Aalborg Universitet

Inaccuracy of Location Information as a Consequence of Data Collection Delay and Presence of Misbehaving and Malicious Nodes

Madsen, Tatiana Kozlova; Nielsen, Jimmy Jessen; Garcia, Mariano; Poblacion, Adrian; Marques, Hugo; Sucasas, Victor

Inaccuracy of Location Information as a Consequence of Data Collection Delay and Presence of Misbehaving and Malicious Nodes

Tatiana K. Madsen*, Jimmy J. Nielsen*, Mariano García†, Adrián Población†, Hugo Marques‡, Victor Sucasas‡
{tatiana,jjn}@es.aau.dk*, {mariano,adrian}@gaps.ssr.upm.es†, {hugo.marques,vsucasas}@av.it.pt‡
*Dept. of Electronic Systems, Aalborg University, Fredrik Bajers vej 7A, 9220 Aalborg, Denmark
†ETSI Telecomunicación, Universidad Politécnica de Madrid, Avda. Complutense 30, 28040 Madrid, Spain
‡Instituto de Telecomunicações, Campus Universitário de Santiago, 3810-193, Aveiro, Portugal

Abstract— **This paper presents an overview of current research investigations within the WHERE-2 Project with respect to possibilities of usage of geo-location information for protocol and system optimizations. In this work we would like to underline that incorporating the knowledge of nodes' location in protocol design and optimizations one should be aware of the associated additional "cost" in terms of increased communication overhead and processing requirements, as well as potential inaccuracy of the location information. We ellaborate on the example where location information becomes stale due to nodes' mobility and delays caused by transmission and propagation delays. Presence of malicious nodes can be another reason for erroneous location estimation. We demonstrate the importance of choosing localization techniques that are robust towards misbehaviours.**

I. INTRODUCTION

Currently, location measurement technologies are becoming more and more widespread and it is becoming common to base smart phones and devices application on the availability of location information, thus offering advanced services and enhanced user experience. What is more, the use of location information can potentially advance mobile communication and mobile computing, e.g., by helping conducting efficient network resource management, and in this way by enhancing the quality of existing connections/communication links. Some of the classical examples where information about nodes' positions can be successfully applied for optimizing protocol design are ad hoc networks and mobile wireless sensor networks. Here nodes are moving autonomously and comprise a distributed network where cooperation between devices is essential in order to provide networking functionalities. In such scenarios, location information is the key for the implementation of geographical routing protocols (such as GPSR [1] and GEAR [2]), which outperform other routing schemes in terms of traffic overhead and scalability. Another class of examples is relaying that is a well-known technique for improving connectivity in wireless networks. By letting neighboring nodes act as relays, two low error rate transmissions can be used instead of a single high error rate transmission, thereby improving the transmission quality. Here, also relying on information about nodes' positions for relay selction is a promising approach due to lower signaling overhead and since it allows for movement

prediction, and thus, provides a good base for optimal relay selection [3], [4].

In order to introduce intelligence in a network based on location information, a system-wide perspective should be taken, which typically requires that knowledge about the positions of various physical objects, mobile devices and different network entities (such as access points; mobile relays or cluster heads) is available and distributed within the network. It raises a question how positioning information can be collected and exchanged efficiently and securely. Information exchange introduces communication overhead and this should be taken into account in the overall performance evaluation. Additionally, in order to ensure confidentiality and integrity the information exchange should be done in a secure and lightweight manner.

How well the location information can be exploited depends also on the accuracy of this information. Position information can be inaccurate for the following reasons:

1) **Inaccurate estimation**: environmental factors in addition to limitations in receiver sensitivity and localization algorithms means that an estimate of a user's position typically has some error relative to the true position. As communication protocol designers we have to deal with this inaccuracy and evaluate performance under realistic assumption about the quality of positioning information; it is important to understand the influence of inaccuracy on the performance improvement.

2) **Delayed and inconsistent position estimation**: many scenarios with a real need for location infomration are networks characterized by a high nodes mobility. This poses a requirement to have timely updated location estimations. Due to mobility, information can quickly become outdated. It is important to understand an impact of the delayed information on the overall system performance.

