Degree-based Clustering Algorithms for Wireless Ad Hoc Networks Under Attack

C. Tselikis, S. Mitropoulos, N. Komninos, *Member, IEEE* C. Douligeris, *Member, IEEE*

***Abstract*—In this paper we investigate the behavior of degree-based clustering algorithms with respect to their stability and attack-resistance. Our attack scenario tries to bias the clustering head selection procedure by sending faulty degree claims. We propose a randomized variant of the highest degree algorithm which is proved, through experimental results, attack-resistant without imposing significant overhead to the clustering performance. In addition, we extend our proposal with a cooperative consistent clustering algorithm which integrates security into the clustering decision achieving attacker identification and classification.**

***Index Terms*—secure clustering, cooperation, simulation.**

I. Introduction

SElf-organization in hierarchical structures with multi-level clustering is appealing in large scale ad hoc networks, MANET and Wireless Sensor Networks. However, two clustering issues remain challenging in dynamic mobile environments: a) how to minimize the re-clustering overhead in the face of network partitions (link or node outages), and b) how to make the clustering procedure attack-resistant without sacrificing clustering and network performance. Regarding (a) many heuristic solutions can be found in the literature [1] which when sub-network merging or split is detected they select new cluster heads (CH). Regarding(b) in [2] a cluster-based cooperative IDS is proposed in which only the fairly and securely selected CHs perform traffic monitoring and intrusion detection. We address (a) by proposing a cooperative weighted clustering scheme, the Consistent Clustering Algorithm (CCA), and we address (b) in two different ways, namely by proposing a randomized version of the highest degree algorithm (RHD) and by integrating into CCA a cooperative mechanism in which any node can act as a detector that correlates the advertized node claims in order to identify the attackers. We concentrate on the protection of the weighted clustering schemes because their merits are numerous, namely they are application-independent, by weight-optimization can be adaptable to different network conditions (e.g., topology changes due to mobility), they are applicable to both centralized and distributed architectures and allow for simultaneous self-organization and self-protection when extended with security components. One disadvantage is that they can introduce significant communications overhead and processing delay (unless the clustering information is exchanged only locally). Our experimental results show that when the CH selection procedure is protected, additional re-clustering overhead is imposed. Particularly, the proposed CCA selected as CH one of the simulated attackers with the least probability but, on average, the CCA CH change rate was found approximately three times more than that of the HD. On the other hand, the proposed randomized CH selection (RHD) can offer protection in the sparse network case with small processing and re-clustering overhead.

Section II presents our conceptual model and the compared clustering algorithms. Next, in section III we present our assumptions, the integrated simulation model and the experimental results. Section IV draws the conclusions.

C. Tselikis, S. Mitropoulos, N. Komninos and C. Douligeris are with University of Piraeus, Greece, e-mail: (ctseliki@unipi.gr).

Manuscript received June 10, 2011; revised December 30, 2011.

II. CONCEPTUAL MODEL

Our conceptual model is based on the ad hoc self-organization concept, as shown in Figure 1. This model demands global agreement (consensus) to be reached for cluster formation and for intruder identification. Also the ad hoc routes are selected after a mutual exchange of opinions amongst the neighboring nodes.

