Localization Anomaly Detection in Wireless Sensor Networks for Non-flat Terrains

Sireesha Krupadanam and Huirong Fu

Member, IEEE

Abstract. In this paper, we develop and evaluate a Localization Anomaly Detection (LAD) scheme for non-flat surfaces for wireless sensor networks. The beacon-less grid localization scheme proposed for non-flat terrains in [1] is used for localization and the localization anomaly detection uses observations of the sensor node at two different reception ranges. Moreover, a new Signal Strength (SS) Metric is proposed and evaluated for LAD. Simulations show that the beacon-less localization method when combined with the LAD scheme gives good detection rates with low false positive rates for the proposed Signal Strength Metric and the Difference Metric. The results show that Signal Strength Metric is comparable to existing metrics while being more difficult to attack.

Keywords: Wireless Sensor Network, LAD, Attack, Localization, Non-flat Terrain.

1 Introduction

Many Wireless Sensor Networks (WSNs) are deployed in unattended and often hostile environments such as those in military and homeland security operations. Therefore, security mechanisms providing confidentiality, authentication, data integrity, and non-repudiation, among other security objectives, are vital to ensure proper network operations. A future WSN is expected to consist of hundreds or even thousands of sensor nodes. This renders it impractical to monitor and protect each individual node from either physical or logical attack.

Location Anomaly Detection (LAD) is the ability of the sensor network to detect anomalies in the reported locations of the sensors. The anomalies may be caused by malicious attacks against the localization scheme to corrupt the sensor locations and thereby render the network measurements worthless. Most of the proposed techniques for location anomaly detection described in the literature cannot be applied to wireless sensor networks as they are computationally intensive. Furthermore, almost all of these are applicable to sensor networks with beacon nodes.

Lazos and Poovendran proposed a Secure Range-independent Localization (SeRLoc) [4] scheme that assumes that the network is comprised of sensor nodes and anchors/locators. In [5], Tang et al. presented a RSSI based cooperative anomaly detection scheme to detect physical displacement attack in Wireless Sensor Networks. However there is not much work done in the area of location anomaly detection using beacon-less localization schemes. The authors Du, Fang and Ning in [2] proposed a

P. Primet et al. (Eds.): GridNets 2008, LNICST 2, pp. 175 – 186, 2009.

general scheme to detect localization anomalies that are caused by adversaries. Their method uses the deployment knowledge and actual observations of the nodes to detect localization anomalies in beacon-less sensor networks. Their scheme has been used as the basis for the proposed grid based location anomaly detection method.

In this research, we propose a LAD method applicable to non-flat terrains based on the Beacon-less grid localization [1]. We discuss various attacks on localization and evaluate the impact of the attacks on the performance of our proposed scheme in terms of ROC curves for different metrics, degrees of damage, percentages of compromised nodes, node density, and different attacks.

Moreover, a new Signal Strength Metric to detect anomalies in the localization is proposed and evaluated. The Difference metric, Add-all metric, Probability metric proposed by Du et al. [2] are modified and evaluated for the various cases. Our simulation results show that the proposed location anomaly detection method gives good results for the sensor networks deployed over non-flat terrains. The results show good detection rates and low false positive rates.

The organization of the paper is as follows: section II describes the proposed Beacon-less Grid Localization method, section III describes the proposed Location Anomaly Detection, section IV, introduces the new Signal Strength metric, section V presents simulation results of the proposed LAD method, and section VI presents the simulation results for the Signal Strength Metric. Section VII compares the localization results and section VII provides a summary and proposed future work.

2 Beacon-Less Grid Localization for Non-flat Terrains

For beacon-less grid localization, two different deployment models are studied for non-flat terrains. The terrain used is shown in Figure 1, which is the terrain for Sea-cliff [3]. In the Static deployment case, it is assumed that the sensor nodes fall around the deployment point in a Gaussian distribution and stay at their drop point on the non-flat terrain. In the Dynamic deployment case, the sensor nodes slide to their final locations based on the surface characteristics. The results presented in this paper are mainly for the more challenging dynamic deployment.

