A K-Anonymity Privacy-Preserving Location Monitoring System for Wireless Sensor Networks with Nymble Secure System
Gayathri M, Bharathi M
Department of Computer Science and Engineering, SJCIT.
Affiliated to Vishvesvaraya Technological University, Chikkaballapur, India.

Abstract—Anonymizing wireless sensor networks allow users to access services privately by using a series of routers to hide the client’s IP address from the server. As a result, administrators block all known exit nodes of anonymizing networks, denying anonymous access to misbehaving. To address this problem, servers can “blacklist” misbehaving users, thereby blocking users without compromising their anonymity. Monitoring personal locations with a potentially untrusted server poses privacy threats to the monitored individuals, a privacy-preserving location monitoring system for wireless sensor networks is adopted. Two innetwork location anonymization algorithms are considered, namely, resource and quality-aware algorithms, that aim to enable the system to provide high-quality location monitoring services for system users, while preserving personal location privacy. Both algorithms rely on the well established k-anonymity privacy concept, that is, a person is indistinguishable among k persons, to enable trusted sensor nodes to provide the aggregate location information of monitored persons. Each aggregate location is in a form of a monitored area A along with the number of monitored persons residing in A, where A contains at least k persons. The resource-aware algorithm aims to minimize communication and computational cost, while the quality-aware algorithm aims to maximize the accuracy of the aggregate locations by minimizing their monitored areas. To utilize the aggregate location information and to provide location monitoring services, a spatial histogram approach is used that estimates the distribution of the monitored persons based on the gathered aggregate location information. Then, the estimated distribution is used to provide location monitoring services through answering range queries.

Keywords—Location privacy, wireless sensor networks, location monitoring system, aggregate query processing, spatial histogram, Anonymous blacklisting, privacy, Misbehaving users.

I. INTRODUCTION

Wireless sensor networks (WSN), is a large collection of densely deployed, spatially distributed, autonomous devices (or nodes) that communicate via wireless and cooperatively monitor physical or environmental conditions. The sensor nodes such networks are deployed over a geographic area by aerial scattering or other means. Each sensor node can only detect events within a very limited distance, called the sensing range. In addition, sensor nodes normally have fairly limited transmission and reception capabilities so that sensing data have to be relayed via a multi-hop path to a distant base station (BS), which is a data collection center with sufficiently powerful processing capabilities and resources. Monitoring personal locations with a potentially untrusted system poses privacy threats to the monitored individuals. This paper proposes a privacy preserving location monitoring system for wireless sensor networks to provide monitoring services. It relies on the well established k-anonymity privacy concept which requires each person is indistinguishable among k persons. In this system each sensor node blurs its sensing area into a cloaked area, in which at least k persons are residing. Each sensor node reports only aggregate location information, which is inform of a cloaked area along with the number of persons, N located in A where N>= k to the server[1]. Two innetwork aggregate location anonymization algorithms namely resource and quality aware algorithms are adopted along with securing the user system by Nymble, which uses the following properties: anonymous authentication, backward unlinkability, subjective blacklisting. In Nymble, users acquire an ordered collection of nymbles, a special type of pseudonym, to connect to Websites. Without additional information, these nymbles are computationally hard to link, and hence, using the stream of nymbles simulates anonymous access to services. Web sites, however, can blacklist users by obtaining a seed for a particular nymble, allowing them to link future nymbles from the same user—those used before the complaint remain unlinkable. Servers can therefore blacklist anonymous users without knowledge of their IP addresses while allowing behaving users to connect anonymously. Our system ensures that users are aware of their blacklist status before they present a nymble, and disconnect immediately if they are blacklisted.

II. SYSTEM MODEL

Fig.1 depicts the architecture of our system, where there are three major entities, sensor nodes, server, and system users. We will define the problem addressed
by our system, and then describe the detail of each entity and the privacy model of our system.

1) **Sensor nodes.** Each sensor node is responsible for determining the number of objects in its sensing area, blurring its sensing area into a cloaked area A, which includes at least k objects, and reporting A with the number of objects located in A as aggregate location information to the server. Each sensor node is also aware of its location and sensing area.

