

ARIMA Prediction Model-based Cluster Algorithm in Ad Hoc Networks

Yingyu Cao

School of Information Science and Engineering,
Northeastern University,
Shenyang, China

Ting Cao

College of Electronics Information Engineering
Shenyang Aerospace University
Shenyang, China

Huang Ye, Yang Yan, Jiafu Chu

School of Information Science and Engineering
Northeastern University
Shenyang, China

Abstract—The paper introduces prediction mechanism in weighted clustering algorithm (WCA). Time series model (ARIMA) is embedded into the algorithm during routing maintenance to predict the network node location. Combining with location information of nodes provided by GPS systems, the algorithm uses ARIMA to predict position of nodes at next time interval. So, it can calculate aggregate holding time of the nodes. Then, compare the predicted aggregate holding time of next moment with time warning threshold. If cluster structure is predicted unstable, recovery process will be activated before the link fails, and it will search appropriate routing in order to avoid frequent failures of network links. In this way, the influence to the routing protocols brought by the dynamic changes of network topology can be reduced. The simulation results show that, compared with LOWID and RLWCA not joined the forecasting mechanism, the proposed algorithm can dramatically improve the network packet delivery rate, reduce the network normalized expenses and the number of routing interruptions significantly, and improve network performance.

Keywords—Ad Hoc network, ARIMA, Clustering algorithm, CBRP, Prediction

I. INTRODUCTION

Ad Hoc network [1] is a special wireless mobile network combined by mobile communication and computer network. The initial structure is flat type, each node is maintaining the routing information with other node, as the quantity of data sent by network node increases, routing maintenance will bring great burden to nodes. To solve the problem, the best way is to use hierarchical structure, which is constructed by clustering, it means that network node is divided into some clusters, each cluster is usually composed of a cluster head and multiple ordinary nodes, cluster head combined with gateway form higher network. That how to select ordinary node from cluster head will produce significant effect to the stability and reliability of the network.

At present, many researchers have studied on clustering algorithm in Ad Hoc network. Grela and Tsai put forward with minimum ID clustering algorithm (LOWID) [2], where, each node is distributed a unique ID, node with the least ID in adjacent nodes is cluster head, the algorithm convergence is fast. But its weakness is tending to choose the node with smaller ID as a cluster head, and not to consider load balancing and other factors. A dynamic clustering algorithm

based on weight (WDCA) is proposed in reference [3], which produced the dynamic clustering algorithm based on weight, combining thoughts of multi-hop clustering based on the WCA, but when node mobility is strong, it is hard to ensure the stability of the cluster structure. In literature[4], weighted clustering algorithm (MPWCA) based on node moving mechanism is proposed, here, the difference between node degree and ideal degree, residual energy and mobility of nodes and other factors were considered, besides, the calculating method of node mobility is improved, but in the improvement process, only relative speed between the nodes be paid attention to, and the concussion phenomenon is not eliminated, which is caused by cluster structure frequently changes when total weight values of each node is close to equal.

The paper proposes an improved Weighted Clustering Algorithm RLWCA (Restrictive Location-based on-demand Weighted Clustering Algorithm) based on geographic Location, considering deviation of nodes degree, residual energy of nodes and maintain time within the scope of communication, the weights of node choosing cluster head is more rational and the cluster structure is more stable.

II. PREDICTION OF NODE POSITION

ARIMA (Auto Regressive Integrated Moving Average) model is also called difference Moving Average model, which is put forward by the statisticians Box and Jenkins, also called B - J model. The model system is a group of model, including autoregressive model AR(p), moving average model MA(q), smooth average moving model ARMA(p,q), nonstationary summation autoregressive moving average model ARMA(p,d,q). In this paper, we use ARMA(p,d,q). When time series model is non-stationary sequences, d times difference can make it smooth, its expression is as follows:

$$(1-B)^d X_t = \phi_1(1-B)^d X_{t-1} + \phi_2(1-B)^d X_{t-2} + \dots + \phi_p(1-B)^d X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

The above model is:

$$\phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t \quad (2)$$

For time series analysis model having defined, we

adopt minimum mean square deviation analyses forecasting method to analysis it.

