Assessment and Event Based Analysis of Dynamic Wireless Networks

Denis Carvin, Guillaume Kremer, Philippe Owezarski, Pascal Berthou

Assessment and Event Based Analysis of Dynamic Wireless Networks

Denis Carvin[1,2a], Guillaume Kremer[1,2b], Philippe Owezarski[1,2c], Pascal Berthou[1,2b]
[1]CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France
[2]Univ de Toulouse, [a]INSA, [b]UPS, [c]LAAS, F-31400 Toulouse, France
Email: {carvin, kremer, owe, berthou}@laas.fr

***Abstract*—The last decade has seen an increasing interest for wireless communications. With the current use of smart-phones and tablets, together with the rise of the Internet of Things, the number of mobile nodes in networks is significantly changing the way they are managed. Indeed, these wireless networks are highly dynamic, mainly, regarding topology and traffic matrices. The high dynamism of mobile nodes can for instance impact the connectivity of the network, or the role of the nodes in the routing graph. Given this increasing complexity, network's service management requires to be as autonomous as possible. However, autonomy can not be achieved unless the network is able to assess and understand its own behavior. In this paper, we propose an assessment index (SA) based on nodes' satisfaction and its self-estimation algorithm for wireless mobile networks. We then provide events' collection and distributed mining methods allowing nodes to analyze the evolution of this index. We illustrate our framework and characterize the estimation error for various network properties under NS3 simulations.**

I. Introduction

Nowadays, wireless mobile networks have invaded our daily lives. Cisco forecasts [1] expect a 7.5-fold increase of mobile traffic in the next 4 years. If managing wired networks to sustain unpredictable traffic is already complex and many issues remain unsolved in this technical domain, the problems raised by wireless networks represent an extra magnitude of complexity. Indeed, wireless infrastructures experience dynamic topologies depending on nodes' position and links' quality. Furthermore, new constraints and traffic might arise with emerging technologies. Thus, the management task needs to be carried out by the networks themselves. For this purpose, we provide a model that serves as reference for any observer to evaluate a system and then perform an assessment-centric analysis. While this model can be applied to systems in general, we consider dynamic wireless networks of collaborative nodes. More specifically, our work is focused on the evaluation of the network layer in terms of packet loss, end-to-end reachability and delay. Thus, we aim at providing the nodes with the ability of assessing and analyzing their own network. Considering that the nodes have partial information on the network state, they need to collaborate and agree on a time varying assessment value. This is a distributed consensus problem. Regarding analysis, nodes will have to search and share the information they dispose in order to understand what has caused the change in the assessment value and determine in which extend they are responsible for it. We treat this as a distributed mining problem.

The remainder of this paper is structured as follows. In Section II the related work on data-mining applied to network management is described. In Section III our assessment model and its estimation algorithm is introduced. In Section IV, an event based analysis framework is provided and it is illustrated with an example scenario. Finally, in the last section, conclusions are drawn and future work is stated.

II. Related Work

Understanding and managing wireless networks is one of the major operator's concerns. Orange Labs have shown interest on the optimal deployment of wireless substitution networks [2]. On its side, AT&T Labs worked on data-mining to analyze its own wireless infrastructure [3]. Data-mining applied to networking has already been investigated by the security community. The main idea is to reduce false positive alerts in the context of intrusion detection or traffic monitoring and anomaly detection. In [4] the authors mentioned that data-mining was a valuable tool but which could not replace human analysis, specially in the process of choosing attributes. Results can be found in [5], where Casas et al. demonstrated the efficiency of clustering techniques to detect traffic anomaly and construct new filtering rules without knowledge. Moreover, understanding cause and effect between network events is not reserved to security. The authors in [6]–[8] have analyzed the network behavior for a better understanding of it. In particular, in [6] the sources of TCP reset anomalies have been highlighted. The field of wireless communication has been investigated in [7] where the key characteristics of the traffic have been captured on several base stations to optimize their coordination. The authors showed a significant enhancement on the downlink delay performance by clustering users in profiles. Finally, the authors of [8] explained the relation between server's response time, round time trip and users' satisfaction on a set of mobile users. While [5], [7] have brought methods to extract information, [6], [8] have tailored their studies towards a more specific goal. These two approaches need to be connected by a common objective which is the network assessment. The main contribution of the present work is to link these two approaches , we extract information out of the system to obtain the system's assessment and to understand this evaluation.