3) **Server.** The server is responsible for collecting the aggregate locations reported from the sensor nodes, using a spatial histogram to estimate the distribution of the monitored objects, and answering range queries based on the estimated object distribution. Furthermore, the administrator can change the anonymized level k of the system at anytime by disseminating a message with a new value of k to all the sensor nodes.

4) **System users.** Authenticated administrators and users can issue range queries to our system through either the server or the sensor nodes, as depicted in Fig. 1. The server uses the spatial histogram to answer their queries.

5) **Privacy model.** In our system, the sensor nodes constitute a trusted zone, where they behave as defined in our algorithm and communicate with each other through a secure network channel to avoid internal network attacks, [6], [11]. Our system also provides anonymous communication between the sensor nodes and the server by employing existing anonymous communication techniques [12], [13]. Thus given an aggregate location R, the server only knows that the sender of R is one of the sensor nodes within R. Furthermore, only authenticated administrators can change the k-anonymity level and the spatial histogram size. In emergency cases, the administrators can set the k-anonymity level to a small value to get more accurate aggregate locations from the sensor nodes, or even set it to zero to disable our algorithm to get the original readings from the sensor nodes, in order to get the best services from the system.

2) **Resource-Based Blocking.** To limit the number of identities a user can obtain, the Nymble system binds nymbles to resources that are sufficiently difficult to in great numbers[19]. For ex. We have used IP address as resource in our implementation.

6) **The Pseudonym Manager.** The user must first contact the Pseudonym Manager (PM) and demonstrate control over a resource; for IP-address blocking, the user must connect to the PM directly.

7) **The Nymble Manager.** After obtaining a pseudonym from the PM, the user connects to the Nymble Manager (NM) through the anonymizing network, and requests nymbles for access to a particular server (such as Wikipedia). A user’s requests to the NM are therefore pseudonymous, and nymbles are generated using the user’s pseudonym and the server’s identity. These nymbles are thus specific to a particular user-server pair.

8) **Time.** Nymble tickets are bound to specific time periods. While a user’s access within a time period is tied to a single nymble ticket, the use of different nymble tickets across time periods grants the user anonymity between time periods. Smaller time periods provide users with higher rates of anonymous authentication, while longer time periods allow servers to rate-limit the number of misbehaviors from a particular user before he or she is blocked. The linkability window allows for dynamism since resources such as IP addresses can get reassigned and it is undesirable to blacklist such resources indefinitely, and it ensures forgiveness of misbehaviour after a certain period of time.

9) **Blacklisting a User.** If a user misbehaves, the server may link any future connection from this user within the current linkability window. A user connects and misbehaves at a server during time period t within linkability window w. The server later detects this
misbehavior and complains to the NM in time period $t_c$ of the same linkability window $w$. As part of the complaint, the server presents the nimble ticket of the misbehaving user and obtains the corresponding seed from the NM. The server is then able to link future connections by the user in time periods of the same linkability window $w$ to the complaint. Therefore, once the server has complained about a user, that user is blacklisted for the rest of the day.

### III. LOCATION ANONYMIZING AND NYMBLE SECURE SYSTEM ALGORITHMS

#### A. The Resource-Aware Algorithm

Algorithm 1 outlines the resource-aware location anonymization algorithm. Fig. 3 gives an example to illustrate the resource-aware algorithm, where there are seven sensor nodes, A to G, and the required anonymity level is five, $k = 5$. The dotted circles represent the sensing area of the sensor nodes, and a line between two sensor nodes indicates that these two sensor nodes can communicate directly with each other. In general, the algorithm has three steps.

1: function RESOURCEAWARE (Integer k, Sensor m, List R)
2: PeerList $\leftarrow \{0\}$
3: // Step 1: The broadcast step
4: Send a message with m’s identity m.ID, sensing area m.Area, and object count m.Count to m’s neighbor peers
5: // Step 2: The cloaked area step
6: Add the message to PeerList
7: if m has found an adequate number of objects then
8: Send a notification message to m’s neighbors
9: else if Some m’s neighbor has not found an adequate number of objects then
10: Forward the message to m’s neighbors
11: end if
12: end if
13: // Step 3: The validation step
14: if no containment relationship with Area and R $\in$ R then
15: Send Area with a cloaked N to the peers within Area and the server
16: else if m’s sensing area is contained by some R $\in$ R then
17: Randomly select a R’ $\in$ R such that R’:Area contains m’s sensing area
18: Send R’ to the peers within R’:Area and the server
19: end if
20: end if