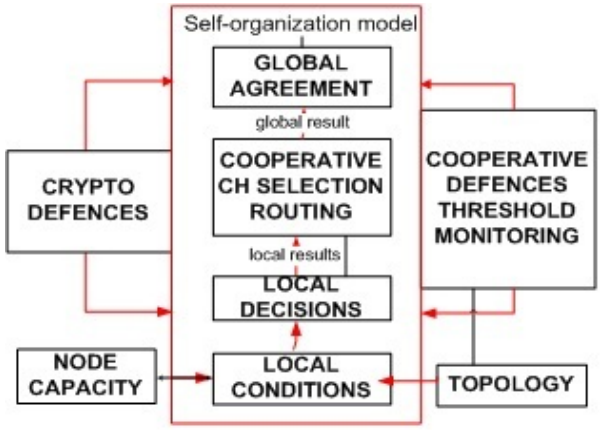

Fig. 1. Conceptual model for self-protection and self-organization

Figure 1 shows two complementary defense blocks, namely the cryptographic (encryption, authentication with digital signatures and key management) and the cooperative block (includes intrusion detection with consensus, reputation/trust, voting and game-based schemes). We concentrate here on the cooperative secure clustering since we want to evaluate the efficiency of such mechanisms (especially consistency thresholds) as substitutes for cryptographic primitives. In that respect, we propose a randomized variant of the highest degree and we extend our proposal with the cooperative CCA. Also, we investigate how the HD and its variant WHD behave under

attack. The CH selection criteria for each algorithm follow. Highest Degree (HD) HD is a well-known from the 90s ad hoc clustering algorithm in which as local CH is selected the node with the maximum connectivity degree, i.e., the node with the maximum number of uncovered in-range neighbors (periodic broadcast messages are used for one-hop neighbor detection).

A. Weighted Highest Degree (WHD)

WHD [4] is a variant of the HD algorithm in which the clustering score V_i for each node i is calculated as the inverse of the sum of the degrees of his j neighbors, Equation (1). WHD gives high priority to low-degree nodes with many neighbors aiming to reduce the number of clusters.

$$V_i = \frac{1}{\sum_{j=1}^{N} deg_j} \quad (1)$$

B. Randomized Highest Degree (RHD)

RHD is part of our proposal in which the $top-k$ neighbors, i.e., the nodes having the k largest advertised degrees are found and the new local CH is drawn randomly (by the old CH) amongst the $top-k$ neighbors. In our tests with RHD we used the uniform random number generator however we recommend the use of parameterized pseudo-random generators for increased guarantees of security.

C. Consistent Clustering Algorithm (CCA)

CCA extends our proposal. CCA for each node i takes into account its degree degi (the number of nodes whose Euclidean distance from i is less than the radio range of i), an energy-related fairness factor Fi (how many times i has previously served as CH), a security-related component and the nodes Euclidean distance L_i from the clusters maximum range (nodes located at the neighborhoods center are more preferable). Equation (2) presents the normalized clustering variable V_i of CCA:

$$V_i = a\times\frac{deg_i}{d_{max}} + b\times\frac{F_i}{F_{max}} + c_t\times\left(\frac{N_f}{deg_i} - \frac{2}{3}\right) + d\times\frac{L_i}{L_{max}} \quad (2)$$

where the coefficients a, b, c_t and d satisfy the following:

$$a + b + c_t + d = 1. \quad (3)$$

The third component in Equation (2) protects the CH selection from nodes that advertise faulty degrees in order to gain the CH role and hence control the network. CCA classifies each node as normal, suspect or attacker and allocates a different value of c_t for each type of node according to Equation (4). According to CCA, the maximum acceptable advertised degree degi equals to the network size which is assumed known. If this threshold is exceeded, the monitored i is marked as attacker and it is immediately excluded from both the clustering and the routing procedures (red alarm raised).

$$c_t = \begin{cases} > 0 & \text{if } \frac{N_f}{deg_i} > \frac{2}{3}, 0.4 \text{ for dense, } 0.2 \text{ for sparse,} \\ < 0 & \text{if } \frac{N_f}{deg_i} <= \frac{2}{3}, \\ = 0 & \text{if } deg_i > \text{network size, } (a,\ b,\ d = 0). \end{cases} \quad (4)$$

TABLE I
CONSISTENT CLUSTERING ALGORITHM (CCA)

For each ad hoc node
Phase I: node set-up
  place the node in the field according to the topology model;
  initiate the node state and the clustering / network elements;
Phase II: CH selection
 start moving the nodes randomly;
 build nodes Neighbour-List (NL);
  if (NL.size ¿ max-cluster-size) then
   truncate NL to max-cluster-size;
  sort NL on the V_i (the neighbors scores)
  for each node i in NL
    if (N_f / degi 2/3) then
      neighbor is suspect, calculate V_i with c ¡ 0; from Eq. (2)
  else if (N_f / deg_i ¿ 2/3) then
    node is normal, calculate V_i with c ¿ 0; from Eq. (2)
  else if (deg_i ¿ network-size) then
    node is attacker, exclude node from decision;
CH = neighbor with maximum V_i;
Phase III: re-clustering
if new selected CH was simple member before then
   Increase the number of CH changes;
   Update the CH states; Update the members state; Update the CH-Table;