III. Assessment Model and Distributed Estimation

As we mentioned in the Introduction, our model can be applied to systems in general but we analyze the particular case of dynamic wireless networks. In this section we introduce a method and define a scheme to evaluate these particular systems.

A. Satisfaction Ratio and Assessment Policy

We consider a system as a set of agents. Each agent i has a satisfaction ratio (SR^i) that represents it's wellness over time. Given a set of agents satisfaction ratio, an observer can choose a policy $AP : [0,1]^N \to [0,1]$ to obtain the assessment of an N-agents system (SA). SR^i and SA are time series where each element (respectively SR^i_t and SA_t) is associated to a time interval and takes value in [0,1]. The value 1 represents the best ratio and 0 the worst one. In our model, a policy AP for a system of N agents must satisfy the conditions $AP(\{1\}^N) = 1$ and $AP(\{0\}^N) \neq 1$. In this setting, a system of fully satisfied agents will be considered as ideal. Typically, SAs have been considered to be weighted averages, where the weights are fixed by the observer (simple projection for a selfish agent, equal weight for a fair policy, class based weighting policy...).

B. Assessment of Wireless Mobile Networks

We define the assessment of a wireless mobile network, where nodes are the agents of our model. Their satisfaction function is based on the delay they experience. Each packet that has an end-to-end delay d is scored with the function

$$score(d) = \max\left(0, \frac{d_{max} - d}{d_{max}}\right).$$

Each packet's score decreases linearly as the delay increases between 0 and a given threshold d_{max}. The score equals 1 for a null delay and 0 in case the delay is either greater than the threshold or the packet is lost. The satisfaction of Node i for the t^{th} time interval is given by the average score of the packets that have been generated during this interval (P^i_t), that is,

$$SR^i_t = 1 \text{ if } |P^i_t| = 0, \text{ and } SR^i_t = \overline{score(d^i_t)}, \text{ otherwise,}$$

where d^i_t is the vector of the delays associated to the packets $p \in P^i_t$, *i.e*, $\mathbf{d^i_t} = (\mathbf{d^i_{t1}}, \ldots, \mathbf{d^i_{t|P^i_t|}})$.

The chosen policy in order to assess a network is the average of the satisfaction ratios over the nodes. Thus, the t^{th} assessment of a network of N nodes is given by $SA_t = \overline{SR_t}$.

C. A Distributed Algorithm for Self-Assessment

The SA and SR that were introduced above, capture delay, loss and end-to-end reachability of the network. Moreover, their linearity properties allow a distributed computation.

The estimation of the satisfaction ratio does not require every packet to be scored. Indeed, since the scoring function is linear on the delay interval $[0, d_{max}]$, we can find a d_{max} for which the average score of delays can be approximated by the score of delays' average, that is,

$$\overline{score(d^i_t)} \approx score(\overline{d^i_t}) = \widehat{SR^i_t}.$$

Given that the end-to-end delay is a sum of transmission time, the Node i can compute the estimated delay d^i_t of the packet it has sent or forwarded during the time interval t, by summing its average transmission time $\overline{l^i_t}$ over the packets generated in this interval, and the expected delay from its next hop during the previous time interval. This estimation can be computed distributively with the following iteration:

$$\overline{d^i_t} \approx d'^i_t = \overline{l^i_t} + \sum_{j \in G^i_t} \rho^{ij}_t . d'^j_{t-1}, \text{ with } d'^i_0 = l^i_0.$$

G^i_t is the set of nodes that should have routed traffic for Node i during the t^{th} time interval, while, ρ^{ij}_t is the proportion of this traffic for the Node j. For each packet the transmission time equals 0 if the local node is the destination, it equals d_{max} if the packet is dropped and in other cases it is the time between the first reception (or transmission if the packet is generated by the local node) of the packet and its last successful transmission to the next hop.