#### B. Quality-aware location anonymization algorithm

1: function QUALITYAWARE (Integer k, Sensor m, Set init_solution, List R)
2: current_min_cloaked_area $\leftarrow$ init_solution
3: // Step 1: The search space step
4: Determine a search space S based on init_solution
5: Collect the information of the peers located in S
6: Add each peer located in S to C[1] as an item
7: Add m to each item set in C[1] as the first item
8: for i = 1; i <= 4; i ++ do
9: for each item set X = {$a_1, \ldots, a_i$} in C[i] do
10: if Area(MBR(X)) < Area(current_min_cloaked_area) then
11: if N(MBR(X)) $\geq$ k then
12: current_min_cloaked_area $\leftarrow$ X
13: Remove X from C[i]
14: else
15: Remove X from C[i]
16: end if
17: end for
18: if i < 4 then
19: for each item set pair X = {$x_1, \ldots, x_i$}, Y = {$y_1, \ldots, y_i$} in C[i] do
20: if $x_1 = y_1, \ldots, x_i = y_i$ and $x_{i+1} \neq y_{i+1}$ then
21: Add an item set {$x_1, \ldots, x_i, y_{i+1}$} to C[i+1]
22: end if
23: end for
24: end if
25: end if
26: end for
27: end for
28: Lines 18 to 25 in Algorithm 1
C. Pseudonyms’ Algorithm

PMCreatePseudonym
Input: (uid, w) e HxN
Persistent state: pmState € S_p
Output: pnym € P
1: Extract nymKeyP, macKeynp from pmState
2: nym := MA.Mac(uid||w, nymKeyp)
3: mac := MA.Mac(nym||w, macKeyNP)
4: return pnym := (nym, mac)

D. NMVerifyPseudonym Algorithm

Input: (pnym, w) e PxN
Persistent state: nmState € SN
Output: b € (true, false)
1: Extract macKeyNP from nmState
2: (nym, mac) := pnym
3: return mac = ? MA.Mac(nym||w, macKeyNP)

E. Seeds and Nymbles Algorithm

NMCreateCredential
Input: (pnym, sid, w) e PxHxN
Persistent state: nmState € SN
Output: cred € D
1: Extract macKeyNS, macKeyN, seedKeyN, encKeyN from keys in nmState
2: seed0 := f(Mac(pnym||sid||w, seedKeyN))
3: nimble* := g(seed0)
4: for t from 1 to L do
5: seedt := f(seedt-1)
6: nymblet := g(seedt)

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we show and analyze the experimental results with respect to the privacy protection and the quality of location monitoring services of our system. We implemented Nymble and collected various performance numbers which verify the linear time and space cost of the various operations and datastructures.

A. Anonymization Strength

When the anonymity level gets stricter, our algorithms generate larger cloaked areas, which reduce the accuracy of the aggregate locations reported to the server. When there are more objects, our algorithms generate smaller cloaked areas, which increase the accuracy of the aggregate locations reported to the server.

B. Effect of Query Region Size

Fig.6 depicts the privacy protection and the quality of our location monitoring system with respect to increasing the query region size ratio from 0.001 to 0.256, where the query region size ratio is the ratio of the query region area to the system area and the query region size ratio 0.001 corresponds to the size of a sensor node’s sensing area. The results give evidence that our system provides low-quality location monitoring services for the range query with a small query region, and better quality services for larger query regions. This is an important feature to protect personal location privacy, because providing the accurate number of objects in a small area could reveal individual location information; therefore, an adversary cannot use our system output to track the monitored objects with any fidelity. The definition of a Parameter Settings. The results also show that the quality-aware algorithm always performs better than the resource-aware algorithm.

C. Effect of the Number of Objects

The broadcast step of the resource-aware algorithm effectively allows each sensor node to find an adequate number of objects to blur its sensing area. When there are more objects, the sensor node finds smaller cloaked areas that satisfy the k-anonymity privacy requirement.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we show and analyze the experimental results with respect to the privacy protection and the quality of location monitoring services of our system. We implemented Nymble and collected various performance numbers which verify the linear time and space cost of the various operations and datastructures.