Further, CCA detects those nodes that send unreasonably high claims by evaluating the ratio of the number Nf of the neighbors found to contain i in their Neighbor Lists over the degree deg_i advertized by i ($log_2 n$ binary search processing delay). When the ratio of the search result (N_f) over deg_i is less or equal than the second threshold (set to 2/3 according to the byzantine agreement requirement [4]) node i is classified as suspect (yellow alarm). According to CCA, a suspect is not immediately excluded but he is penalized by reducing his V_i. CCA consists of three phases, namely the set-up phase, the CH selection and the re-clustering phase. Table I presents the pseudo-code of the proposed CCA.

In Phase I the network state is initialized and during Phase II the nodes select the CHs. If the previous state of a new selected CH was simple member, re-clustering is performed (Phase III) i.e., the CH changes are increased, and the node states and the clustering tables are updated. In addition, the neighbors have to associate with the new announced CH by sending him a join message and the new CH has to acknowledge each one of them. The three-phase clustering structure is also followed by the compared RHD, HD and WHD however in the respective implementations each algorithm makes decisions according to its own CH selection criteria (as described previously).

III. EXPERIMENTAL RESULTS

We simulated nodes moving randomly according to the Random Waypoint model and broadcasting their degree (true or not), their NL (true or not) and their coordinates (only true, known via GPS or other localization means). Any node can be a CH (peers). The clustering procedure yields two-hop clusters and two types of nodes: a) simple nodes (e.g., tiny sensors) which perform nothing more than default routing to their CH, and b) CHs which aggregate, filter, secure and route the received messages to the final destination via the other CHs. No two selected CHs must be in range. Every node is covered

by a CH. There is a maximum on the cluster membership (25 nodes). We assume the clustering of L legitimate nodes (L=95) is threatened by two types of N in total attackers (N=5): a) by class A attackers who advertise a degree which is larger than the network size, and b) by class B attackers who advertise degrees lower than the network size but inconsistently high. We generated two random models of the node degree d in order to evaluate the impact of the initial topology on the clustering performance. We used a) the uniform distribution (U) to simulate sparse scenarios in which the nodes with sufficient energy are weighted more (e.g., home ad hoc applications), and b) the heavy tail (HT) Pareto distribution (P) to simulate groupe-dense scenarios, e.g., military ad hoc applications in which the nodes with higher connectivity are more important. For the Pareto model we set the coefficients (a, b, c, d) of Eq. (2) to (*0.4*, *0.2*, *0.4*, *0*) so that the connectivity and security are weighted more during the CH selection phase.

Fig. 2 shows the CH change rate with respect to the ad hoc radio range (each point is the average of 50 runs). Low radio range values correspond to a sparse network while high radio ranges to a dense.

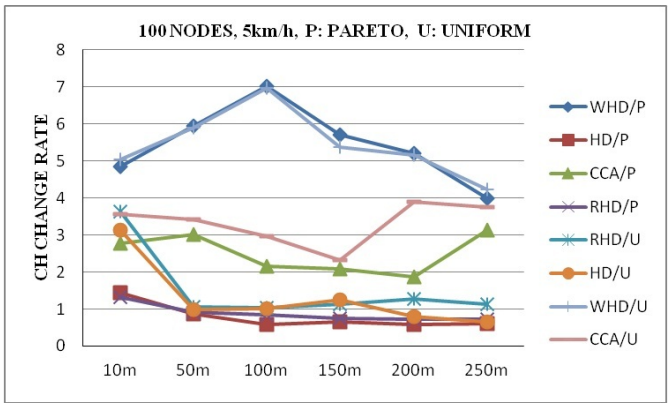

Fig. 2. The CH change rates per ad hoc radio range.