Obtained by formula (2),

$$X_t = \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \dots + \Psi_p \varepsilon_{t-p} + \varepsilon_t = \Psi(B) \varepsilon_t$$

here, the value of Ψ_1, Ψ_2, \dots is determined by the following equation:

$$\Phi(B)(1-B)\Psi(B) = \Theta(B) \quad (3)$$

If $\Phi^*(B)$ is defined as generalized autocorrelation function, there is:

$$\Phi^*(B) = \Phi(B)(1-B)^d = 1 - \phi_1 B - \phi_2 B^2 + \dots \quad (4)$$

It is easy to verify that the value of Ψ_1, Ψ_2, \dots meets recursive formula below:

$$\begin{cases} \Psi_1 = \phi_1 - \theta_1 \\ \Psi_2 = \phi_1 \Psi_1 + \phi_2 - \theta_2 \\ \dots \\ \Psi_j = \phi_1 \Psi_{j-1} + \dots + \phi_{p+d} \Psi_{j-p-d} - \theta_j \end{cases} \quad (5)$$

$$\text{Here, } \Psi_j = \begin{cases} 0, j < 1 \\ 1, j = 0 \end{cases}, \theta_j = 0, j > q.$$

Assuming the known condition is time t and the observed value $x_t, x_{t-1}, x_{t-2}, \dots$ before t , we will use the known observed value to predict the observed value $x_{t+l} (l > 0)$ after time t , noted for $\hat{x}_l(l)$, and time series $\{X_t\}$ is the predictive value of the l step.

Then, the true value of X_{t+l} is:

$$X_{t+l} = (\varepsilon_{t+l} + \Psi_1 \varepsilon_{t+l-1} + \dots + \Psi_{l-1} \varepsilon_{t+1}) + (\Psi_l \varepsilon_t + \Psi_{l+1} \varepsilon_{t-1} + \dots) \quad (6)$$

Because of $\varepsilon_{t+l}, \varepsilon_{t+l-1}, \dots, \varepsilon_{t-1}$ being not acquired, so

the only estimate value of X_{t+l} is:

$$\hat{x}_l(l) = \Psi_0^* \varepsilon_t + \Psi_1^* \varepsilon_{t-1} + \Psi_2^* \varepsilon_{t-2} \dots \quad (7)$$

The mean square error between Real value and the forecast value is:

$$E[x_{t+l} - \hat{x}_l(l)]^2 = (1 + \Psi_1^2 + \dots + \Psi_{l-1}^2) \sigma^2 + \sum_{j=0}^{\infty} (\Psi_{l+j} - \Psi_j^*)^2 \sigma^2 \quad (8)$$

To make mean square error least, when and only when:

$$\Psi_j^* = \Psi_{l+j} \quad (8)$$

So, under the principle of least-square error, forecast values of l period is:

$$\hat{x}_l(l) = \Psi_l \varepsilon_t + \Psi_{l+1} \varepsilon_{t-1} + \Psi_{l+2} \varepsilon_{t-2} + \dots \quad (9)$$

forecast error values of l period is:

$$e_l(l) = \varepsilon_{t+l} + \Psi_1 \varepsilon_{t+l-1} + \dots + \Psi_{l-1} \varepsilon_{t+1} \quad (10)$$

The real value is forecasted error values plus forecasted values:

$$X_{t+l} = (\varepsilon_{t+l} + \Psi_1 \varepsilon_{t+l-1} + \dots + \Psi_{l-1} \varepsilon_{t+1}) + (\Psi_l \varepsilon_t + \Psi_{l+1} \varepsilon_{t-1} + \dots)$$

$$= \hat{x}_l(l) + e_l(l) \quad (11)$$

Forecasted square error value of l period is:

$$\text{Var}[e_l(l)] = (1 + \Psi_1^2 + \dots + \Psi_{l-1}^2) \sigma^2 \quad (12)$$

Here, the node's geographical location information of thirty moments is extracted in equal time interval, drawing time sequence graph of the node's horizontal y-coordinate respectively. To predict the minimum mean square error based on the above model can get the next moment geographical location information of the node. Compare actual value to true value, if the absolute variation is small, the effect of the prediction model will be better.

