for unique data.
- The search initiator is unaware of the locations of target information replicas.
- Events occur uniformly in the given terrain.
- To relay the search packet along the *Medians* of the rings, we assume that the density of sensor nodes is high.

III. Protocol Design

The basic principle of coverage based increasing ray search is that if a subset of the total sensor nodes transmit the search packet such that, the entire circular terrain area is covered by these transmissions, then at least one target node will denitely receive the search packet. The selection of the subset of nodes which transmit the search packet are performed in a distributed way. However if the search packets are broadcast to entire terrain, even though the target information found, the number of messages are large. To minimizes the number of messages we divide the circular terrain into narrow rectangular regions called *rays*. In IRS, the rectangular regions are covered one after another until the target information is found. Each ray is formed by dividing the circumference of the circular terrain into archs of the length equal to twice to that of transmission radius of sensor nodes and attaching to the two end points of arc to the two end points of the transmission diameter of the sink node as illustrated in Fig.1. The width of the ray is equal to the twice the transmission radius of sensor nodes. The median of the ray is a line joining the mid point of the arc and the sink node.

The sink node broadcast search packet by embedding the information of the first ray in it with $Angle = 30^0$. A node which receives the search packet is referred as *CurrentNode*. All CurrentNodes evaluate the following two conditions to check whether they are eligible to forward the search packet or not:

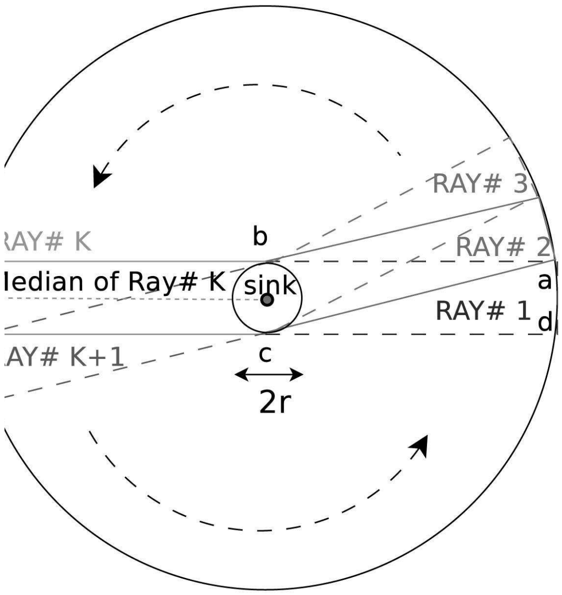

Fig. 1. Circular terrain divided into rays

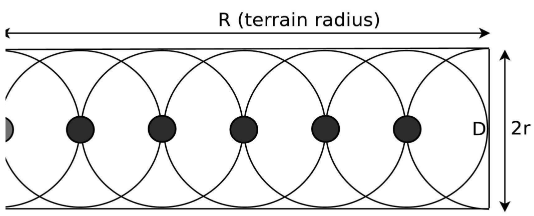

Fig. 2. The area covered by a single ray

1) is distance between current node and destination point ($Dist_{c,d}$) less than distance between relay node and destination point ($Dist_{r,d}$)?
2) is $\angle DRC$ less than $Angle$

A node which satisfies both these conditions is referred as *EligibleNode*. The first condition makes sure that the EligibleNode is closer to *Destination Point* compared to *RelayedNode*. The second condition makes sure of the following: 1) The EligibleNodes are closer to the Median of ray and 2) When the $Angle = 30^0$, all nodes in the *EligibleNodes* set are in the transmission range of each other. An *EligibleNode* has to wait for a time proportional to its proximity to the Desti nationPoint and the Median of ray before relaying the packet. When an EligibleNode relays the search packet, all other EligibleNodes which receive this packet, drop the packet which should be relayed by them. The time to wait before relaying is given by

$$T_{wait} = W_1 \times Dist_{c,d} + W_2 \times Dist_{c,m}$$

where W_1 and W_2 are the weight factors and $Dist_{c,m}$ is the distance between current node and median of the ray. By choosing appropriate value of these weighting factor, the weight of the $Dist_{c,d}$ and $Dist_{r,m}$ can be adjusted.