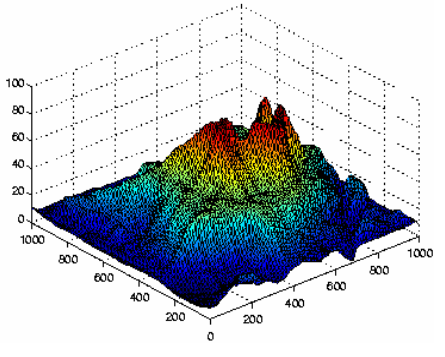

Fig. 1. Sea-cliff Terrain

2.1 Static Node Deployment

For a deployment over a 1000mx1000m area divided into a 10x10 grid of 100mx100m each. The deployment points would be the centroids of each grid section.

2.2 Dynamic Node Deployment

In this case, the initial distributions of the nodes follow the Gaussian distribution as in the case of Static Node Deployment. However, due to the surface characteristics of the terrain, the nodes will not remain at the same location after the deployment. The nodes slide/roll to their final locations. This is modeled using the information about the terrain and the characteristics of the nodes and is described in [1].

2.3 Sensor Node Localization

After sensors are deployed, each sensor broadcasts its group id to its neighbors, and each sensor can count the number of neighbors from group G_i, for i = 1,….n, within a radius of the transmission range R. Assume that a sensor finds out that it has $o_1,...,o_n$ neighbors from group $G_1,...,G_i$, respectively. The actual observation of the sensor is $o = (o_1,...,o_n)$, where n is the number of deployment points. Based on the actual observation of the sensor, and using the deployment knowledge, it finds the nearest location where the expected observation $\mu = (\mu_1,..., \mu_n)$ is closest to o. This location is the localization. The difference between the actual and the localization is the localization error.

In our localization method, the deployment area is divided into a grid of sufficiently high resolution. The resolution is chosen such that the constraints on sensor node memory are satisfied. Each sensor is equipped with the expected observations only for the localization grid points. The expected observation of each grid point p_i, for i=1… l, where l is the number of grid points, is defined as $\xi = (\xi_1,.., \xi_n)$ where $\xi_i = \{\mu \mid \text{location} = p_i\}$. The sensor finds the grid point whose expected observation ξ is closest to its actual observation o. This grid point location p is taken as the localized position of the sensor node.

3 Localization Anomaly Detection for Non-flat Terrains

The Localization Anomaly Detection (LAD) proposed here uses deployment knowledge to find anomalies in the localization information of the nodes. The observations of sensor nodes within a certain radius are used for Localization. For the detection of location anomalies a different range, called the LAD Range is used for observations. Based on the deployment knowledge and its localization the sensor calculates the number of nodes from each group that should be observed within the LAD range. The

difference between this expected observation for the LAD range and the actual observation provides a measure of the location anomaly. Metrics to quantify the location anomaly are then calculated using this error. The metrics are compared against preset thresholds to determine if there is an anomaly.

After sensors are deployed, each sensor broadcasts its group id to its neighbors, and each sensor can count the number of neighbors from group G_i, for $i = 1,....m$. Assume that a sensor finds out that it has $o_{1,...,}o_n$ neighbors from group $G_1,...,G_i$, respectively within its LAD range. $o = (o_1,...,o_n)$ is the actual observation of the sensor. Based on the localization of the sensor, and using on-board deployment knowledge, it finds the expected observation within the LAD range $\mu = (\mu_1,...,\mu_n)$.

3.1 Attacks on Localization

For localization anomaly detection, we first attack the sensor network by compromising *N* sensor nodes by dec-bounded and dec-only attacks [2]. To simulate the attacks with the degree of damage D, we use the following procedure.

1. A node v at the location L_a is randomly picked, and the actual observation, a obtained.
2. To stimulate the attack against the localization, we add nodes to the neighborhood of L_a such that the localization observation p_i equals the localization observation g_i at a grid point a distance D away. A tolerance of $\sqrt{2}\mathrm{d}_\mathrm{g}$ is used to get grid points in all directions around the attacked node at a distance D where d_g is the grid size. This is done as the node is not exactly at a grid point and so we need to get a list of nodes in a circular annulus close to the intended degree of damage D. From this list of nodes that are compromised a random attack location is chosen. Boundary nodes are not considered in this case.
3. For the Dec-bounded attack, based on the above, the observation of the node **o** is obtained from its actual observation **a** by compromising nodes and using multi-impersonating nodes. The nodes which are compromised are not considered for the localization and are dropped.
4. For the Dec-only attack, the observation of the node o is obtained from its actual observation a by decreasing the nodes used to cause the localization error (**g - p**) in only a decreasing manner.
5. For the locations attacked the expected observations within 100m radius, **μ**, are used for LAD metrics.