The estimation of the system assessment is an average consensus problem. In our case, the average evolves over time, therefore, we modified the scheme presented in [9] for a fixed average. The estimation $\widehat{SA^i_t}$ of SA_t is computed by Node i with the following iteration:

$$\widehat{SA^i_t} = \alpha^{ii}_t . \widehat{SR^i_t} + \sum_{j \in N^i_t} \alpha^{ij}_t . \widehat{SA^j_{t-1}}, \quad \widehat{SA^i_0} = \widehat{SR^i_0}.$$

The value of α^{ij}_t represents the influence of Node j in the neighborhood of Node i for the t^{th} interval of time. This value could be set with the metropolis weighting method described in [10]. In this scheme, the $\widehat{SR^i_t}$ term captures the variability of the satisfaction over time, which was not the case in [9], whereas the summation allows to integrate the propagation of the estimation over the network.

D. Estimation's Results

We conducted NS3 simulations to study the impact of network properties on the quality of our estimation with the Metropolis weights. In these simulations, the width of the time interval is fixed to 1 second. The networks that were considered have ten mobile nodes (when moving, the speed of the nodes is between 5 and 7 m/s). The routing protocol used in the analysis is AODV. Nodes are either source or server, each source has a constant flow (1 Mb/s) towards one of the servers, packet's size is fixed to 1470 bytes, d_{max} equals 10 ms. Each simulation is a combination of the different values for the parameters in table I, while Fig. 1. illustrates the case of 9 sources with two levels of mobility (D $\leq$ 50 and D>50). For all scenarios, we computed the average absolute estimation error at each iteration. For clearness, we did not plot the mid-spreads which were under 0.20 for both curves. The initial error is null since nodes are all fully satisfied, it increases

Network properties		
Random seed	0,1,2	
Number of sources	1,4,7,9	
Initial spacing (D)	20,45,65,75	
Mobility Model	Random Walk	Random Waypoint
Mobility area size	D/2xD/2	DxD (pause duration: 25s)

TABLE I
SIMULATION PARAMETERS AND NETWORK PROPERTIES

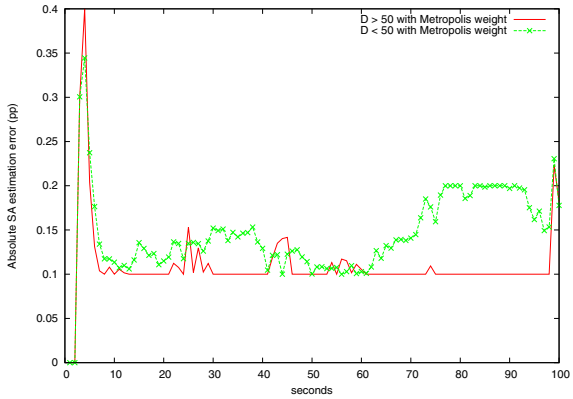


Fig. 1. Average error of the estimated SA depending on network profiles.

brutally when sources are started. However, the error decreases over time. The algorithm performs better under a very dynamic topology. We identified two reasons: (1) when the network is too dynamic, the performance of AODV decreases and all nodes tend to be unsatisfied. (2) dynamism increases the number of known neighbors satisfaction ratio. In this section we showed how nodes could estimate their SA. In the next section we give them a method to analyze the evolution of this SA.

IV. EVENT BASED ANALYSIS OF A SYSTEM ASSESSMENTS

We assume that each agent of a system can produce and observe events. Our approach consists in building system's properties from events to understand their impact on the SA.

A. Observation and Event Definitions

An observation, the perception of an event by an agent at a specific time, can be seen as a tri-dimensional point, see Fig. 2. An event is a *perceptible* modification of the system's state, it is an N-dimensional point which can be represented by a frame. The first field in this frame is the event's type (*eType*) which determines the validity and the meaning of the following fields. Table II illustrates the observation space and gives examples of events to be considered in the case of dynamic wireless networks.

B. Features' Construction

We call feature a property of a cluster of observations. Thus, we will create clusters of observations, compute clusters' properties that vary over time and then study the association between these properties and the SA. As a result, we need metrics to cluster observations by similarity.

(a) Example of Observations

Observations						
Time	**Agent**	**Event**				
		Type	**Source**	**Speed**	**Length**	...
float	*int*	*string*	*int*	*float*	*int*	...
1.2	0	'Packet'	1	-	1500	...
1.25	1	'Move'	1	5.6	-	...