Fig. 2 shows that each algorithm is more stable in the P than the U placement case. HD/P achieved the most stable clustering followed by RHD/P (by 11.91 increase in the rate). The CCA/P performance lies between RHD/P and WHD/P (219.6 overall increase of the HD/P rate). The curves exhibit fluctuations due to our setting of clustering with restricted membership. Under the same conditions each point in Fig. 3 shows the average probability to select an attacker as CH (including stdev which increases with the radio range). CCA/P achieved the best performance, especially when the network is highly connected (range between 200-250 meters). CCA/P was by *58.04* more attack resistant than HD/P. Fig. 4 shows the average number of created clusters. HD/P achieved the least number of clusters followed by CCA/P, RHD/P and WHD/P.

IV. Discussion

All four degree-based algorithms were found more stable for placements with HT characteristics. The RHD achieved encouraging results. The CCA achieved to identify the faulty claims and hence can avoid the impact of CH compromisation

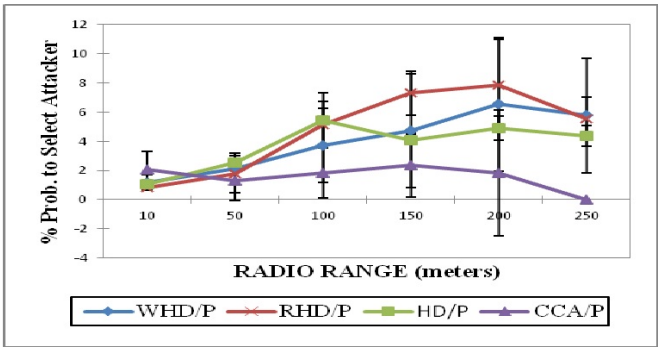

Fig. 3. The average probability of selecting an attacker as CH.

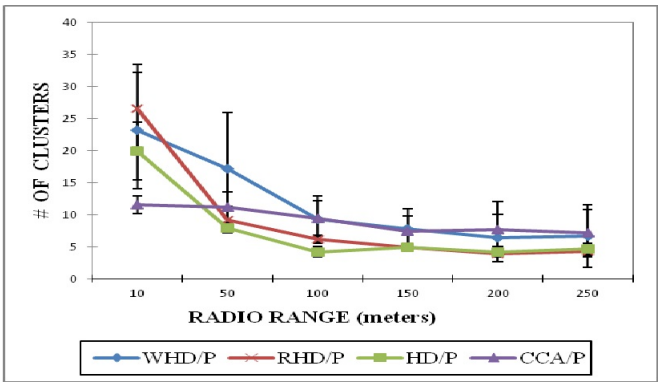

Fig. 4. The average number of created clusters.

(such as packet loss). However, CCA imposed re-clustering overhead. We conclude that the applicability of a specific cooperative mechanism depends on the ad hoc application, the conditions, the resources, and the type and level of threats.

References

[1] J. Y. Yu, P. H. J. Chong, *A survey of clustering schemes for mobile ad hoc networks*, IEEE Communications Surveys and Tutorials (2005) Vol. 7, Issue 1, pp. 32-48.

[2] Y. Huang, W. Lee, *A Cooperative Intrusion Detection System for Ad Hoc Networks*, Proceedings of the 1st ACM Workshop Security of Ad Hoc and Sensor Networks, SASN 03, Fairfax, Virginia.

[3] H. Taniguchi, M. Inoue, T. Masuzawa, and H. Fujiwara, *Clustering Algorithms in Ad Hoc Networks*, Electronics and Communications in Japan, Part 2, Vol. 88, No. 1, 2005, pp. 51-59.

[4] J. S. Baras et al., *Intrusion Detection System Resiliency to Byzantine Attacks: The Case Study of Wormholes in OLSR*, Military Communications Conference, MILCOM 2007, Orlando, FL, USA, pp. 1-7.