3.2 LAD Scheme

For the LAD scheme the deployment of nodes on the surface is simulated a priori for *m* nodes at each deployment point on the surface. The number of sensors from

different groups observed at each grid point is calculated and stored as the expected observation at the grid point.

For the node to determine its location, the sensor finds the actual observation of the node using the neighbor information within the Localization Range, and determines its location as the grid point with the closest expected observation. Using this localization, the sensor finds the expected observation within the LAD range using the a priori simulation data.

Now the sensor compares the actual LAD observation within the expected LAD observation and detects anomalies using the metrics by specifying a threshold. For example, if the sensor node at location A observes the same set of neighbors as the sensor node at location B within the transmission range of $60m$, it is not possible for the LAD scheme [2] to detect the accurate location of the sensor node. However, in the proposed LAD method, it is possible to detect the location quite accurately as this method is using different transmission ranges for the Localization and for the LAD. Thus, if we look at the broader range, i.e. $100m$, the observations are different and hence it is possible to detect the anomaly.

3.3 Metrics

In this research, we extend and implement the *Difference*, *Add-all*, and *Probability* metrics [2]. A new metric called *Signal Strength Metric* (SSM) is proposed and the results of the simulations are described in the following sections.

3.3.1 Difference Metric

The difference metric, *DM*, uses the sum of the differences in the nodes observed from each group for the expected and actual observations.

$$DM = \sum_{i=1}^{n} \left| o_i - \mu_i \right|$$

3.3.2 Add-All Metric

The Add-all metric, *AM*, takes the union of the expected and actual observations and obtains the maximum value of the nodes observed for each group. These larger observations for each node group are summed to arrive at the final metric. The metric shows an increase with increasing localization error.

$$AM = \sum_{i=1}^{n} \max \left\{ o_i , \mu_i \right\}$$

3.3.3 Probability Metric

The Probability metric, *PM*, computes the probability that the localization of a node sees exactly o_i neighbors from Group G_i. We approximate the probability for nodes from each Group G_i as μ_i /m. Thus, if the localization is L_e and X_i represents the number of neighbors that come from the Group G_i, the probability metric is calculated as

$$PM_i = \Pr\left(X_i = o_i \middle| L_e\right) = \binom{m}{o_i} (\mu_i / m)^{o_i} (1 - \mu_i / m)^{(m - o_i)}$$

$$PM = \Sigma PM_i$$

In [8], PM_i is used as the probability metric and any abnormal drop is used to trigger the anomaly alarm. With our approximation of the probability density functions for each node group, probabilities for some node groups are very low both in the case of regular nodes localized to a nearby grid point or attacked nodes incorrectly localized. However the probabilities are very easy to calculate. Hence we modify the probability metric as the sum $PM = \sum PM_i$ for our simulations.

4 Signal Strength Metric

We propose a new metric, the Signal Strength Metric, described by Figure 2, to utilize the information from the strength of the signal received by the sensor node. The strength of signals received from each node can be computed. Using this signal strength, an estimated distance can be obtained for each node within the transmission range. The difference between the expected and observed distribution of nodes is then used as the metric. In the Location Anomaly Detection method, the LAD range observations could themselves be attacked. However, using the distribution of all the nodes against distance ensures that any attack that adds or removes nodes from a particular distance is easily identified by a comparison of the expected and actual distributions. The signal strength metric performs such a comparison.

The distances of all observed nodes from the center to the edge of the transmission range, for a group *i*, are obtained as $\{d_{o_k}\}_i$. Similarly, the expected nodes from group *i* are obtained as $\{d_{\mu_k}\}_i$. The metric in outward direction is computed as

$$SS_{ci} = \int_0^R abs\left(sum\left(\{d_{o_k}\}_i < z\right) - sum\left(\{d_{\mu_k}\}_i < z\right)\right) dz$$

However, this term is biased towards nodes which are close to the node being considered. In order to eliminate this bias, we integrate the cumulative error in observed and expected number of nodes for each group from the edge of the transmission range inwards towards the center. The metric for this calculation is

$$SS_{ei} = \int_R^0 abs\left(sum\left(\{d_{o_k}\}_i > z\right) - sum\left(\{d_{\mu_k}\}_i > z\right)\right) dz$$