(b) Example of Event Types

Type	**Information**
Packet	Packet capture in promiscuous mode
Rtam	Routing table's attribute modification (size...)
Move	Speed vector modification
Ipv4Drop	Packet Drop for a routing reason
PhyRxDrop	Frame dropped during reception
MacTxDataFailed	Data packet transmission failed at mac layer

TABLE II
OBSERVATIONS AND EVENTS IN DYNAMIC WIRELESS NETWORKS

Regarding the distance in time, timestamps difference is the natural way to proceed but it could be meaningful to use difference between hours of the day or between days of the week. Distance between agents, could be geographical, logical (number of hops) or a state comparison. Events similarity can be based on the string distance between their type names, their number of common fields or the values of their fields.

In this paper, each node groups its observations by time interval and event field value. Like nodes did in Fig. 2. with the number of observed events by unit of time, they build time series of features by applying an aggregate function on these groups (such as count, or average over a field). Then, they study the delayed correlations between their time series and the SA over a period to determine the features that might have impacted the SA. When they collaborate, nodes only need to exchange the correlation coefficients of highest magnitudes.

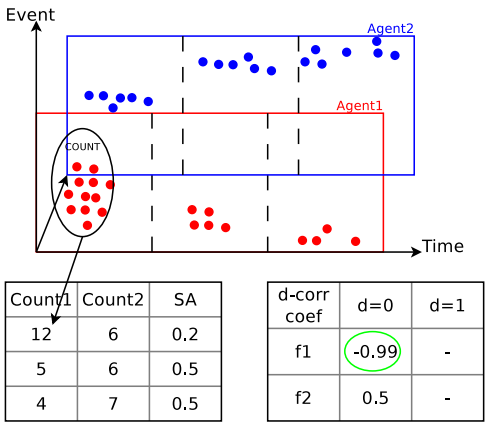


Fig. 2. Features and System Assessment Delayed Correlation.

Name	Information
AvgnbValid	Average # valid entry in the routing table
CountMyRetry	# transmitted frame with a retry flag
CountAllFlow	# IP flow sent, received or forwarded
CountPhyRxDrop	# PhyRxDrop events
CountDropRouteErr	# IPv4 Drop events for a route error reason

TABLE III
FEATURES DESCRIPTION

C. Analysis of a Dynamic Wireless Networks Assessment

For a better understanding, we analyze the scenario given in Fig. 3. and we provide the most relevant features, which were constructed for this study, in Table III. Each feature is related to a node, the # stands for "*number of*".

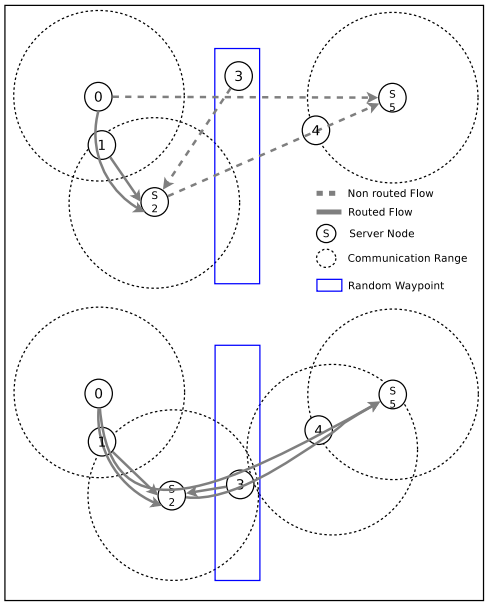
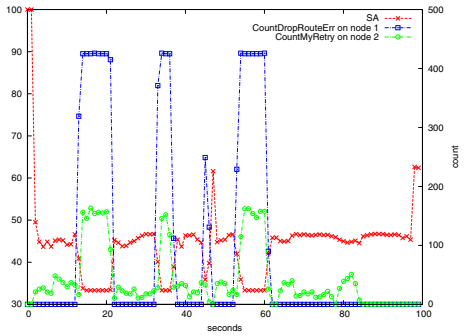
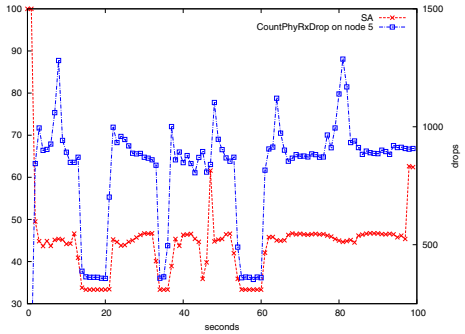


Fig. 3. Topology of the Studied Network.