A. Anonymization Strength

When the anonymity level gets stricter, our algorithms generate larger cloaked areas, which reduce the accuracy of the aggregate locations reported to the server. When there are more objects, our algorithms generate smaller cloaked areas, which increase the accuracy of the aggregate locations reported to the server.

B. Effect of Query Region Size

Fig.6 depicts the privacy protection and the quality of our location monitoring system with respect to increasing the query region size ratio from 0.001 to 0.256, where the query region size ratio is the ratio of the query region area to the system area and the query region size ratio 0.001 corresponds to the size of a sensor node’s sensing area. The results give evidence that our system provides low-quality location monitoring services for the range query with a small query region, and better quality services for larger query regions. This is an important feature to protect personal location privacy, because providing the accurate number of objects in a small area could reveal individual location information; therefore, an adversary cannot use our system output to track the monitored objects with any fidelity. The definition of a Parameter Settings. The results also show that the quality-aware algorithm always performs better than the resource-aware algorithm.

C. Effect of the Number of Objects

The broadcast step of the resource-aware algorithm effectively allows each sensor node to find an adequate number of objects to blur its sensing area. When there are more objects, the sensor node finds smaller cloaked areas that satisfy the k-anonymity privacy requirement,
Fig. 6. Query region size (a) Resource-aware algorithm (b) Quality-aware algorithm.

D. Effect of Privacy Requirements

Fig. 7 depicts the performance of our system with respect to varying the required anonymity level $k$ from 10 to 30. When the k-anonymity privacy requirement gets stricter, the sensor nodes have to enlist more peers for help to blur their sensing areas; therefore, the communication cost of our algorithms increases (Fig. 7a). To satisfy the stricter anonymity levels, our algorithms generate larger cloaked areas, as depicted in Fig. 7b. For the quality-aware algorithm, since there are more peers in the required search space when the input (resource-aware) cloaked area gets larger, the computational cost of computing the minimal cloaked area by the quality-aware algorithm and the basic approach gets worse (Fig. 7c). However, the quality-aware algorithm reduces the computational cost of the basic approach by at least four orders of magnitude. Larger cloaked areas give more inaccurate aggregate location information to the system, so the estimation error increases as the required k-anonymity increases (Fig. 7d). The quality-aware algorithm provides much better quality location monitoring services than the resource-aware algorithm, when the required anonymity level gets stricter.

An honest PM and NM will issue a coalition of $c$ unique users at most $c$ valid credentials for a given server. Because of the security of HMAC, only the NM can issue valid tickets, and for any time period, the coalition has at most $c$ valid tickets, and can thus make at most $c$ connections to the server in any time period regardless of the server’s blacklisting. It suffices to show that if each of the $c$ users has been blacklisted in some previous time period of the current linkability window, the coalition cannot authenticate in the current time period $k'$.

F. Nonframeability

Assume the contrary that the adversary successfully framed honest user $i^*$ with respect to an honest server in time period $t^*$, and thus, user $i^*$ was unable to connect in time period $t^*$ using ticket* even though none of his tickets were previously blacklisted. Because of the security of HMAC, and since the PM and NM are honest, the adversary cannot forge tickets for user $i^*$ and the server cannot already have seen ticket*; it must be that ticket* was linked to an entry in the linking list. Thus, there exists an entry in the server’s linking list, such that the nymble in ticket* equals nimble*. The server must have obtained this entry in a successful blacklist update for some valid ticketb, implying the NM had created this ticket for some user $i'$.

G. Anonymity

We show that an adversary learns only that some legitimate user connected or that some illegitimate user’s connection failed, i.e., there are two anonymity sets of legitimate and illegitimate users. Since all honest users execute the Nymble connection establishment protocol in exactly the same manner up until the end of the Blacklist validation stage, it suffices to show that every illegitimate user will evaluate safe to false, and hence, terminate the protocol with failure at the end of the Privacy check stage. The authenticity of the channel implies that a legitimate user knows the correct identity of the server, and thus, Boolean disclosed for the server remains false.