The signal strength metric for each observed group is thus computed by adding the center to edge and, edge to center metrics as shown in Figure 2. The values for each node group are then added together to obtain the final metric.

$$SS_i = \sum_i \left(SS_{ci} + SS_{ei} \right)$$

As described earlier, the advantage of the signal strength metric over the previous metrics is that it is able to detect changes in the distribution of the nodes even though the overall number of nodes from each node group is the same. Increasing differences between the observed and expected distribution of distances of the nodes would lead to larger values for the metric.

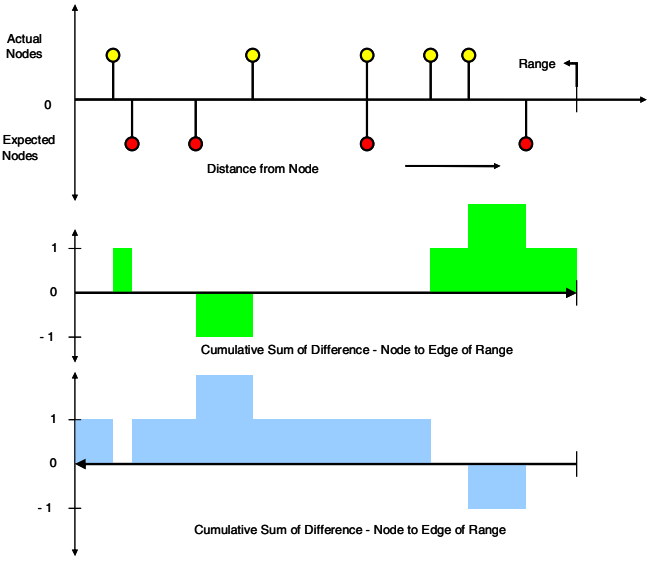

Fig. 2. Signal Strength Metric

The Signal Strength Metric outperforms the difference metric for short transmission ranges as the difference metric only considers the total number of observed nodes from each group without regard to their spatial distribution. With increasing transmission range and node density, more information about the distribution is available through $\{o_i\}$ thereby reducing the advantage of the signal strength metric.

5 Simulation Results

The simulations of the location anomaly detection on the Sea-cliff terrain are described in this section.

5.1 Attacks on Localization

A total of 1000 randomly selected nodes are simulated for values of Degree of Damage, D of 80m, 120m and 160m respectively. The nodes are attacked by dec-bounded and dec-only attacks. Several experiments by varying certain parameters such as

network density (*m*), Degree of Damage (D), percentage of compromised nodes (x) are carried out and ROC curves are plotted for different cases.

5.2 ROC Curves for Different Metrics

Figure 3 shows ROC curve for three metrics for dynamic deployment of nodes on a non-flat terrain. These ROC curves are for different detection metrics and different degrees of damage. The percentage of compromised nodes is set to 10% and the number of nodes deployed at each deployment point is set to 300.

The results show that the LAD method gives better results for attacks with high degree of damage. The Difference metric shows good anomaly detection rates for low false positive rates. Even though, the experiment was carried out on a non-flat terrain, the results are comparable to the results of static deployment on a flat terrain. As the degree of damage increases, the detection rate increases.

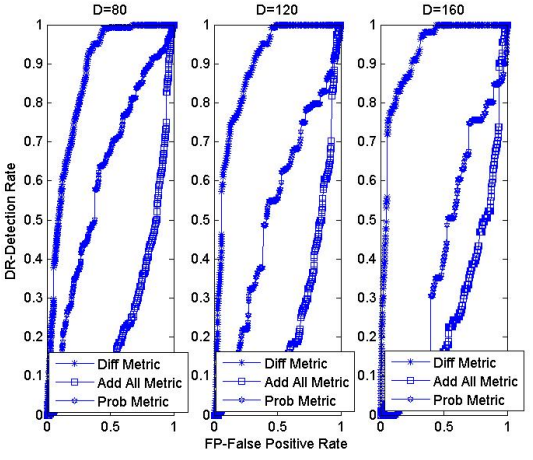

Fig. 3. False Positive Rate vs. Detection Rate for Dynamic Deployment of Nodes on the Non-flat Terrain

5.3 ROC Curves for Different Attacks

Figure 4 shows the ROC curves for the Diff metric for dec-bounded and dec-only attacks for the dynamic deployment. The results shown are for degree of damages D = 40m and 80m.