3) **Malicious attacks**: localization protocols are vulnerable in hostile environments and traditional authentication methods might not be very efficient; efficient methods to tolerate malicious attacks against beacon-based location discovery should be designed and evaluated to make
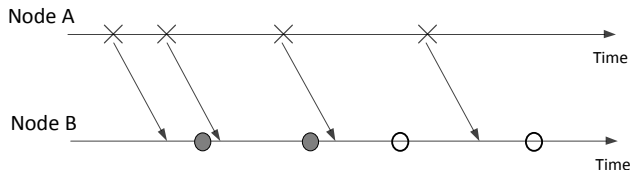
Fig. 1. Illustration of impact of data collection delays on event state estimation: crosses indicate moments when a state change has taken places; grey circles correspond to the wrong state estimation, while white circles show correct estimation.

localization algorithms robust towards malicious data manipulations.

The first mentioned reason for inaccuracy is already well-understood and documented in the literature (see e.g. [5]). It is becoming common to model the error in position estimation due to localization algorithm as a zero-mean Gaussian noise. In this paper our focus is on the two other reasons that are often neglected. In the following we elaborate on how nodes mobility and time required for infomation exchange impact accuracy of information about nodes' positions and illustrate it by an example of geo-based clustering. Additionally, we show that localization techniques are sensitive to malicious attacks, and advocate that robust procedures should be employed when there is suspicion about the behavior of any node that serves as a reference position in the localization process.

II. DELAYED AND INCONSISTENT POSITION ESTIMATION DUE TO MOBILITY

A. General note on Mismatch probability

In distributed systems we typically face a problem that information about an event that has happened somewhere in the system should be propagated to all network entities. An event can for example be the latest measurement of a physical phenomenon such as a temperature measurement, indicator of an element state such as an indicator of an empty printer cartridge or, as we will consider later, information about geographical location of network nodes. Delays associated with processing and transmission of information can potentially cause a mismatch between a received state and the true state. It happens in a situation when during the time it takes to propagate the information about the current state, the state actually changes, and thus the received information becomes outdated. This is illustrated in Figure 1. In case of a discrete state space, we can speak about mismatch probability, i.e. the probability that a state estimate at time t is different from the true state at that time. In case of continuous variables an error in state estimation is of interest.

In the following we will use location information as a state information. In mobile scenarios, a mismatch between the actual position of a device and other nodes' belief of the node's position, can easily occur, since a device continues to move while information about its position is dissiminated. We further illustrate this problem with an example of geo-location clustering taking a closer look at which information flows are in this scenario.

B. Geo-location based clustering

A problem of organizing a set of mobile, radio-equipped nodes into a connected network is a general networking problem. We require that a reliable structure can be achieved in a distributed manner that is robust towards any topological changes due to nodes' mobility or failure. In cellular networks resource allocation is managed by a base station. In a somewhat similar manner this solution can be extended to ad hoc and sensor networks by creating clusters of nodes. Most hierarchical clustering architectures for mobile radio networks are based on the concept of clusterhead, a node that acts as a local coordinator within the cluster.

The importance of clustering techniques comes from the scalability problems of reactive and proactive routing protocols. The complexity in terms of overhead is highly increased with the size of the networks. Concretely, in proactive routing protocols the overhead is in the order of n^2 where n is the number of nodes, while in reactive routing protocols the delay produced by the route search can be considerable.

With clustering techniques we can reduce the routing information by setting a small group of nodes to form a backbone. These will be the *clusterheads* and the rest of nodes will be tied to one of the members of the backbone forming a cluster. In this way we can significantly reduce the number of routes. Thus, every node will have to store less routing information and the topology itself will be more stable. The mobility of one node will only affect the nodes of the cluster(s) it is leaving or joining. With non-overlapping clusters we also get better reuse of resources. Two clusters can use the same channel if they are not in touch. Besides, since nodes are better coordinated this reduces the number of collisions.

Due to the availability of GPS, Wi-Fi and cellular based location systems in most of the new mobile devices, it is possible to include the geographic location of the node as a parameter in the cluster formation algorithm. In the following we briefly describe the most currently known algorithms that use location as metric to form clusters.