III. THE WEIGHTED CLUSTERING ALGORITHM BASED ON GEOGRAPHIC LOCATION(RLWCA)

In traditional WCA algorithm, there is not enough prediction to node mobility, besides, position, velocity, and moving direction are not considered comprehensively, so, cluster structure is hard to be stable. We draw lessons from the ideas proposed in reference [5], in which an algorithm based on mobile maintaining time is put forward, combining the distance from a node to its neighbors and the average mobility of nodes mentioned in WCA algorithm, making cumulative maintaining time of nodes within the limit of cluster's communication as a main standard used to measure the stability of cluster structure, we put forward an improved weighted clustering algorithm based on geographic position (RLWCA). Its formula of computing combined weight is:

$$\text{Weight} = c_1 \times D_v + c_2 \times P_v - c_3 \times T_v \quad (14)$$

Among them, the specific meanings of parameters are as follows:

(1) D_v Says the difference between node degree (the number of neighbor nodes) and ideal degree.

(2) P_v is the battery energy expended by the node itself, which is the difference between initial energy and the current energy that can be achieved from the energy model.

(3) T_v is the accumulative time of the node's removing beyond the range of cluster head, and T_v is also known as the cumulative keep time.

Calculating processes are as follows:

Each node sends broadcast at power rating before startup time, proclaims its own ID to neighbors and the present speed, then detects the existent broadcast of neighbors, records their current speed (v_{xi}, v_{yi}), counts neighbors' number, and monitors position (x_j, y_j) that neighbor node j relatives to itself taking itself as the center. Assuming that the speed of a node is (v_{x0}, v_{y0}), it may be calculated for the speed that the node relatives to its neighbor node j :

$$(v_x, v_y) = (v_{xj} - v_{x0}, v_{yj} - v_{y0}) \quad (15)$$

Assuming that the power coverage of a node is R , when $v_x \neq 0$, let the slope of the trajectory j is $a = v_y / v_x$, we can calculate the coordinate of intersection $C(x_1, y_1)$ that the node j reaches out of its round intersection,

the formula is as follows:

$$\begin{cases} x_1 = \frac{a(ax_j - y_j) \pm \sqrt{(2a^2 - 1)(ax_j - y_j)^2 + R^2(1 - a^2)}}{1 + a^2}; \\ y_1 = ax_1 - ax_j + y_j; \end{cases} \quad (16)$$

In formula (16), if $v_x > 0$, a larger value will be selected in (16), otherwise a reduction will be selected. When $v_x = 0$, the movement track of node j is a line. At this point, the coordinates of intersection point C is seen in equation (17):

$$\begin{cases} x_1 = x_j; \\ y_1 = \pm \sqrt{R^2 - x_j^2}; \end{cases} \quad (17)$$

In formula (17), when $v_y > 0$, y_1 is the positive item, or else y_1 is the negative item. Then calculate the time used by node J moving from (x_j, y_j) to the intersection of C along the trajectory at the relative velocity, which is the holding time within the scope of that node j is communicating.

$$t_j = \frac{\sqrt{(y_1 - y_j)^2 + (x_1 - x_j)^2}}{|v|} \quad (18)$$

Finally, according to the collected data of neighbor nodes, we can calculate the cumulative time that each neighbor node reaches out of the communication range :

$$T_v = \sum_{j=1}^N t_j z \quad (19)$$

In RLWCA algorithm, the combined weights formula considers mainly the following three factors: The stability of nodes, the nodes' energy state and the deviation of nodes' degree. Among them, the smaller the difference between ideal degree and true degree and energy consumed by nodes is, the more possible node being cluster head is. The longer cumulative time that nodes removed from communication range is, the greater stability of the node is. Therefore, by communicating the formula, we should select the node whose weight is smaller as a cluster head.

