PROCEEDINGS

55. IWK

Internationales Wissenschaftliches Kolloquium
International Scientific Colloquium

13 - 17 September 2010

Crossing Borders within the ABC

Automation,

Biomedical Engineering and

Computer Science

Faculty of
Computer Science and Automation

www.tu-ilmenau.de

TECHNISCHE UNIVERSITÄT
ILMENAU

Impressum
Published by

Publisher:	Rector of the Ilmenau University of Technology Univ.-Prof. Dr. rer. nat. habil. Dr. h. c. Prof. h. c. Peter Scharff
Editor:	Marketing Department (Phone: +49 3677 69-2520) Andrea Schneider (conferences@tu-ilmenau.de) Faculty of Computer Science and Automation (Phone: +49 3677 69-2860) Univ.-Prof. Dr.-Ing. habil. Jens Haueisen
Editorial Deadline:	20. August 2010
Implementation:	Ilmenau University of Technology Felix Böckelmann Philipp Schmidt

USB-Flash-Version.

Publishing House:	Verlag ISLE, Betriebsstätte des ISLE e.V. Werner-von-Siemens-Str. 16 98693 Ilmenau
Production:	CDA Datenträger Albrechts GmbH, 98529 Suhl/Albrechts
Order trough:	Marketing Department (+49 3677 69-2520) Andrea Schneider (conferences@tu-ilmenau.de)

ISBN: 978-3-938843-53-6 (USB-Flash Version)

Online-Version:

Publisher:	Universitätsbibliothek Ilmenau ilmedia Postfach 10 05 65 98684 Ilmenau

Home / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=16739

SECURE MULTI-HOP LOCALIZATION IN WIRELESS AD HOC NETWORKS

Sander Wozniak [1] *and Tobias Gerlach* [2] *and Guenter Schaefer* [1]

Ilmenau University of Technology, Germany
[1] Telematics and Computer Networks Research Group
[2] Operations Research and Stochastics Research Group

ABSTRACT

The problem of localizing nodes without relying on GPS by employing a small fraction of anchor nodes, which are aware of their positions, is considered to be an important service for a wide variety of applications in wireless ad hoc networks. Several approaches aiming at defeating attackers by means of robustness instead of pure cryptographic measures have been proposed. Yet they all are making assumptions primarily focused on single-hop based localization schemes. Hence, we discuss threats and requirements to be fulfilled by multi-hop based localization mechanisms and compile a comprehensive attacker model. Furthermore, we discuss shortcomings of existing evaluation techniques and propose a unified approach for evaluating multi-hop based localization schemes. Finally, utilizing the mean and the median to estimate and select a hop length, we evaluate the suitability of Linear Least Squares (LLS) based on the l_2-Norm and provide a comparative evaluation with LLS based on the l_1-Norm, a l_1-Norm based filtering scheme and the well-known Least Median of Squares (LMS) approximation employing each norm. We show that the l_1-Norm is able to reduce the mean estimation error and that LMS as well can benefit from the l_1-Norm. Additionally, we argue that, assuming a DV-hop scheme, LMS is not able to meet its requirements when using the mean for hop length estimation and selection. We conclude that employing the median in this context leads to more accurate results.

Index Terms— Secure localization, Optimization, Multi-hop

1. INTRODUCTION

Wireless Ad Hoc Networks offer a wide variety of possible applications in the context of sensor networks ranging from environmental monitoring to intrusion detection and battlefield surveillance. An important service for these applications is the localization of the participants without relying on GPS. Thus the problem of localizing sensors using only a small fraction of anchor or beacon nodes, which are aware of their location, has gained much attention from researchers in the past decade. A lot of proposed schemes assume cooperative behavior among the nodes to achieve localization. Since many applications require the deployment of nodes in an adversarial environment, the problem of providing secure and robust localization has led to a variety of mechanisms aiming at either preventing the attacker from disturbing the process of localization, detecting and repairing such intervention or using robust statistical methods to offer graceful degradation in the face of an attack. Yet most approaches require single-hop communication between nodes and anchors to conduct distance measurements. In contrast, multi-hop schemes require only a small amount of anchor to enable localization. Such mechanisms may provide a rough estimation by having the anchors flood their location into the network. Nodes receiving this information forward it to their neighbors while either incrementing a hop count value or summing up a distance measurement contained in the packets traveling through the network. Participants receiving the location (x_i, y_i) and the measured distance d_i to at least three anchor nodes are able to estimate their own location by applying lateration to the set of references (x_i, y_i, d_i). The nodes may subsequently refine this estimate by iteratively exchanging locations and distance measurements with their neighbors. Despite their promising nature, not much effort has been made in the past to secure multi-hop localization schemes.