One of the most known mobility-aware algorithms is the *Weighted Clustering Algorithm* (WCA) [6]. In WCA the election of the *clusterhead* is influenced by four parameters: the degree difference w.r.t. a defined threshold Δ_v, the sum of distances with all neighbours D_v, the speed average M_v and the time serving as a cluster head P_v (this is related with the battery consumption, the more the time being *clusterhead* the more energy is being drained by the coordination tasks). Performance evaluation indicates that the proposed metric works well, however, each node is supposed to gather sufficient accurate information for a period of time that does not fit in a highly changeable scenario–the so called frozen period is longer than in other algorithms. The message exchange overhead is also bigger since the nodes have to share the metric and also compute the distances from their neighbours.

There are other proposals focused on improving this metric by including other parameters like remaining battery level or the mean connectivity degree [7]. With respect to the mobility

parameter, M_v, the Entropy Based Weighted Clustering Algorithm [8] improves the stability of the topology by measuring the level of disorder of the set of neighbouring nodes of each node. This is done by comparing the change of the current position of the node with each neighbour and choosing the one with a more uniform distribution.

Other studies face the problem of the optimization of a composed metric. The work proposed by [9] and [10] use a genetic algorithm and a simulated annealing technique to optimize the election of the dominating set. These techniques, however, just optimize the metric as a function of only one variable, the weight, and choosing the value of the weights is not a simple task. In [11] a method is proposed to optimize the election of the *clusterheads* by optimizing multiple parameters and not only the outcome of the weight function. In [12] the Distributed Group Mobility Adaptive clustering algorithm is proposed based on the measurement of the current location. It records the location of the node in discrete times. If the newest location differs more than a threshold from the previous location it computes the direction (angle w.r.t. the coordinate system) and the speed (magnitude of the movement) and records it. By comparing these parameters with the neighbours it computes the Total Spatial Dependency (TSD). TSD is the addition of the Spatial Dependency (SD) of each neighbour, thus a higher TSD implies the node has a bigger set of neighbours with similar mobility. This algorithm is suited for scenarios where nodes are moving in groups.

The Stable Cluster Protocol (SCP) in [13] proposes a new metric based on the speed of the nodes: a Stable Factor is calculated comparing a node's speed with the average speed of neighbours. The Mobility Based Clustering (MBC) algorithm proposed in [14], as the previous ones, is also based on the speed of the nodes. MBC takes the velocity vector of each node and computes the relative mobility of each pair of nodes.

The Mobile Clustering (MC) algorithm described in [15] is based on a collection of location updates at a server. In MC all nodes update to a server an information vector that includes location information. Each node is described by its identity, position, velocity and time of the sample. With this information the server can assess the level of dissimilarity of two nodes using 'only' the position recorded at several times, and weighting the samples to give more relevance to the most current ones.

The algorithm KCMBC (k-hop Compound Metric Based Clustering Algorithm) [16] uses a different approach. It takes into account the connectivity degree, the *Id* of the node and the mobility, but the latter is used to predict the time expiration of a link between two nodes. It uses the position and the velocity of the nodes to compute, in pairs, the time of availability of a given link and computes the average of this time with respect to the nodes in the neighbouring set. The most suitable node to be *clusterhead* will be the one with bigger time average. The assumption that the velocity will be invariable for a given period of time is necessary. After this time interval a *hello* message is needed between the nodes to update the new values of velocity and position. KCMBC uses a dynamic

time interval that can be tunable depending on the mobility of the nodes. The higher the mobility, the more frequent the data is exchanged. This is an interesting and intuitively clear observation on which we will ellaborate in the next subsection.

C. Is location information reliable?

From the literature overview presented above one can see that there is a lot of focus on ways to construct suitable metrics for clusterhead selection assuming the availability of location information and much less considerations are done on efficient ways to disseminate this information among the nodes as well as to evaluating the impact of a stale location information on the system performance. In the following we illustrate that delays in location information dissemination can have a drastic impact on the accuracy of the information.