IV. WEIGHTED CLUSTERING ALGORITHM BASED ON ARIMA PREDICTION MODEL (ARP-LWCA)

A. THE ESTABLISHMENT OF CLUSTER

The process of building clusters is the process of electing cluster head, the main task of which is to cluster the network. The specific steps are as follows:

1) When system startups, each node named i declares its own ID and current rate (v_{xi}, v_{yi}) to neighbors, then monitors existence radio of neighbors, records their speed, counts number of neighbors, and tests place (x_j, y_j) each neighbor named j relative to the node by taking itself as the center.

2) The node puts its own ID, information of current location and speed direction into its HELLO packet, and broadcasts them to one hop neighbors around. Once

receiving HELLO message package of the node, neighbors will check the list of their own neighbors. If the node already exists in neighbor list, node's information in neighbor list will be update, including node's location, node's deportation and the time that node updates. If the node is not in the neighbor list, the node will be added to neighbor list. After a period of time that the node sent HELLO packets, assuming that it has received all the HELLO message packets of neighbors, that is to say, there is no new neighbors added to neighbor lists within a period of time. At this point, according to the weight of node calculated through equation (14). Taking into account that it is important to take cumulative maintain time of the node as a factor keeping the cluster structure stable in this algorithm, so the weight distribution is: $c_3=0.6$, $c_1=c_2=0.2$.

3) Each node will put combination weights and node ID into the weight packet of the node, then broadcast them to one hop neighbor around. After receiving the weight message of neighbors, each node will update its neighbor list information. Within a period of time, if the node does not receive the weight messages of nodes in the original neighbor list, the node in the original neighbor's list will be removed, or else, the weights of neighbor will be update. If the node information message received is not in the original neighbor's list, it will be discarded.

4) After a certain time that node sent weight message, if weight message in the list of neighbors has been updated, the node will determine whether its own weight is smaller than the weight of all the neighbors. If weights are equal, then see whether the node ID is smaller than that of neighbor's. If it is, the node will send a message indicating itself being a cluster head, or else wait for the message that a neighbor node becomes cluster head.

5) If the general nodes not being able to become cluster head receive news that its neighbor become cluster head, their status will change to cluster members and update the cluster head table, and no longer involve in the process of cluster head election. Otherwise, the node will set itself as a cluster head and send a message of being the cluster head.

6) If there are nodes which are still in initial state, the process of electing cluster head will not stop until all nodes' state become cluster head or cluster member.

B. The prediction to node aggregate holding time and the warning threshold of aggregate holding time

We can obtain node location of t times before by establishing a historical data table. Assuming that location of cluster head node i , previous $n-1$ times and the current time t location of members of node j are known, model parameters can be obtained by way of the least square estimation, then we can calculate the location of node i $(\hat{x}_i(t+1), \hat{y}_i(t+1))$ and the location of node j $(\hat{x}_j(t+1), \hat{y}_j(t+1))$ in $t+1$ moment using the above predicting principle of minimum mean square error.

By (16), we can calculate the predicted value $(\hat{x}_1(t+1), \hat{y}_1(t+1))$ of the intersection C_j , which is the intersection that members node j reach out of the

communication range of the cluster head node i . By (18), we can calculate the time spent by that members node j reached out of the communication range of the cluster head node i , that is the hold time predicted value \hat{t}_j of node i at the moment $t+1$:

$$\hat{t}_j = \frac{\sqrt{(\hat{y}_1(t+1) - \hat{y}_j(t+1))^2 + (\hat{x}_1(t+1) - \hat{x}_j(t+1))^2}}{|v|} \quad (20)$$

$\hat{T} = \hat{t}_j$ is the hold time predicted value.