In this paper, we make the following contributions. First, we discuss threats and requirements to be fulfilled by multi-hop based localization mechanisms and compile a comprehensive attacker model. Second, we discuss shortcomings of existing evaluation techniques and propose a unified approach for evaluating multi-hop based localization schemes. Finally, we investigate the feasibility of the l_1-Norm in contrast to the widely-used l_2-Norm to enhance the robustness of existing multi-hop based localization schemes. Additionally, we provide a formal description of Linear Least Squares (LLS) motivating the use of the l_1-Norm. We are able to support our assumption in a comparative evaluation with Linear Least Squares based on the l_1-Norm and the l_2-Norm, a new l_1-Norm based filtering approach and the Least Median of Squares (LMS) approximation [1]. By employing both the mean and me-

dian to estimate and select the hop length required by the multi-hop based DV-hop, we are able to show that, by using the mean, LMS is not able to hold against a basic attack. Hence we investigate the source of this behavior and are able to show that only by applying the median, LMS based schemes may be considered a robust solution.

2. MULTI-HOP LOCALIZATION

In contrast to single-hop localization schemes, which require connectivity to at least three anchor nodes, multi-hop based techniques aim at supplying participants with the necessary beacon coordinates and distance measurements aggregated over several hops. Three of the most well-known approaches for multi-hop localization are evaluated by LANGENDOEN and REIJERS [2], identifying the following common phases: Distance measurement phase, location estimation phase and an optional refinement phase. Since the refinement phase depends on the correctness of first two phases, we concentrate our efforts on the inspection of their vulnerabilities. Nevertheless, it should be noted that the refinement phase may very well provide an attacker with additional opportunities for tampering with the results of the localization process. We also subdivide the description of the distance measurement phase in the following section into several sub-sections to emphasize the different steps of the protocol.

2.1. Distance measurement

Location dissemination: In the distance measurement phase references containing the location of the corresponding anchor nodes are flooded into the network. Since flooding is very expensive, SAVARESE et al. propose for each node to stop broadcasting further messages once it has received a sufficient number of references to estimate its own location [3]. Another way to reduce the communication overhead is proposed by ZENG et al. [4]. They suggest to use a fixed hop limit known by all participants and have nodes drop a packet once its hop count value exceeds the given hop limit. Upon receiving a location references, participating nodes (i.e. nodes and anchors) adjust an aggregated measurement value in the received packet by adding the measured distance to its sender. In case the received message offers some new or useful information to the participant (reception of a reference from a previously unknown beacon node or a shorter route to an already known anchor), the node then broadcasts the modified message. Nodes may either use range-based methods requiring specialized hardware, possibility involving a ranging protocol and a measurement message exchange (*DV-distance* [5]) or connectivity information to sum up hops in order to measure the distance to another node (*DV-hop* [5, 3]). The second option additionally demands each node to obtain an estimate of the length of a hop in order to translate the number of hops to a distance.

Hop length estimation: The above mentioned hop length estimate required by DV-hop is calculated by anchors using the information received from other beacon nodes. Since anchors know their own location, upon the reception of the coordinate of another anchor, a beacon node is able to compute the euclidean distance between itself and its counterpart and divide the calculated distance by the number of hops the corresponding packet has traveled. An anchor may also receive a packet from more than just one other anchor. Thus multiple hop length estimates resulting from anchors are merged by calculating the mean of estimates. Anchors may also use the median instead of the mean to estimate the hop length to limit the potential influnce of an attacker [4]. After having estimated a hop length, the anchor broadcasts its estimate into the network using a second message to provide nodes with the necessary information. Like proposed by ZENG et al. [4], it is also possible to have anchors broadcast messages in regular intervals and piggy-back the hop length estimate (if available) with the regular location information.

Hop length selection: Although only one hop length estimate is required at each node, it typically receives multiple hop length estimates from different anchors. They are combined calculating the mean of the corresponding values, although, as we will show, using the median instead of the mean is assumed to be more robust in the face of an attack.

2.2. Location estimation

After having received at least three references, the nodes estimate their own location using one of the approaches described in more detail in section 3. If there are only one or two references available, a limited guess about the rough position of the node might still be possible, but is generally considered to be too error-prone to be used for location estimation. Therefore any estimation technique running at a node is presumed to fail with a node holding less than three references.

3. LOCATION ESTIMATION TECHNIQUES

In order to estimate a location from a set of references (x_i, y_i, d_i), several mechanisms based on solving an optimization problem have been proposed [5, 3, 6]. For this class of algorithms, to determine its position, each node performs multilateration. In this work we only consider the two-dimensional case without loss of generality. In case of N given references (x_i, y_i, d_i) with $N \geq 3$, not all lying on a straight line, each reference is interpreted as a circle with center (x_i, y_i) and radius d_i, $\forall i \in \{1, \ldots, N\}$. Ideally, assuming distance measurements free of errors and manipulation, this would

result in a set of circles intersecting at the coordinate, at which the node resides. Hence it would be sufficient to solve a system of non-linear equations to find this point of intersection:

$$(x_i - x)^2 + (y_i - y)^2 = d_i^2 \;\; (i = 1, \ldots, N)$$

One approach for being able to deal with error-prone measurements is to consider the above over-determined system of equations. Due to the measurement errors or manipulated distance values, this system is in general unsolvable. Therefore, we define the *assumed position* $(\hat{x}, \hat{y})$ as:

$$\begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} := \arg\min_{x,y} \sum_{i=1}^{N} \left[\sqrt{(x - x_i)^2 + (y - y_i)^2} - d_i \right]^2$$

Solving this non-linear optimization problem is too expensive as it usually requires techniques of global optimization. Therefore, this problem is approximated as described below.