5.4 Detection Rate vs. Node Compromise Ratio

In this experiment, the ROC curves are plotted for detection rare versus node compromise ratio. Figure 5 shows the ROC curve for dynamic deployment of nodes on the non-flat terrain.

As expected, the detection rate decreases as the percentage of compromised nodes increases. When the number of compromised nodes increases, the localization error increases which results in the anomalies. The detection rates are lower for the dynamic deployment when compared to static deployments as a result of the depletion of nodes from the slopes of the terrain.

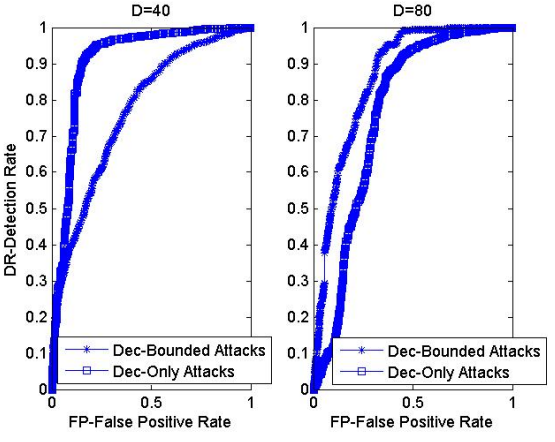

Fig. 4. False Positive Rate vs. Detection Rate for Different Attacks for Dynamic Deployment of Nodes

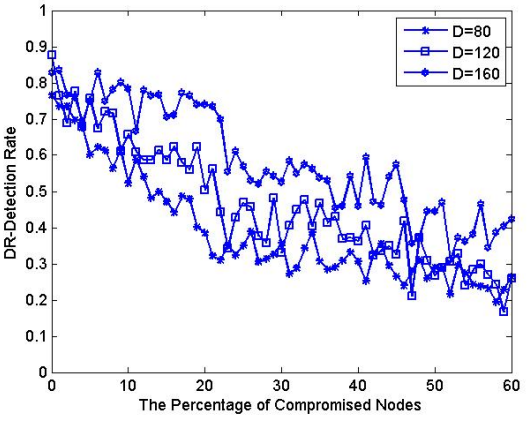

Fig. 5. Detection Rate vs. Node Compromise Ratio (Degree of Damage = 80m, 120m, 160m) for Static Deployment of Nodes

5.5 Detection Rate vs. Network Density

The localization becomes more accurate when the number of nodes deployed at each deployment point in the sensor network increases. In order to demonstrate this, the false positive rate is set to 0.1, and the results show the detection rate for D*iff* metric when the attack is *Dec-bounded.*

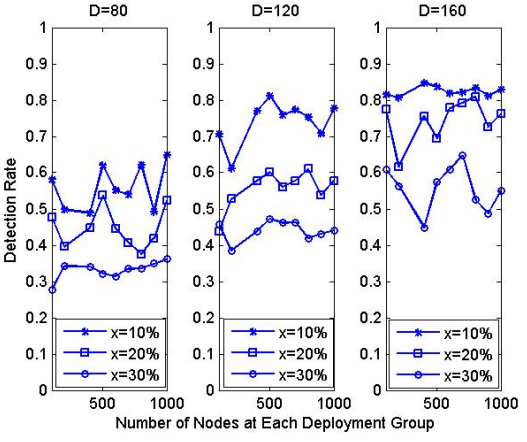

Fig. 6. Detection Rate vs. Network Density (Percentage of Compromised Nodes = 10%, 20%, 30%) for Dynamic Deployment

As the Degree of Damage and the number of nodes deployed at each deployment point increases, the detection rate increases. Figure 6 shows good detection rates when the nodes are deployed dynamically which causes larger localization errors. The false positive rate is set to 0.1 in this case. In all these cases, the trend of the detection rate is observed and it increases as the network density increases.

6 Simulation Results of Signal Strength Metric

This section describes the simulation results of the signal strength metric. Here we compare the results of the signal strength metric with the results of the difference metric. The simulation results, when 30% of the nodes are compromised are compared for the two metrics.