Let us consider a scenario when each node in a network periodically sends updates containing information about its current position. Upon receiving this broadcast the neighboring nodes update their entries and recalculate the distances among each pair of nodes in a network. In such situation, while making a distance estimation, only information about one node can be assumed to be sufficiently correct; the information about other nodes' positions can be slightly outdated depending on the time when the broadcast from a particular node was received last. To model this situation we assume that there are n nodes in a network exchanging information either directly or via multiple hops. From a point of view of a single node, location updates from other nodes are received at some random points in time and interarrival times between consecutive updates are exponentially distributed with rate λ. It means that the time that elapsed from receiving k-last update is gamma distributed with parameters (k, λ). We assume that all nodes move with the same constant speed v, but no information is available about the direction of their movement. Thus, in a time Δt that has elapsed since the last update from a node, at the current time it could be anywhere on a circle with a centre at his old position and radius $r = \Delta t \cdot v$. Figure 2 shows the average values for distance error $\epsilon_d = |d_{\text{new}}^{ij} - d_{\text{old}}^{ij}|$ where d_{new}^{ij} is a true distance between nodes i and j, whereas d_{old}^{ij} is a distance calculated based on the stale position information available at a node. The horisontal axis corresponds to a rate at which location updates are performed. Three curves are presented for different speed values. Increase in location update rate, thus reducing the interval between consecutive updates, leads to significant reduction of distance estimation error. However, it also means increase in control overhead as the number of broadcasted messages is increased. Speed of nodes' movement also plays an important role: with low speed even infrequent location updates are sufficient to keep estimation error at an acceptable level, while high speed requires more often updates.

D. Discussions

With the simple example presented in the previous subsection we have motivated a balanced approach where an update rate of location information exchange is correlated with
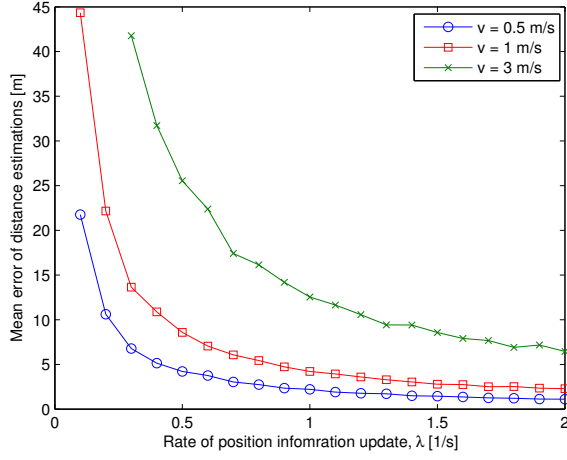
Fig. 2. Illustration of impact of data collection delays on event state estimation.
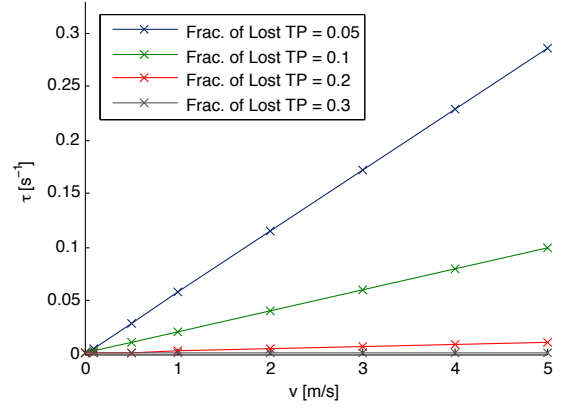

Fig. 3. Results showing how the proposed Markov Chain model of location based relaying can be used to determine the optimal information update frequency for a given scenario, where τ is the update rate and v is the mobility speed.

the required estimation error and nodes' mobility. Using the dependency between these parameters like the one presented on Figure 2, one can choose optimal values for λ. An additional optimization constraint in this case is minimization of resource consumption required for location information dissimilation.