The system of non-linear equations described above can be linearized by subtracting the mean of all left and right parts [1]

$$\frac{1}{N}\sum_{i=1}^{N}(x_i - x)^2 + (y_i - y)^2 = \frac{1}{N}\sum_{i=1}^{N} d_i^2$$

resulting in the matrix form $\mathbf{Ax} = \mathbf{b}$ of the system, where

$$\mathbf{A} = \begin{pmatrix} x_1 - \frac{1}{N}\sum_{i=1}^{N} x_i & y_1 - \frac{1}{N}\sum_{i=1}^{N} y_i \\ \vdots & \vdots \\ x_N - \frac{1}{N}\sum_{i=1}^{N} x_i & y_N - \frac{1}{N}\sum_{i=1}^{N} y_i \end{pmatrix} \in \mathbb{R}^{(N,2)}$$

and

$$\mathbf{b} = \frac{1}{2}\begin{pmatrix} (x_1^2 - \frac{1}{N}\sum_{i=1}^{N} x_i^2) + (y_1^2 - \frac{1}{N}\sum_{i=1}^{N} y_i^2) - \\ (d_1^2 - \frac{1}{N}\sum_{i=1}^{N} d_i^2) \\ \vdots \\ (x_N^2 - \frac{1}{N}\sum_{i=1}^{N} x_i^2) + (y_N^2 - \frac{1}{N}\sum_{i=1}^{N} y_i^2) - \\ (d_N^2 - \frac{1}{N}\sum_{i=1}^{N} d_i^2) \end{pmatrix} \in \mathbb{R}^N$$

Another way to linearize this overdetermined system of equations is to subtract one single equation from every other [5]. However, in this case the risk of a malicious reference to strongly affect the benign ones might be higher. We therefore assume that it is more desirable to use the mean of all equations in order to reduce the potential influence of one or more attackers.

For $1 \leq p < \infty$ the l_p-Norm of a vector $\mathbf{r} = (r_1, \ldots, r_N) \in \mathbb{R}^N$ is defined as:

$$\|\mathbf{r}\|_p := \left(\sum_{i=1}^{N} |r_i|^p \right)^{\frac{1}{p}}$$

With this, we estimate the assumed position according to:

$$\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \arg\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_p$$

Furthermore, it is easy to see that $\bar{\mathbf{x}}$ is a solution of

$$\underset{\mathbf{x}}{\text{MIN}} \|\mathbf{Ax} - \mathbf{b}\|_p$$

if and only if $\bar{\mathbf{x}}$ is a solution of

$$\underset{\mathbf{x}}{\text{MIN}} \|\mathbf{Ax} - \mathbf{b}\|_p^p$$

This allows to use a simplified way of computing $\bar{\mathbf{x}}$ as described below.

3.1. l_2 – Linear Least Square (l_2-LLS)

The value of p may be selected accordingly to fit a specific application and is, in terms of localization in wireless ad hoc networks, typically set to $p = 2$. The solution of the resulting problem

$$\begin{aligned} \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} &= \arg\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2 \\ &= \arg\min_{\mathbf{x}} \left\{ \sum_{i=1}^{N} (\mathbf{a}_i\mathbf{x} - b_i)^2 \right\} \end{aligned}$$

with $\mathbf{a_i}$ denoting the i^{th} row of matrix $\mathbf{A}$ and b_i denoting the i^{th} component of vector $\mathbf{b}$ can be computed by solving a system of linear equations. For this, QR-factorization is applied to $\mathbf{A}$:

$$\begin{aligned} &\mathbf{A} = \mathbf{QR} \text{ where } \mathbf{Q} \in \mathbb{R}^{(N,N)}, \mathbf{R} \in \mathbb{R}^{(N,2)} \\ &\mathbf{Q}^{-1} = \mathbf{Q}^T, \; \mathbf{R} = \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{pmatrix} \\ &\mathbf{A} = \mathbf{Q}\begin{pmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{pmatrix} = (\mathbf{Q}_1, \mathbf{Q}_2)\begin{pmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{pmatrix} = \mathbf{Q}_1\mathbf{R}_1 \end{aligned}$$

where R_1 is an upper triangular matrix and (Q_1, Q_2) is an appropriate decomposition of Q. Hence, we obtain:

$$\begin{aligned} \|\mathbf{Ax} - \mathbf{b}\|_2^2 &= (\mathbf{Rx} - \mathbf{Q}^T\mathbf{b})^T\mathbf{Q}^T\mathbf{Q}(\mathbf{Rx} - \mathbf{Q}^T\mathbf{b}) \\ &= (\mathbf{Rx} - \mathbf{Q}^T\mathbf{b})^T(\mathbf{Rx} - \mathbf{Q}^T\mathbf{b}) \\ &= \left|\left| \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{pmatrix}\mathbf{x} - \begin{pmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{pmatrix}\mathbf{b} \right|\right|_2^2 \\ &= \left|\left|\mathbf{R}_1\mathbf{x} - \mathbf{Q}_1^T\mathbf{b}\right|\right|_2^2 + \left|\left|\mathbf{Q}_2^T\mathbf{b}\right|\right|_2^2 \end{aligned}$$