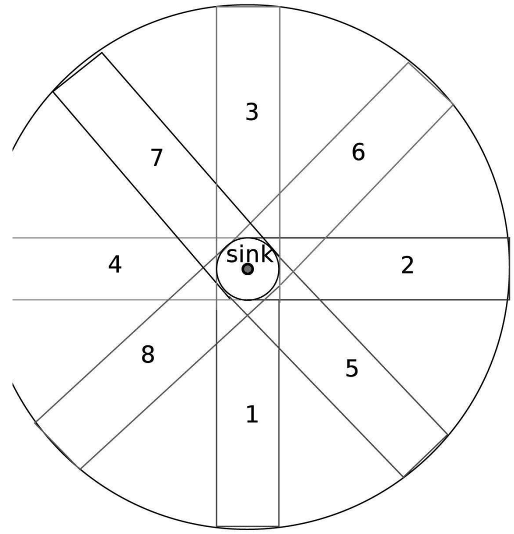

Fig. 3. The search order of the rays

Based on the values of W_1 and W_2, the *EligibleNode* which is closer to the Median of current ray and among them the one closer to the *DestinationPoint* will have lesser waiting time compared to the other *EligibleNodes*. The *EligibleNode* with smaller waiting time relays the search packet while others drop it. This continues until the search packet reaches the DestinationPoint or there is no other node to relay the search packet further.All nodes are forwarding through the median ray. The idea behind this is to keep track of which areas of terrain are covered and also to cover the entire region of ray as illustrated in Fig.2. From one can observe that some portions of the rectangle are not covered by transmissions of any of the sensor nodes. Because of this, IRS are not deterministic protocols, but the probability of finding target information is very high as long as the distance between the nodes and the angular conditions can be calculated accurately. But, if the coordinates are not very accurate, then the trajectory of IRS variants might deviate, due to which, the search efficiency in terms of the cost and latency might be affected.

A. Ordering of the Rays

We sort the rays in decreasing order of unexplored area covered by them. After the division of circular terrain into rays, the area covered by each ray is equal. However, the area covered by a ray which is not covered by any of the rays previously searched is not same for all the rays. We call this area as unexplored area. For example, in Fig. 1, let us assume that rays are ordered according to ray number. Clearly, the unexplored area covered by $Ray - 1$ is more than the unexplored area covered by $Ray-2$. At any point in the order, the next ray is the one which covers the maximum unexplored area of all the remaining unsearched rays. In this way, the rays are ordered in decreasing order of unexplored area covered by them. In Fig. 1, the next ray in the order after $Ray - 1$ is definitely not $Ray - 2$, as there are other rays which cover more unexplored area than $Ray - 2$. $Ray - K$ covers the most unexplored area compared to all other unsearched rays. There may be multiple rays that cover the most unexplored area at a given point in the order, in this case, one of them is selected as the next ray to be searched. This pattern of ray ordering or ray growth is called Greedy Ray Growth (GRG), as it tries to maximize the probability of finding the target node in rays searched as early as possible. GRG is illustrated in Fig. 3. The numbers on rays in this figure indicate the order of search.

B. Grouping of the Ray

We make the rays into groups ordered according to the unexplored area covered. All rays in a group cover equal unexplored area. The first ray explored will be part of Group1 and this will be the only ray in the group. The first ray covers maximum unexplored area and this is defined as the total area of the ray which is given as

$$A_1 = (R + r) \times 2r = 2(n + 1)r^2 \quad (1)$$

where R is the radius of the circular terrain, r is the transmission radius of the sensor nodes and n is the minimum number of hops required to reach the destination point from the sing node.

While deriving the unexplored area covered by a ray after first ray, we have to consider the overlap with previous rays. For any of the rays in other groups, the unexplored area covered should exclude the area covered by previous rays. In deriving the unexplored area covered by the rays in second group, the length of the rectangle (ray) should be taken as $(R - r)$ whereas for the ray in first group, the length of the rectangle is $(R + r)$. After the first ray, the next maximum unexplored area is covered by rays which are either 90^0 or 180^0 to the first ray, i.e., Rays 2, 3, and 4 in Fig. 3. We add these rays to Group2 for which the area covered is given by

$$A_2 = (R - r) \times 2r = 2(n - 1)r^2 \quad (2)$$

So Group2 has four rays in the circular terrain. These four rays divide the circular terrain into four approximately equal sectors. The next ray to maximize the unexplored area is always the one with its Median stretching from the sink node to the midpoint of circumference of any one of the newly formed sectors.The four additional rays will create eight new sectors and this process continues until the entire circular terrain area is covered. Based on the above pattern, we can generalize the following: The maximum number of rays in Group1; Group2; Group3; Group4; Group5; . . . ; etc: will be 1; 3; 4; 8; 16; . . . ; etc., respectively. The total number of groups:

$$G = \lceil log_2[\pi n] \rceil \quad (3)$$

C. Coverage Based Increasing Ray Search

Coverage Based IRS explores the rays one after the other according to GRG. When IRS is initiated, the sink node sends the search packet to the ray which covers the maximum unexplored area. Then the sink node waits for a time-out *N Rwait* before sending the search packet to the next ray. The *N Rwait* value should be carefully estimated based on the radius of circular terrain as a high value of *N Rwait* results in high latency of search and a low value of *N Rwait* results in high cost of search. If the sink node receives acknowledgement from the target node within this time-out value, it stops the search; otherwise it continues with the next ray. IRS explores groups starting with Group1 and in a single group, no specific ordering is followed and rays are explored sequentially. The protocol is designed in such a way that it always chooses the sensor nodes closest to the Medians of rays for forwarding the search packet. One possible way of alleviating this problem is to find a different set of rays for each search so that the *DestinationPoints* and the *Medians* of rays will be different for each search, thereby, load gets distributed among all the sensor nodes. This protocol has high latency for saving cost and this makes it suitable mainly for delay insensitive applications.

IV. THEORETICAL ANALYSIS

In this section, we provide a theoretical analysis of the proposed IRS according to the analysis of the existing protocols such as ERS, Random walk, and Gossip. We consider a dense WSN in our theoretical analysis. Our main goals for theoretical analysis is to show that the cost and latency of IRS are independent of node density

A. Cost Analysis of IRS

Cost is modeled as the number of messages transmitted to find the target information. We assume that sensor nodes are densely deployed and area covered by each ray forms a rectangular structure as shown in Fig. 2. Length of ray = R, Breadth of ray = 2r, since we always assume dense deployment and IRS always reaches the Destination Point in minimum number of hops, so minimum number of hops required to reach the destination point is defined as

$$n = \left\lceil \frac{R}{r} \right\rceil$$

Total number of rays required to cover the terrain is

$$n = \left\lceil \frac{2\pi R}{2r} \right\rceil = \lceil \pi n \rceil$$

By doing the analysis for cost estimation finally we get the following expression for the average number of messages required to find the target information using IRS i.e. the expected value of cost is

$$E[C]_i = \frac{1}{\pi n^2} \times \left[(n^2 \times (4^G)) + nB + C \right] \quad (4)$$

where,

$$B = \left(4^G \times \left(\frac{11}{3} - 2G \right) + 2^G - \frac{26}{3} \right)$$

$$C = (2^G \times (5 - 2G) - 6)$$

we can see that $\frac{d(E[C]_i)}{dn} > 0, \quad \forall n > 2$, so that $E[C]_i$ is a monotonically increasing function of n in the intervals of $(2, \infty)$ Therefore the cost of IRS depends only on R (

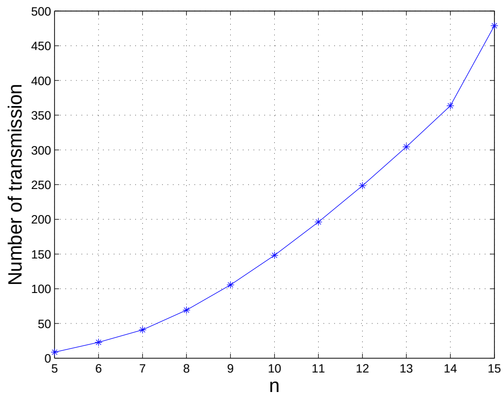

Fig. 4. Effect of terrain size on cost of search

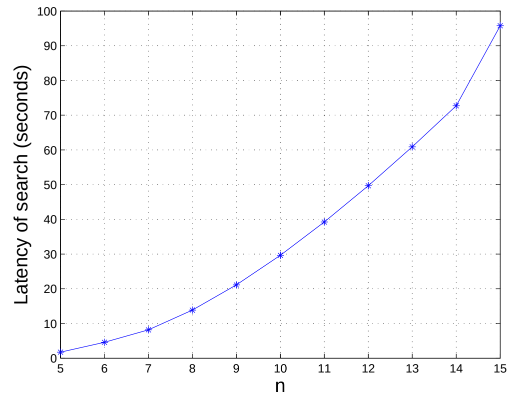

Fig. 5. Effect of terrain size on latency of search

because of $n \propto R$). so that the radius of the circular terrain is independent of the density nodes in WSN.