C. THE ROUTE DISCOVERY MAINTENANCE OF ARP-LWCA

After the clustering structure is formed, when a node needs to communicate with another but not possessing an effective route to the destination, it will start the process of route discovery. Source node initializes RREQ packet first, record node's current location, and broadcast RREQ to its neighbors. If a member receives the packet, the node will discard it. If cluster head receives the packet, it will query routing tables to determine whether there is some reverse route to cluster head of the destination, if yes, reverse route will be established. If not, determine whether it is the first time to receive RREQ, if it is, discard the RREQ, if not, then determine that whether the destination is in the cluster. If yes, determine whether itself is the destination, if it is, respond route response packet RREP, if not, use unicast to the destination and return routing message response RREP. If the destination is not in the cluster, then determine whether there is valid route to the destination, if yes, the return route response message, if not, RREQ will be broadcasted to cluster head of neighbor clusters. Cluster heads receiving the packet doesn't stop transmitting to neighbor cluster heads until the packet reaches the destination belongs to the cluster head node, which unicast the RREQ to destination node. Establishment of actual routing is completed in the process of source node sending RREP to destination node. When RREP is transmitted along reverse cluster head recorded in RREQ, it will add gateway nodes passed to the routing table.

In the route maintenance phase, taking total hold time early warning threshold of nodes as an early warning basis, when we predict that retention time prediction of node $\hat{T} < T'$, it will send warning signals to its upstream nodes and start pre-repair process. When upstream node receives warned signals, it starts route discovery process, according to that the next time the location information forecasted is established a relatively stable the new route, then switch routing, to achieve to select the more stable link to communicate without data transmission interrupting.

V. SIMULATION AND PERFORMANCE ANALYSIS

We use NS2 to simulate routing protocols of LOWID, RLWCA and ARP-RLWCA clustering algorithm on the same conditions. Table 1 shows simulation parameters used in the environment. Each node uses the same wireless transceiver. The wireless transmission radius is 250m, using a single gain omnidirectional antenna with channel capacity 2Mbit / s.

In the simulation, the maximum moving speed of nodes changes in the range of 1m / s to 20m / s, the movement residence time of nodes is 0 seconds, the number of nodes is 40, other parameters are shown in Table 5.1, data obtained are shown in Figure 1. to Figure 3.

As is shown in Figure 5.1, with node's speed changing from 1m/s change to 20m/s, packet delivery rates of routing protocol in three clustering algorithms are decreased, but remained above 85%, and ARP-LWCA delivery rate is higher. When the node moves at the speed of 20m/s, packet delivery rate of RLWCA is higher than LOWID about 2%, which of ARP-LWCA is more than LOWID about 8%, more than RLWCA about 6%. It is because ARP-LWCA is based on RLWCA and increases pre-repair process in the phase of route maintaining, switches to the new routing to communicate before the routing being interrupted, so it ensures data transmission reliable, and reduces the probability of packet loss. Therefore, packet delivery rate has been improved significantly.

As it is shown in Figure 5.2, with the node speed increases, the interrupted number of routing protocol in three clustering algorithms routing is on the rise. The routing interrupted number of RP-LWCA is significantly less than that of RLWCA and LOWID. This is because ARP-LWCA started the pre-repair process in phase of route maintaining and selected the appropriate route in advance to avoid the frequent link breakage. Especially in the case that the node's moving speed is larger, because the protocols do real-time forecast, they are able to quickly adapt to the changes in network topology, reducing the number of routing disruption.

As is shown above, with the maximum node speed increases, the controlled overhead of routing protocols in three clustering algorithm is on rise. Network overhead of ARP-LWCA is the largest, it is because that it adds a pre-repair process in ARP-LWCA and warning control packets, and new route discovering mechanism is enable before routing is interrupted, and thus switches of routing increase, which results in the increase of routing controlled overhead, but the increase of routing controlled overhead in cost is not great, it is acceptable for the application.