Finally, by solving $\mathbf{R}_1\bar{\mathbf{x}} - \mathbf{Q}_1^T\mathbf{b} = \mathbf{0}$, an estimate for the assumed position $\hat{\mathbf{x}}$ of the node can be retrieved with $\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2 = \left|\left|\mathbf{Q}_2^T\mathbf{b}\right|\right|_2^2$.

3.2. l_1 – Linear Least Square (l_1-LLS)

While following the method described above is considered to be the most feasible of all optimization-based approaches for use in wireless ad hoc networks, it is

vulnerable to malicious references forging the location or the distance to an anchor. Compared to the l_2-Norm, the l_1-Norm is generally less vulnerable to outliers contained in the data, e.g. caused by measurement errors. Therefore, the use of the l_1-Norm might increase the robustness of existing localization schemes against attackers.

Employing the l_1-Norm, the assumed position of a node is estimated by solving the following optimization problem:

$$\begin{aligned}\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix} &= \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 \\ &= \arg\min_{\mathbf{x}} \left\{\sum_{i=1}^{N} |\mathbf{a}_i\mathbf{x}-b_i|\right\}\end{aligned}$$

This problem can be formulated as a linear optimization problem by introducing a vector

$$\mathbf{h} = (h_1, h_2, \ldots, h_N) \in \mathbb{R}^N$$

of auxiliary variables $h_i \geq 0, \forall i \in \{1, \ldots, N\}$. With this, the optimization problem can be rewritten as follows:

$$\begin{aligned}&\underset{\mathbf{x}}{\text{MIN}} \left\{\sum_{i=1}^{N} |\mathbf{a}_i\mathbf{x}-b_i|\right\} \\ \Leftrightarrow\quad &\underset{\mathbf{x},\mathbf{h}}{\text{MIN}} \left\{\sum_{i=1}^{N} h_i \;\middle|\; -\mathbf{h} \leq \mathbf{A}\mathbf{x}-\mathbf{b} \leq \mathbf{h}\right\}\end{aligned}$$

With this, we obtain

$$\underset{\mathbf{x},\mathbf{h}}{\text{MIN}} \left\{ (\mathbf{0}^T, \mathbf{1}^T)\begin{pmatrix}\mathbf{x}\\ \mathbf{h}\end{pmatrix} \middle| \begin{pmatrix}\mathbf{A} & -\mathbf{E}\\ -\mathbf{A} & -\mathbf{E}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{h}\end{pmatrix} \leq \begin{pmatrix}\mathbf{b}\\ -\mathbf{b}\end{pmatrix}\right\}$$

where $\mathbf{E}$ is the identity matrix of dimension N, $\mathbf{0} = (0,0)^T \in \mathbb{R}^2$ and $\mathbf{1} = (1,1,\ldots,1)^T \in \mathbb{R}^N$.

Although estimating the assumed position using the l_1-Norm involves solving a linear optimization problem, we note that the complexity of this problem is comparatively small with up to $N+2$ variables and $2N$ constraints and already solvable in a few millisecond on a present 2.4 GHz machine for a set of $N = 50$ two-dimensional references. We thus assume that this kind of processing power might very well be available to future applications, which are not as resource constrained as nodes in today's sensor networks.

4. THREAT ANALYSIS

There are a number of known attacks, which can be launched against multi-hop schemes [4], assuming that the goal of an attacker is to msilead nodes to obtain an incorrect position. Here it is important to differentiate *how many* nodes the attacker wants to manipulate. He or she might try to influence as many nodes as possible or target a few specific nodes. Therefore, when evaluating secure localization schemes, either the mean or maximum estimation error can be of interest.

We now give a short overview of the possibilities of an attacker with the above goals in a multi-hop localization environment. It should be noted that the following attacks may also be launched by multiple colluding nodes in order to increase their strength. In this case the ratio of malicious participants among benign ones is defined as *contamination ratio* ϵ. This ratio usually refers to the corresponding originator of the attack, i.e. if launched by an anchor, the contamination ratio refers to the ratio of malicious anchors among benign ones.

Sybil attack: Robust localization schemes trying to filter malicious data instead of employing means of cryptography (section 5) usually rely on the majority of references (more than 50%) originating from benign anchors. Therefore, one of the potentially most harmful threats is the Sybil attack, where an attacker forges multiple identities. By distributing faked messages for a specific number of anchors, an attacker might be able to control the majority of references in the network. Benign anchor references then might be classified as outliers by the robust localization algorithm, possibly even augmenting the strength of the attack.