Similar problems concerning the impact of mobility on the accuracy of location information appears in other situation as well, typically in a dynamic environments and when location information is used for protocol and system performance optimizations. In [4] the authors consider the problem of a mobile relay selection based on collected location information–specifically on the impact of node mobility and information collection delays. It is shown that selecting good values for particularly the information update frequency, is crucial for achieving good system throughput with a location based relay selection scheme. However, setting the update frequency too high, leads to a waste of of network resources due to excess signaling overhead. This work has been extended within the WHERE2 project in order to analytically model this impact. Specifically we have proposed a Markov Chain model that takes into account relay mobility and information collection delays, which is described in details in [17]. A key outcome of this contribution is that the proposed model allows to calculate the optimal information update frequency for a given scenario in terms of mobility speed and required Quality of Service (QoS). The QoS is specified in terms of a maximum allowed fraction of lost throughput, which is defined relative to the hypothetical case of having perfect information. This result is shown in Fig. 3.

III. Inaccurate position estimation due to malicious attacks

The availability of accurate node localization is critical for the successful implementation of communication protocols that are resilient to different kinds of attacks, such as link spoofing, wormholes, etc [18]. Localization protocols assume the existence of a set of nodes (anchor nodes or beacons) that are aware of their own positions and make such information available to the network. Then, any other node could be able to locate itself using these reference positions along with some measurements obtained by the node that can be related to its relative position with respect to the beacons. However, the localization process can be compromised by nodes that deliberately report incorrect position references (i.e. malicious anchor nodes) so that the measured distances are inconsistent with such fake locations and thus lead to biased positions estimations.

A. Secure localization

Let us assume that there are N beacons located at points in the plane $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$, and the node to be located is able to estimate the distances $d_1, d_2, \ldots, d_N$ to the beacons. Then, the absolute position (x, y) of the node can be estimated via a trilateration procedure as

$$(\hat{x}, \hat{y}) = \arg\min_{(x,y)} C(x, y | x_i, y_i, d_i, i = 1, 2, \ldots, N) \quad (1)$$

where $C(\cdot)$ is a suitable "cost function". An example of cost function is the one used in the method of least squares (LS)

$$C_{LS} = \sum_{i=1}^{N} \varepsilon_i^2 \quad (2)$$

where ε_i is a "residual" that depends on the difference $d_i - \sqrt{(x - x_i)^2 + (y - y_i)^2}$ and is, therefore, only related to the information obtained from the i-th anchor.

If there are malicious beacons reporting fake location references, we face a problem of estimating an unknown vector of parameters (i.e. the node position) using a set of observations (distances to the beacons) some of which are "outliers" (those given by the malicious beacons). It is well-known that the conventional LS approach is especially sensitive to the presence of outliers, so a solution to mitigate the effects of such kind

of attacks is to resort to robust estimation techniques, which are resilient to the effects of outliers [19]–[21].

Among those robust estimators, we have:

a) Least Absolute Deviations (LAD): In this method, the cost function is the $\mathcal{L}_1$-norm of the vector of residuals, instead of the quadratic norm used in the LS procedure

$$C_{LAD} = \sum_{i=1}^{N} |\varepsilon_i| \qquad (3)$$

b) Least Trimmed Squares (LTS): The cost function is quadratic, but only considers the subset of the q (with $q \leq N$) smallest squared residuals, so that up to $N - q$ observations could be outliers without reaching the breakdown point

$$C_{LTS} = \sum_{i=1}^{q} \varepsilon_{(i)}^2 \qquad (4)$$

c) Least Median of Squares (LMS): It is well-known that the median is a robust measure of centrality; by using a cost function based on this statistic, the location estimation can tolerate up to 50% of malicious beacons

$$C_{LMS} = \underset{i=1,2,\ldots,N}{\text{median}} \varepsilon_i^2 \qquad (5)$$

B. Simulation Results

We have simulated a network composed of 10 beacon nodes randomly deployed in a square room of 30m×30m, and a node actually located in the center of the room, whose position is to be estimated from measurements of distances to the beacon nodes. These measurements are contaminated with zero-mean Gaussian noise of standard deviation proportional to the actual distance in meters (to simulate the effect of errors in TOA-based distance estimations with limited bandwidth), with a proportionality constant of 0.0648. Furthermore, we have assumed that M randomly chosen malicious beacons are supposed to declare false locations that are 20 m away from their real positions. We have tested the estimations given by the LS, LAD, LTS and LMS methods of (2), (3), (4) and (5), respectively. For the LTS algorithm, we have chosen the number of residual terms $q = 6$.