VI. CONCLUSION

Weighted clustering algorithm ARP-LWCA based on ARIMA prediction model proposed in the paper, according to node's location provided by GPS system, is able to predict node's location of next time using the established ARIMA model in real time, thus calculate total hold time predicted value of nodes. During route maintained, using ARIMA model to predict aggregate holding time of cluster head, compare time warning threshold value with predicted node's aggregate holding time at next moment, if it predicts the link is unstable, pre-repair process will be enabled in advance before link is broken and reconstruction of the cluster will be completed in order to avoid frequent link-broken in the network, reduce the impact of dynamic changes of network topology on the routing protocols, and maintain the stability of cluster structure. Simulation results

show that, ARP-LWCA has much significant improved than RLWCA and LOWID in choosing stable paths, the number of route broken reduced significantly, making the packet delivery rate is much better than the algorithm RLWCA and LOWID.

Especially when the speed of nodes moving is faster, the performance advantages of ARP-LWCA above become more pronounced. Thus it is proved that the clustering algorithm ARP-LWCA added prediction mechanism is feasible and effective.

REFERENCES

- [1] Frodigh M, Johansson P, Larsson P. Wireless ad hoc networking—the art of networking without a network[J]. Ericsson Review, 2000 (4):248-262.
- [2] C.H.R Lin, M Gerla. A distributed architecture for multimedia in dynamic wireless networks[C], IEEE GLOBECOM, 1995, 2: 1468-1472.
- [3] ZHANG Wei-gang1; HE Qing1; CHEN Hao-liang2; XU Jun1. A weighted dynamic clustering algorithm [J], Journal of Jinan University(Natural Science & Medicine Edition), 2009, (2): 35-39.
- [4] GAO LEI, YAN HAN. On-demand Weighted Clustering Algorithm based on Mobility Prediction in MANET [D], Nanjing University of Science and Technology , Master thesis, 2008.
- [5] Huang Weihong, Li Renfa, Peng Xianwu. Clustering Algorithms for Wireless Self- Organized Networks Based on Mobile Holding Time[J], Networks and Communication, 2007, 23(3): 95-97.

TABLE 1. THE SIMULATION ENVIRONMENT PARAMETERS

Parameter	Value	Parameter	Value
Routing protocol	CBRP	Propagation model	TwoRayGround
Channel	WirelessChannel	network interface	Phy/WirelessPhy
Moving model	RWP	Omni-directional antenna	OmniAntenna
Interface queues	PriQueue	Business types	CBR
Moving scene	800m×800m	Queue length	50
Simulation time	600s	Packet-sent rate	1packet/s

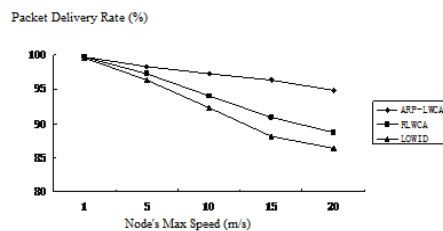


Figure 1. Packet delivery rate-maximum speed

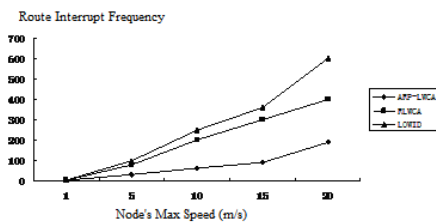


Figure 2. Route interrupt frequency-maximum speed

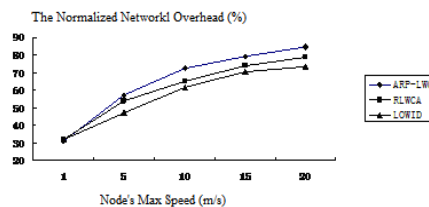


Figure 3. The normalized network overhead-maximum speed