Faked anchor location: Assuming a mechanism to prevent the Sybil attack, a trusted anchor might still get compromised. Hence an attacker could be able to fabricate messages for a corresponding anchor. By announcing a forged location or a faked initial distance measurement value (e.g. increased initial hop count), a node trying to localize itself might be mislead about the position of the anchor, assuming itself to reside at a forged location.

Faked hop length: Apart from faking its location, in DV-hop based localization schemes, a compromised anchor might distribute a manipulated hop length estimate. Since this estimate is used by nodes to translate the number of hops to anchors to a distance value, a node trying to estimate its location might be mislead about its whereabouts.

Reference modification: In addition to the attacks involving anchors described above, malicious nodes may pose a possible threat as well. Apart from the risk of attackers modifying the location or hop length estimate inside anchor messages being forwarded, which could be countered by data origin authentication, participants are required to modify the contents of anchor messages for the sake of adjusting the distance measurement value accordingly (i.e. increase the hop count by one). Hence, even assuming data origin authentication, the manipulation of this value (incrementing, decrementing or setting the distance resp. hops) is considered a threat to the process of localization.

Wormhole attack: Assuming protection against the threats mentioned above, an attacker might still employ a wormhole attack [7] to manipulate a great number of benign references. Since in a wormhole attack an attacker may tunnel and replay anchor messages at a remote location without modifying the distance measurement value, nodes may mistakenly assume a shorter route to the respective beacon node actually residing at a more distant location. This also poses a possible threat to robust localization schemes since any message from any benign anchor gets manipulated implicitly by traveling through the wormhole. Hence, for an attacker, manipulating a great number of the references is just a matter of tunneling messages from as many different anchors as possible.

Pollution attack: Concerning the pollution attack [8], an attacker might make use of the fact, that subsets of benign references may agree upon a node possibly residing at a location different from the actual estimated coordinate. This especially is a problem for overlap-based approaches, where shapes typically overlap at different locations, with the estimate being selected by the area where most references overlap. Thus, by forging or manipulating anchor messages and using a different location where a subset of benign references overlap, an attacker might be able to create an area with a greater number of overlapping references (compared to the benign area of intersection).

5. COUNTERMEASURES

Several methods to defeat one or more of the threats described in section 4 have been proposed. They can be divided in the following two classes:

- *Attack detection and response* aiming at detecting malicious nodes and responding accordingly to their presence.
- *Robust localization* techniques trying to filter malicious references by employing statistical methods for outlier detection before computing a location estimate.

In this work, we focus on the second class of robust localization schemes. With this class of approaches, any manipulation of the references by one or more attackers should ideally be unsuccessful, as long as more than 50% of the references are still benign. Consequently, this is only possible if attackers are unable to launch a Sybil attack and forge references for one or more non-existing anchors. In the past, several robust localization schemes have been proposed [1, 9, 10, 11]. Here, we only consider the well-known LMS algorithm [1], as well as a new technique described in section 5.2, focusing on possible advantages of using the l_1-Norm compared to the commonly used l_2-Norm in the context of a multi-hop based localization scheme.

5.1. Least Median of Squares (LMS)

The optimization problem defined in section 3 is not robust against outliers. Therefore LI et al. [1] propose the use of the following optimization problem as a starting point for estimating the assumed position:

$$\begin{pmatrix}\hat{x}\\ \hat{y}\end{pmatrix} = \arg\min_{x,y} med_i \sum_{i=1}^{N} \left[\sqrt{(x-x_i)^2+(y-y_i)^2} - d_i\right]^2$$

Solving this non-linear problem exactly is again too complex. Thus this problem is approximated by using the following algorithm, which is based on estimating assumed positions for a series of randomly selected subsets of the N given references [1, 12]:

1. Select a subset size n (typically $n = 4$).
2. Randomly draw M subsets of size n from the set of given references. Compute a location estimate for each subset $j = 1, \ldots, M$ using l_2-LLS and calculate the median of the estimation residuals r_{ij}^2 to each anchor $i = 1, \ldots, N$.
3. Define $m = \arg\min_j med_i \left\{r_{ij}^2\right\}$ (least median of all medians of each subset).
4. Calculate $s_0 = 1.4826(1 + \frac{5}{N-2})\sqrt{med_i r_{im}^2}$.
5. Assign a weight w_i to each reference:

$$w_i = \begin{cases} 1 & |\frac{r_i}{s_0}| \leq 2.5 \\ 0 & otherwise \end{cases}$$

6. Compute a weighted least squares of all given references using weights w_i. This is equivalent to computing l_2-LLS with only the references with weight $w_i = 1$.

The probability of drawing at least one benign subset of size n from M drawings, given the contamination ratio ϵ, can be calculated as follows:

$$P = 1 - \left(1 - (1-\epsilon)^n\right)^M$$

Therefore, the required number of drawings M can be computed for an assumed or known contamination ratio ϵ and a selected probability P:

$$M = \left\lceil \frac{\log(1-P)}{\log(1-(1-\epsilon)^n)} \right\rceil$$

For example, given $P = 0.99$, one would have to select $M = 17$ subsets for $\epsilon = 0.3$ and $M = 34$ subsets for for $\epsilon = 0.4$.