In the scenario of the Fig. 3., Nodes 2 and 5 are UDP sinks. Node 3 can move in the rectangle area and impact the network topology. On the top, route to 5 is down. At the bottom, Node 2 might be overloaded. After having computed the delayed correlation matrix we found high values for the three features illustrated in Fig. 4. *CountDropRouteErr* on Node 1, *CountPhyRxDrop* on Node 5 and *CountMyRetry* on Node 2 scores are respectively -0.92, 0.79 and -0.88. In Fig. 4(a)., we clearly show that the main fluctuation of the SA is correlated to a routing error. Indeed, Node 1 can not find a route to Node 5 since Node 3 has left the path. The retries experienced by Node 2 are detailed in Fig. 4(a). It impacts the SA when the route is up with a bad communication link between 2 and 3. At first, one can think that transmission retries of Node 2 are introduced by the physical drops on Node 5, but in fact those events are negatively correlated. Indeed, these nodes can not reach each other due to their relative distances, but Node 5 could still be in the carrier range of Node 2. Fig. 4(b). confirms that Node 5 does not drop packets for low SA, indeed, Node 2 does not have to send them because of the routing errors on Node 1.



(a) CountDropRouteErr on Node 1 and CountMyRetry on Node 2



(b) CountPhyRxDrop on Node 5

Fig. 4. Temporal Evolution of SA Regarding 3 Features.

V. CONCLUSION AND FUTURE WORK

The invasion of wireless mobile communication in our network and their increasing complexity forces the management to be mainly delegated to the network itself. In this paper we introduce a way to assess a wireless mobile network and provide the distributed algorithm for nodes to compute this assessment. Our algorithm is derived from existing average consensus schemes. We evaluated this algorithm under various networking conditions to describe its sensitivity to the load and to the nodes' mobility. Then, we proposed a method to analyze the evolution of this assessment. We collected event observations to construct time series of features that we correlate with the network assessment. Using simple features based on event counts, we were able to diagnose the assessment fluctuation, these features can be distributively computed and exchanged in real time by nodes to analyze their situation. Our approach is general enough and could be applied to other multi-agent systems (wired networks, farm of servers or social networks) for their assessment and analysis. For the future we address the following research lines: (1) sensitivity analysis of the estimation algorithm, (2) reduction of the estimation error by refining the metropolis heuristic for the dynamic average consensus algorithm.

ACKNOWLEDGMENT

This work is partially funded by the French National Research Agency (ANR) under the project ANR VERSO RESCUE (ANR-10-VERS-003)

REFERENCES

[1] "Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017," White Paper, Cisco, February 2013.

[2] "Substitution networks based on software defined networking," in *Ad Hoc Networks*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, J. Zheng, N. Mitton, J. Li, and P. Lorenz, Eds., 2013, vol. 111.

[3] The AT&T Labs Research website. [Online]. Available: http://www.research.att.com/projects

[4] E. Bloedorn, A. D. Christiansen, W. Hill, C. Skorupka, L. M. Talbot, and J. Tivel, "Data mining for network intrusion detection: How to get started," Tech. Rep., 2001.

[5] P. Casas, J. Mazel, and P. Owezarski, "Unsupervised network intrusion detection systems: Detecting the unknown without knowledge," *Comput. Commun.*, vol. 35, no. 7, pp. 772–783, Apr. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.comcom.2012.01.016

[6] M. Arlitt and C. Williamson, "An analysis of tcp reset behaviour on the internet," *ACM SIGCOMM Computer Communication Review*, vol. 35, pp. 37–44, 2005.

[7] B. Rengarajan, "Data mining and coordination to avoid interference in wireless networks."

[8] S. Ickin, K. Wac, M. Fiedler, L. Janowski, J.-H. Hong, and A. Dey, "Factors influencing quality of experience of commonly used mobile applications," *Communications Magazine, IEEE*, vol. 50, no. 4, pp. 48–56, 2012.

[9] K. Topley and V. Krishnamurthy, "Average-consensus in a deterministic framework ;part i: Strong connectivity," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6590–6603, 2012.

[10] L. Xiao, S. Boyd, and S. jean Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, pp. 33–46, 2005.