6.1 ROC Curves for Different Metrics

Figure 7 shows the ROC curve for detection rate vs. false positive rate. When compared to the difference metric, the Signal Strength metric shows good detection rates for smaller transmission ranges while being close in performance overall.

The main advantage of the signal strength metric is that it is not susceptible to an attack on the observations in the LAD range. That is, if both the Localization and LAD ranges are attacked the Diff metric would not be able to detect the attack. The entire distribution of nodes against distance needs to be exactly replicated in order to defeat the signal strength metric.

6.2 Detection Rate vs. Degree of Damage

Figure 8 shows the results of detection rate vs. degree of damage. In both the cases, the results of the Signal Strength metric are comparable to the results of the Difference metric. Signal Strength metric performs better for smaller ranges.

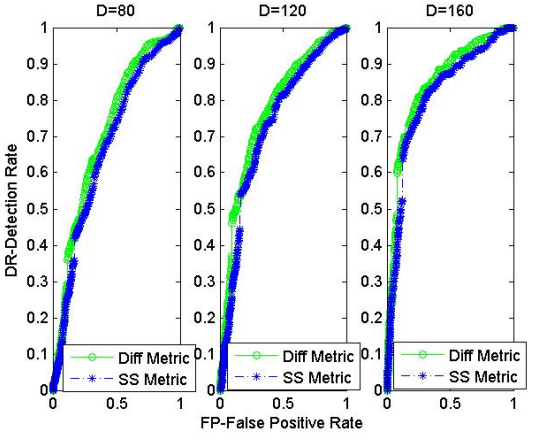

Fig. 7. Detection Rate vs. False Positive Rate for Diff Metric and Signal Strength Metric for Dynamic Deployment of Nodes

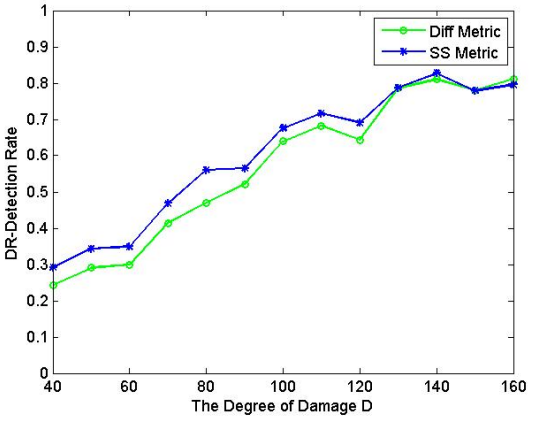

Fig. 8. Detection Rate vs. Degree of Damage for Diff Metric and Signal Strength Metric for Dynamic Deployment of Nodes

7 Conclusion and Future Work

In this research, a LAD method for non-flat terrains is proposed and evaluated. The beacon-less localization scheme proposed for non-flat terrains in Krupadanam and Fu [1], is used to find the location of the sensor nodes for the LAD algorithm. A new metric based on signal strength is proposed for LAD. This metric achieves better detection rates with low false positives for smaller signal ranges while being less susceptible to attack.

Moreover, the LAD method developed in this research demonstrates significant robustness with a sparse localization grid. In future work, the impact of errors in the

deployment information on the performance of the LAD method needs to be quantified and analyzed. Other characteristics of the deployment such as wind effects and, non-uniform deployment grids need to be modeled.

References

1. Krupadanam, S., Fu, H.: Beacon-less Location Detection in Wireless Sensor Networks for Non-flat Terrain. In: International Conference on Future Generation Communication and Networking (FGCN 2007), Jeju Island, Korea, December 6-8 (2007)
2. Du, W., Fang, L., Ning, P.: LAD: Localization Anomaly Detection for Wireless Sensor Networks. In: 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2005) (2005)
3. The sample data for Non-flat terrain us taken, `http://www.spinmass.com/2life/docs/Seacliff_Terrain.txt`
4. Lazos, L., Poovendran, R.: SeRLoc: Secure Range-Independent Localization for Wireless Sensor Networks. In: ACM WiSe, pp. 21–30 (2004)
5. Tang, J., Fan, P., Tang, X.: A RSSI-Based Cooperative Anomaly Detection Scheme for Wireless Sensor Networks. In: International Conference on WiCom 2007, Shanghai, China (2007)