In order to capture the basic idea behind this algorithm, please note that it computes an assumed position for each subset of anchors, and that for each assumed position it computes the median of the differences of distances to all anchor points (step 2). Step 3 selects the minimum of all these medians which will be smaller or equal to a value obtained with the assumed position for

one of the benign subsets (under the assumption that less than 50% of anchors are manipulated). Based on this value, steps 4 to 6 decide which anchor points will be ignored in the position estimation (see [1, 12] for further explanations).

Apart from evaluating LMS using l_2-LLS (which from now on we will refer to as l_2-*LMS*), we also consider a variation of LMS employing l_1-LLS which we call l_1-*LMS*.

5.2. $\bar{\epsilon} - l_1 - LLS$ – filtering

In addition to the existing approaches, we describe a simple new optimization-based technique called $\bar{\epsilon} - l_1 - LLS$ – *filtering*, which uses a preprocessing step to discard the most suspicious location references using the l_1-Norm. As mentioned previously in section 3.2, when minimizing the residues, the l_1-Norm is known to be more robust against outliers than any other l_p-Norm. Thus, after estimating the coordinate of a node from all given references in a first step, a specific ratio of $\bar{\epsilon}$ references with the largest residues is discarded in a second step ($\bar{\epsilon}$ is a fixed contamination ratio that is assumed by the nodes of the network; note that the true contamination ratio is unknown to the nodes). Finally, the location of a node is estimated using the remaining references. With the outlier detection capability of the l_1-Norm, we expect the residues of the malicious references to be greater than the benign ones. Therefore when filtering the references with the $\lceil N \cdot \bar{\epsilon} \rceil$ largest residues, we assume to already achieve a notable improvement in reducing the influence of an attacker.

6. EVALUATION

To evaluate the suitability of the l_1-Norm and compare it to the l_2-Norm, we implemented DV-hop using OMNeT++ (http://www.omnetpp.org/) and the MiXiM framework (http://mixim.sourceforge.net/) and incorporated the LLS and LMS approaches for both the l_1 and l_2-Norm as well as the filtering approach described in section 5.2. Each LMS variant was evaluated using $M = 17$ and $M = 34$ subsets, which corresponds to a probability $P = 0.99$ for contamination ratio $\epsilon = 0.3$ and $\epsilon = 0.4$. Accordingly, we analyzed our filtering approach discarding a fixed ratio of $\bar{\epsilon} = 0.3$ and $\bar{\epsilon} = 0.4$.

The evaluation of existing approaches typically only involved a simple Matlab based simulation employing a fixed radio range, comparatively high node densities of 300 nodes on a field of 100m $\times$ 100m [4] and simple distance measurement error models using a Gaussian distribution [1] as well as single-hop based experiments with about 15 nodes using a very small field size of about 18m $\times$ 18m [9]. These approaches use different settings and assumptions, lacking the simulation of the process of reference distribution, hop length estimation and selection, as well as mechanisms for reducing the traffic caused by flooding in the network (i.e. by using a hop limit). Furthermore, experiments deploying motes on a relatively small field only consider a single-hop scenario and are not suited for evaluating the behavior of robust localization schemes in a large-scale multi-hop environment. In addition, high node densities might not be a very realistic assumption for future applications.

Hence, we decided to head for a unified and more realistic approach. Therefore, in addition to implementing the multi-hop based DV-hop approach, we incorporated the IEEE 802.11 radio model provided by MiXiM with its pathloss model and randomly placed 400 nodes on a field of size 2000m $\times$ 2000m using the uniform distribution. In our scenario the applied transmission power of 110.11mW roughly corresponds to a transmission range of about 200m up to a maximum of 300m in very rare cases. We selected the above mentioned field size to be 10 times the size of the approximate radio range, so that many nodes could only hear from anchors over multiple hops. Also, we selected the number of nodes to be 400 to provide a stable yet relatively low node density required to allow all participants to be part of one connected communication graph. Furthermore, we employed a fixed anchor ratio of 10% ($N = 40$). A smaller anchor ratio (e.g. 5%) would have been possible, yet we decided to use 10% to enable a fair comparison between the different algorithms by making sure there were enough references left for localization after some being discarded by the robust schemes.