H. Across Multiple Linkability Windows

With multiple linkability windows, our Nymble construction still has Accountability and Nonframeability because each ticket is valid for and only for a specific linkability window; it still has Anonymity because pseudonyms are an output of a collision-resistant function that takes the linkability window as input.

Fig 7. Anonymity levels. (a) Communication cost (b) Cloaked area size (c) Computational cost (d) Estimation error

E. Blacklistability
Fig. 8. Nymble’s performance at (a) the NM and (b) the user and the server when performing various protocols. (a) Blacklist updates take several milliseconds and credentials can be generated in 9 ms for the suggested parameter of $L = 288$. (b) The bottleneck operation of server blacklist update is less than 1 ms and validating the blacklist takes the user only a few ms.

V. RELATED WORK

Straightforward approaches for preserving users’ location privacy include enforcing privacy policies to restrict the use of collected location information [15], [16] and anonymizing the stored data before any disclosure [17]. However, these approaches fail to prevent internal data thefts or inadvertent disclosure. Recently, location anonymization techniques have been widely used to anonymize personal location information before any server gathers the location information, in order to preserve personal location privacy in location-based services. These techniques are based on one of the three concepts. 1) False locations. Instead of reporting the monitored object’s exact location, the object reports $n$ different locations, where only one of them is the object’s actual location while the rest are false locations [18]. 2) Spatial cloaking. The spatial cloaking technique blurs a user’s location into a cloaked spatial area that satisfy the user’s specified privacy requirements [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]. 3) Space transformation. This technique transforms the location information of queries and data into another space, where the spatial relationship among the query and data are encoded [29]. Among these three privacy concepts, only the spatial cloaking technique can be applied to our problem. The main reasons for this are that 1) the false location techniques cannot provide high-quality monitoring services due to a large amount of false location information, 2) the space transformation techniques cannot provide privacy preserving monitoring services as it reveals the monitored object’s exact location information to the query issuer, and 3) the spatial cloaking techniques can provide aggregate location information to the server and balance a trade-off between privacy protection and the quality of services by tuning the specified privacy requirements, for example, $k$ anonymity and minimum area privacy requirements [17], [27]. Thus, we adopt the spatial cloaking technique to reserve the monitored object’s location privacy in our location monitoring system.

IP-address blocking. By picking IP addresses as the resource for limiting the Sybil attack, our current implementation closely mimics IP-address blocking employed by Internet services. There are, however, some inherent limitations to using IP addresses as the scarce resource. If a user can obtain multiple addresses, she can circumvent both nymble-based and regular IP-address blocking.

VI. CONCLUSION

In this paper, we propose a privacy-preserving location monitoring system for wireless sensor networks. We adopt two in-network location anonymization algorithms, namely, resource and quality-aware algorithms, that preserve personal location privacy, while enabling the system to provide location monitoring services. Both algorithms rely on the well-established $k$-anonymity privacy concept that requires a person is indistinguishable among $k$ persons. In our system, sensor nodes execute our location anonymization algorithms to provide $k$-anonymous aggregate locations, in which each aggregate location is a cloaked area $A$ with the number of monitored objects, $N$, located in $A$, where $N \geq k$, for the system. The resource-aware algorithm aims to minimize communication and computational cost, while the quality-aware algorithm aims to minimize the size of cloaked areas in order to generate more accurate aggregate locations. To provide location monitoring services based on the aggregate location information, we adopt a spatial histogram approach that analyzes the aggregate locations reported from the sensor nodes to estimate the distribution of the monitored objects. The estimated distribution is used to provide location monitoring services through answering range queries. We evaluate our system through simulated experiments. The results show that our system provides high-quality location monitoring services (the accuracy of the resource-aware algorithm is about 75 percent and the accuracy of the quality-aware algorithm is about 90 percent), while preserving the monitored object’s location privacy. We have proposed and built a comprehensive credential system called Nymble, which can be used to add a layer of accountability to any publicly known anonymizing network. Servers can blacklist misbehaving users while maintaining their privacy, and we show how these properties can be attained in a way that is practical, efficient, and sensitive to the needs of both users and services.
INTERNATIONAL JOURNAL OF COMPUTER & ORGANIZATION TRENDS – VOLUME 2 ISSUE 2 MARCH TO APRIL 2012

REFERENCES