B. Latency analysis of IRS

The derivation of expression for latency of IRS is similar to the way we derived expression for $E[C]_i$. The only difference being, transmission time per hop should be considered instead of cost per hop. In deriving $E[C]_i$, we calculated the average number of transmissions required to find the target information assuming each transmission incurs unit cost. Here, we assume the transmission time per hop to be h seconds.

Let $E[L]_i$ be the expected value of latency for finding the target information using IRS:

$$E[L]_i = \frac{h}{\pi n^2} \times \left[(n^2 \times (4^G)) + nB + C \right] \quad (5)$$

where,

$$B = \left(4^G \times \left(\frac{11}{3} - 2G \right) + 2^G - \frac{26}{3} \right)$$

$$C = (2^G \times (5 - 2G) - 6)$$

Based on the results of cost analysis, the latency of IRS also depends only on R i.e. the radius of the circular terrain and is independent of density of nodes in WSN.

V. Numerical Results

Initial simulation results are given in this paper. Terrain size is plotted in terms of n where n is defined as the minimum number of hops required to reach the destination point from the sing node. We also assume that the node density is also uniform through out the terrain. Figs. 4 and 5 show the plots for the effect of terrain size on the cost and latency of IRS. We fixed h at 200 ms to make it more realistic as beacon-less forwarding incurs high delays in forwarding packets. Here we were not compared our results with existing techniques, but we will try to provide detailed results during the presentation.

VI. Conclusion

In this paper, we presented coverage based increasing ray search is energy efficient and scalable query resolution protocols for simple, one-shot queries for unique data. The conclusion drawn from the paper is that under high node density, this protocol consume much less cost and it is unaffected by the variation in node density. Energy is the most premium resource in WSNs and described protocol achieve significant energy savings for dense WSNs.

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks,” *Communications Magazine, IEEE*, vol. 40, no. 8, pp. 102–114, Aug 2002.

[2] X. Liu, Q. Huang, and Y. Zhang, “Combs, Needles, Haystacks: Balancing push and pull for discovery in large scale sensor network,” in *Proceedings of Second International Conference Embedded Network Sensor System*, NOV. 2004, pp. 122–133.

[3] S. Madden, M. Franklin, J. Hellerstein, and W. Wong, “TAG: a tiny aggregation service for ad-hoc sensor network.” in *Proceedings of the ACM Symposium on Operating System Design and Implementation (OSDI)*, 2002.

[4] ——, “The design of an acqustional query processor for sensor networks,” in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 2003.

[5] N. sadagopan, B. Krischnamochari, and A. helmy, “The ACQUIRE mechanism for efficient querying in sensor networks.” in *SNPA 2003: Proceedings of the 1st IEEE International Workshop on Sensor Network Protocols and Applications*, MAY 2003, pp. 149–159.

[6] R. Govindan, J. M. Hellerstein, W. Hong, S. Madden, M. Franklin, and S. Shenker, “The sensor network as a database.” in *Technical Report 02-771, Computer Science Dept., Unversity of southern California*, SEP. 2002.

[7] J. Ahn and B. Krishnamachari, “Modelling search cost in wireless sensor network,” in *Technical Report CENG-2007-1, Computer Science Dept., Unversity of southern California*, 2007.

[8] N. B. Chang and M. Liu, “Controlled flooding search in a large network,” *IEEE/ACM Trans. Networking*, vol. 15, no. 2, pp. 436–449, APRIL. 2007.

[9] J. Hassan and S. Jha, “Optimizing expanding ring search for multi-hop wireless network.” in *Proc. 47th Ann. IEEE Global telecomm Conf. (GLOBECOM’04)*, Nov. 2004, pp. 1061–1065.

[10] H. Tian, H. Sen, and T. Matsuzawa, “Random walk routing for wireless sensor network.” in *Proc. sixth International conf. parallel and Distributed Computing Aplications and Technologies (PDCAT’05)*, DEC. 2005, pp. 196–200.

[11] L.Lima and J. Borros, “Random walk on sensor network.” in *Proc. fifth International Symposium Modelling and optimization in mobile, Ad-Hoc and Wireless Networks (WiOpt’07)*, Apr. 2007, pp. 1–5.

[12] Z. J. Haas, J. Y. Halperen, and L. Li, “Gossip based ad-hoc routing,” *IEEE/ACM Trans. Networking*, vol. 14, no. 3, pp. 479–491, March. 2006.