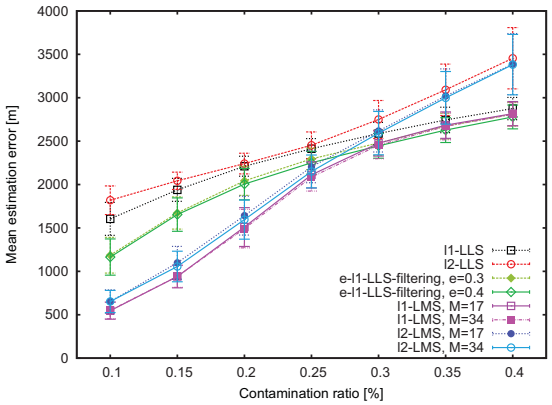
According to the phases of multi-hop localization described in section 2, we implemented DV-hop as follows: Each anchor broadcasts a message containing its location in a total of 10 rounds separated by intervals of roughly 60 seconds for providing the network with enough time to distribute the anchor messages. After anchors having heard of other anchors, they calculate a hop length estimate using either the mean or median of the possible estimates and include this information in the message sent out in the next round. Although two rounds should therefore be sufficient, we used 10 rounds to make sure messages have a good chance of being distributed in the network and enough anchors being able to estimate a hop length. Furthermore, as mentioned in section 2.1, we employed a hop limit of 8 hops, which was selected to allow every node to localize itself by obtaining references from about 60% of the anchors. Nodes and anchors maintain a table of shortest paths to known anchors. Additionally, nodes keep the most recent hop length estimate announced by the corresponding anchor. Finally, with anchors having finished the 10 rounds, nodes estimate their location by selecting a hop length from the list of available estimates and running each of the localization algorithms with the parameter combinations mentioned above. When selecting a hop length estimate, nodes always employ the same technique currently in use by

the anchors, i.e. with anchors using the median of hop length estimates, nodes also use the median.

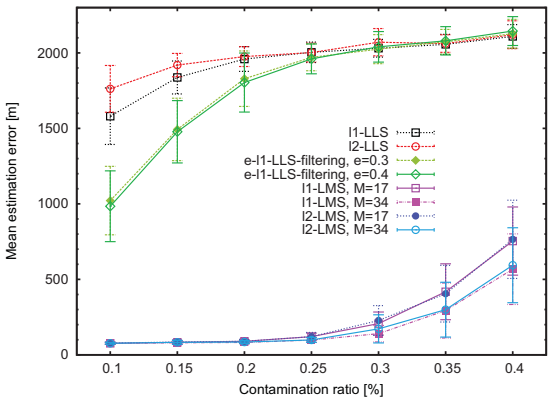
To understand the importance of employing the median instead of the mean, it is necessary to explain the effect of attackers indirectly exploiting honest nodes and anchors to support the attack. Malicious anchors increase the euclidean distance computed at benign anchors, while the number of hops between the anchors remains untouched. Therefore, when estimating the hop length using the mean, attackers may be able to affect benign anchors to announce hop lengths increased due to the influence of the malicious beacon nodes, resulting in a possibly augmented strength of the attack. Consequently, employing the mean, faked hop length estimates resulting from forged anchor locations affect the selected hop length at nodes. With the median being able to ignore outliers up to 50%, it seems reasonable to use the median instead of the mean for hop length estimation and selection to weaken the influence of malicious anchors on benign participants. While the idea of anchors using the median for estimating the hop length is mentioned by ZENG et al. [4], the authors just state the assumption without an evaluation to support it. Furthermore, they only explicitly consider anchors to use the median. In contrast, we employ the median on both nodes and anchors believing that this will allow us to see the full benefit of the median compared to the mean.

To evaluate the robustness of the schemes employing the mean and median, we implemented the *faked anchor location* attack described in section 4. Out of N anchors, $\lceil N \cdot \epsilon \rceil$ are randomly selected to be malicious. The contamination ratio ϵ is varied from 10% to 40%. Malicious anchors fake their location by adding a vector defined by a common direction and length of 4000m to their actual coordinates. To measure the influence of the attacking anchors, we use the *mean estimation error*, which corresponds to the euclidean distance between the actual and estimated location of a node.

First, we evaluated the influence of attacking anchors on the robust localization techniques using the mean for hop length estimation and selection. The results averaged over 15 runs including the 99% confidence intervals (computed with the Student's *t*-distribution) are shown in fig. 6 a. In this scenario, all algorithms are heavily influenced by the attackers with the mean estimation error ranging from 500m to 3500m. As expected, l_2-LLS is most vulnerable to the faked locations, while l_1-LLS is able to maintain a smaller estimation error. This behavior is even more obvious when ϵ is above 25%, where the error for l_2-LLS shows a strong increase compared to l_1-LLS. While the LMS schemes perform better for ϵ below 25%, for contamination ratios above this value, they converge to the curve of their respective LLS variant. We assume this behavior is due to the effect of malicious anchors indirectly influencing benign participants as described above.



(a) Hop length estimated and selected using mean



(b) Hop length estimated and selected using median

Fig. 1. Mean estimation error for faked location attackers declaring their location 4000m away from their actual coordinates. The lines are shown for the sake of enhanced readability.