The quality of the estimation is measured through statistics of the "location error", defined as

$$e = \sqrt{(\hat{x} - x)^2 + (\hat{y} - y)^2} \qquad (6)$$

where (x, y) and $(\hat{x}, \hat{y})$ are the actual and estimated positions, respectively. The location error is characterized by its cumulative distribution function (CDF): $F_e(x) = P(e \leq x)$.

Some results are represented in Fig. 4, where we can see that, when no attacks are present ($M = 0$), the LS, LAD, LTS and LMS algorithms give approximately the same performance. However, in the presence of two malicious beacons ($M = 2$), the behavior of the LS method quickly deteriorates, while the LAD approach is somewhat less affected by the measurement biases, although its overall performance is poor. On the other hand, the influence of the outliers introduced by the compromised beacons on the LMS method is negligible
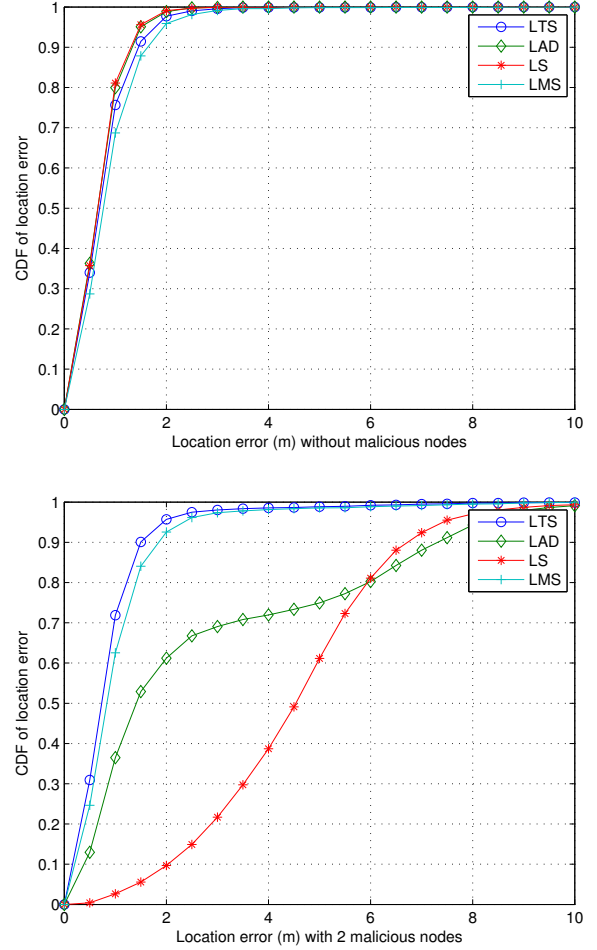

Fig. 4. CDF of location error

(there are less than 50% malicious beacons), and also the LTS performs very well, because the condition $M \leq N - q$ is fulfilled.

IV. CONCLUSION

The goal of this paper is to attract attention of researchers to the issue of a source of inaccurate and unreliable location information that can be easily overseen. And here we do not mean inaccuracy of location estimation obtained by one or another localization algorithm, but we consider location information that becomes outdated while it is disseminated in a network and thus becomes inuaccurate, and unreliable estimation if an active attack is launched by e.g. modifying beaconing messages. We have provided insights into both problems. For the first problem a discussion of trade-off and how different parameters influence each other is given; in the second part we have reviewed localization techniques that are robust towards the presence of malicious nodes.

ACKNOWLEDGMENT

REFERENCES

[1] b. Karp and H. T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," in *Proc. of the 6th ACM Int. Conf. on Mobile Computing and Networking*, 2000.

[2] Y. Yu, D. Estrin, and R. Govindan, "Geographical and energy aware routing: a recursive data dissemination protocol for wireless sensor networks," in *UCLA Computer Science Dept. technical Report, UCLA-CSD TR-01-0023*, 2001.

[3] J. J. Nielsen, T. K. Madsen, and H.-P. Schwefel, "Mobility impact on centralized selection of mobile relays," in *Proc. of the 6th IEEE Consumer Communications and Networking Conference - CCNC*, 2010.