In this case the actual contamination ratio due to this effect might be considerably higher than declared by ϵ, resulting in the LMS techniques to loose their filtering ability. The advantage of using a higher number of $M = 34$ subsets is also visible, yet considering the necessary double of computational overhead, the advantage of drawing more subsets is not as clear as to be expected. Similarly, our filtering approach performs about the same for $\bar{\epsilon} = 0.3$ and $\bar{\epsilon} = 0.4$. Although it perform worse than the LMS approaches for $\epsilon < 0.25$, it is able to perform better compared to l_1-LLS. Like l_1-LMS, it converges to l_1-LLS for $\epsilon > 0.25$.

As mentioned above, although showing individual behavior, when using the mean to estimate and select the hop length, no technique is able to contain the attack. In contrast, by employing the median, all algorithms show a reduction of the mean estimation error (fig. 6 b). Yet there are significant differences in the amount of decrease. While l_1-LLS, l_2-LLS and our filtering approach are contained at an error of about 2100m for $\epsilon = 0.4$, the LMS schemes are able to keep

an error of about 750m. They perform best for $\epsilon < 0.25$ with an error of about only 100m independent of the norm and number of subsets. In case of $\epsilon > 0.25$, l_1-LMS shows a slight advantage, while a higher number of subsets shows the expected behavior of improved outlier detection. Our filtering approach is not able to compete with the LMS variants in this setting. Nevertheless, for contamination ratios below 20%, it is able to reduce the mean estimation error. This might be due to the fact, that, in this case, by discarding 30% or 40% of the references with the largest residues, a significant fraction of the malicious data is filtered out correctly.

7. CONCLUSIONS AND FUTURE WORK

In this work, we described possible attacks concerning multi-hop based approaches and proposed the use of the l_1-Norm instead of the l_2-Norm to increase the robustness of LLS and LMS methods. We evaluated this proposal employing a unified and more realistic simulation model and were able to clearly show the advantage of the l_1-Norm assuming the mean for estimating and selecting the hop length. Nevertheless, in this setting even LMS-based approaches were unable to contain the influence of a basic *faked anchor location* attack. Therefore, we considered employing the median to estimate and select the hop length to filter the influence of benign participants indirectly supporting the attack. We were able to show that, by using the median, the estimation error decreases for all algorithms, with the LMS schemes most notably benefiting from this change. Hence, we conclude that in multi-hop environments LMS leads to more accurate results applying the median instead of the mean to estimate and select hop lengths.

In our future work, we plan to further investigate the performance of robust localization schemes according to the unique properties and threats of multi-hop based approaches. Therefore, we aim at compiling a more comprehensive survey of existing methods including a decomposition of approaches and a comparative evaluation under unified assumptions.

8. REFERENCES

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

[2] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks: a quantitative comparison," *Computer Networks*, vol. 43, no. 4, pp. 499–518, 2003.

[3] Chris Savarese, Jan M. Rabaey, and Koen Langendoen, "Robust Positioning Algorithms for Distributed Ad-Hoc Wireless Sensor Networks," in *ATEC '02: Proceedings of the General Track of the annual conference on USENIX Annual Technical Conference*, Berkeley, CA, USA, 2002, pp. 317–327, USENIX Association.

[4] Yingpei Zeng, Shigeng Zhang, Shanqing Guo, and Xie Li, "Secure hop-count based localization in wireless sensor networks," in *CIS '07: Proceedings of the 2007 International Conference on Computational Intelligence and Security*, Washington, DC, USA, 2007, pp. 907–911, IEEE Computer Society.

[5] Andreas Savvides, Chih-Chieh Han, and Mani B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *MobiCom '01: Proceedings of the 7th annual international conference on Mobile computing and networking*, New York, NY, USA, 2001, pp. 166–179, ACM.

[6] A. Savvides, H. Park, and M.B. Srivastava, "The n-hop multilateration primitive for node localization problems," *Mobile Networks and Applications*, vol. 8, no. 4, pp. 443–451, 2003.

[7] Y.C. Hu, A. Perrig, and D.B. Johnson, "Packet leashes: a defense against wormhole attacks in wireless networks," in *IEEE Societies INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications*, 2003, pp. 1976–1986.

[8] Yingpei Zeng, Jiannong Cao, Shigeng Zhang, Shanqing Guo, and Li Xie, "Pollution attack: a new attack against localization in wireless sensor networks," in *WCNC'09: Proceedings of the 2009 IEEE conference on Wireless Communications & Networking Conference*, Piscataway, NJ, USA, 2009, pp. 2038–2043, IEEE Press.

[9] Donggang Liu, Peng Ning, and Wenliang Kevin Du, "Attack-resistant location estimation in sensor networks," in *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, Piscataway, NJ, USA, 2005, p. 13, IEEE Press.

[10] N. Kiyavash and F. Koushanfar, "Anti-collusion position estimation in wireless sensor networks," in *IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)*, 2007, pp. 1–9.

[11] X. Li, B. Hua, Y. Shang, and Y. Xiong, "A robust localization algorithm in wireless sensor networks," *Frontiers of Computer Science in China*, vol. 2, no. 4, pp. 438–450, 2008.

[12] P.J. Rousseeuw and A.M. Leroy, *Robust regression and outlier detection*, Wiley-IEEE, 2003.