[4] ——, "Location-based mobile relay selection and impact of inaccurate path loss model parameters," in *Proc. of the IEEE Wireless Communications and Networking Conference - WCNC*, 2010.

[5] C. Mensing, S. Sand, M. Laaraiedh, B. Uguen, B. Denis, M. Garcia, J. Casajus, T. Pedersen, G. S. X. Yin, B. Fleury, S. Mayrargue, and D. Slock, "Where d2.1 - performance assessment of hybrid data fusion and tracking algorithms," ICT-217033 WHERE, Tech. Rep., 2008.

[6] M. Chatterjee, S. Das, and D. Turgut, "Wca: A weighted clustering algorithm for mobile ad hoc networks," *Cluster Computing*, vol. 5, no. 2, pp. 193–204, 2002.

[7] W. Yang and G. Zhang, "A weight-based clustering algorithm for mobile ad hoc network," in *Wireless and Mobile Communications, 2007. ICWMC'07. Third International Conference on*. IEEE, 2007, pp. 3–3.

[8] Y. Wang and F. Bao, "An entropy-based weighted clustering algorithm and its optimization for ad hoc networks," in *wimob*. IEEE Computer Society, 2007, p. 56.

[9] D. Turgut, S. Das, R. Elmasri, and B. Turgut, "Optimizing clustering algorithm in mobile ad hoc networks using genetic algorithmic approach," in *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE*, vol. 1. IEEE, 2002, pp. 62–66.

[10] D. Turgut, B. Turgut, R. Elmasri, and T. Le, "Optimizing clustering algorithm in mobile ad hoc networks using simulated annealing," in *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, vol. 3. IEEE, 2003, pp. 1492–1497.

[11] H. Cheng, J. Cao, X. Wang, S. Das, and S. Yang, "Stability-aware multi-metric clustering in mobile ad hoc networks with group mobility," *Wireless Communications and Mobile Computing*, vol. 9, no. 6, pp. 759–771, 2009.

[12] Y. Zhang, J. Ng, and C. Low, "A distributed group mobility adaptive clustering algorithm for mobile ad hoc networks," *Computer Communications*, vol. 32, no. 1, pp. 189–202, 2009.

[13] K. Liu, J. Su, J. Zhang, F. Liu, and C. Gong, "A novel stable cluster protocol for mobile ad hoc networks," in *Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications, 2005. MAPE 2005. IEEE International Symposium on*, vol. 2. IEEE, pp. 1328–1332.

[14] B. An and S. Papavassiliou, "A mobility-based clustering approach to support mobility management and multicast routing in mobile ad-hoc wireless networks," *International Journal of Network Management*, vol. 11, no. 6, pp. 387–395, 2001.

[15] C. Jensen, D. Lin, and B. Ooi, "Continuous clustering of moving objects," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1161–1174, 2007.

[16] S. Leng, Y. Zhang, H. Chen, L. Zhang, and K. Liu, "A novel k-hop compound metric based clustering scheme for ad hoc wireless networks," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 1, pp. 367–375, 2009.

[17] J. J. Nielsen, R. L. Olsen, T. K. Madsen, and H.-P. Schwefel, "On the impact of information delay on location-based relaying: A markov modeling approach," in *Proc. of the IEEE Wireless Communications and Networking Conference - WCNC*, 2012.

[18] X. Du and H.-H. Chen, "Security in wireless sensor networks," *Wireless Communications, IEEE*, vol. 15, no. 4, pp. 60 –66, aug. 2008.

[19] Z. Li, W. Trappe, Y. Zhang, and B. Nath, "Robust statistical methods for securing wireless localization in sensor networks," in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, april 2005, pp. 91 – 98.

[20] D. Liu, P. Ning, and W. Du, "Attack-resistant location estimation in sensor networks," in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, april 2005, pp. 99 – 106.

[21] M. Garcia-Otero, F. Alvarez-Garcia, and F. Casajus-Quiros, "Securing wireless sensor networks by using location information," in *Systems, Signals and Image Processing, 2009. IWSSIP 2009. 16th International Conference on*, june 2009, pp. 1 –4.