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*—This last decade has seen an increasing interest for wireless communications. With the current use of smart-phones and tablets coupled to the rise of the Internet of Things, the number of mobile nodes in networks will significantly change the way we manage them. Indeed, these wireless networks are highly dynamic, especially concerning topology and traffic matrices. The fast moves of mobile nodes can for instance impact the connexity of the network, or the importance of one of the nodes in the routing graph. Given this increasing complexity, network management regarding the provided service will need to be as autonomous as possible. However this can not be done unless the network is able to assess and understand it's own behavior. In this paper, we propose an assessment index (called 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, one can say that wireless mobile networks have definitely invaded our daily lives. When looking at forecasts from Cisco [1], mobile traffic is expected to increase 7.5-fold in the 4 next years. If managing wired networks to sustain unpredictable traffic is still a complex and unsolved technical domain, the problems raised by wireless networks have a level of magnitude of complexity. Indeed, wireless infrastructures will experience dynamic topologies depending on nodes positions and links qualities, furthermore, new constraints and traffic might arise with emerging technologies. Thus, the management task needs to be mainly delegated to networks themselves. For these purposes, we provide a reference model for any observer to assess a system and then conduct an assessment-centric analysis. While this model could be applied to systems in general, we consider dynamic wireless network of collaborative nodes. More specifically, our work is focused on the assessment of the network layer in term of packet loss, end-to-end reachability and delay. Thus, our aim is to give nodes the assessment and analysis capabilities of their own network. Considering that nodes only have partial information on the network state, they need to collaborate and agree on an assessment value which is time varying, this is a distributed consensus problem. Regarding analysis, nodes will have to search and share the information they dispose to understand the reason that have driven the value evolution and determine whether they are responsible for it. This could be seen as a distributed data-mining problem. Therefore, the remainder of this paper is structured as follows. The next section describes the related work on data-mining applied to network management. We then introduce in section III our assessment model and its estimation algorithm. In section IV, we provide an event based analysis framework and illustrate it through an example scenario. We will conclude on future work in a last section.

II. Related Work

Understanding and Managing wireless network is one of the operator concern. 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, in particular by the security community which is quite fond of this angle. 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 was not about making human analysis unnecessary, specifically in the choice of attributes. Results can be found in [5] where Casas & al. demonstrated the efficiency of clustering techniques to detect traffic anomaly and construct new filtering rules without knowledge. Also, understanding cause and effect between network events is not devoted to security. We found in [6]–[8] analysis concerned by the understanding of network behavior. In [6] authors highlighted the sources of TCP reset anomalies. The field of wireless communication is investigated in [7] where the key characteristics of the traffic are captured on several base stations to optimize their coordination. Authors showed a significant enhancement on the downlink delay performance by clustering users in profiles. Finally authors of [8] explained the relation between server response time, round time trip and users satisfactions on a set of mobile users. While [5], [7] have brought methods to extract information, [6], [8] have tailored their studies towards a very specific goals. These two approaches need to be linked by a common objective which is the network assessment. Therefore we insert our work in between, with the motivation to only extract clues on a system relatively to its assessment. Consequently, the next section introduces a way to assess a system and defines an assessment scheme for wireless mobile networks.

III. Assessment Model and Distributed Estimation

A. Assessment Policy and Satisfaction Ratio

As we mention earlier our model can be applied to systems in general. We consider a system as a set of agents that have a satisfaction ratio (SR) representing their wellness over time. A SR takes its value in the real interval $[0,1]$, where 0 is the worst ratio and 1 is the best one. The SR is given by a satisfaction function (SF) depending on the agent.

A system assessment (SA) takes its value in the real interval $[0,1]$, where 0 is the worst assessment and 1 is the best one. The SA is given by an assessment policy (AP) depending on the observer who assess the system. An AP is based on agent satisfactions only and the space of assessment policies for a system of N agents is given by:

$$\{f : [0,1]^N \to [0,1] \mid \quad f(\{1\}^N) = 1 \; and \; f(\{0\}^N) \neq 1\}$$

So that a system of fully satisfied agents is ideal contrary to a system of fully unsatisfied agents. Typical AP are the weighted averages were the weight 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 assess a wireless mobile network as a system of agents where agents are nodes. Our satisfaction function is based on the delay experienced by each node. Each packet p that has an end-to-end delay d is scored with the function :

$$score(d) = max\Big(0, \frac{d_{max} - d}{d_{max}}\Big)$$

The packet score linearly decreases when the delay increases between 0 and a given threshold d_{max}. It equals 1 for a null delay and 0 if the delay is greater than a threshold (or if the packet is lost). The satisfaction of node i for the time interval T is given by the average score of the packets that have been generated by i during T:

$$SF_i(T) = \overline{score(d)}_{d \in D}, \quad SF_i(T) = 1 \quad if \quad |D| = 0$$

With D $= \{delay(p) \mid p.ipSrc = ip(i) \;\cap\; p.time \in T\}$

Our assessment policy is the average of the satisfaction ratios and consider all nodes being of equal importance. Thus, the assessment of a N nodes network for the time interval T is given by:

$$AP(T) = \overline{SF_i(T)}_{i \in [1,N]}$$

C. A Distributed Algorithm for Self-Assessment

The introduced AP and SF have been built to capture delay, loss and end-to-end reachability of the network. Moreover they offer linearity properties so that they can be computed in a distributed way.

Satisfaction ratio estimation 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} where average delay score can approximate the average score of delays:

$$\widehat{SF_i}(T) = score(\bar{d_i}) \approx \overline{score(d_i)}$$

As a result, the average delay can be used to compute the satisfaction ratio. The end-to-end delay is a sum of transmission time (queuing and medium access times). We estimate the average delay that a packet can experience when leaving a node i with the following heuristic:

$$\widehat{d_i}(T) = l_i(T) + \sum_{j \in G_i(T)} \rho_j(T).\widehat{d_j}(T-1)$$

$$with \quad \widehat{d_i}(0) = l_i(0)$$

$l_i(T)$ is the average transmission time for i over T. For each packet, it equals 0 if the local node is the destination, it equals d_{max} if the packet is dropped, in other cases it is the time between the first reception of the packet and its last successful transmission to the next hop. $G_i(T)$ is the set of gateway used by node i during T. Finally, $\rho_j(T)$ is the percentage of data traffic sent/forwarded by node i during T that should be forwarded by node j. Given a packet leaving node i, its expected delay is the sum of the expected transmission time from i and the expected delay from the next hop. The expected delay from the next hop depends on the mac level traffic matrix materialized by ρ.

System assessment estimation 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 and suggest the following iteration:

$$\widehat{SA}_i(T) = \alpha_i.\widehat{SF}_i(T) + \sum_{j \in N_i(t)} \alpha_j.\widehat{SA}_j(t)$$

$$\widehat{SA}_i(0) = \widehat{SF}_i(0)$$

In this scheme, the SF term introduces the variability of the satisfaction over time which was not the case in [9], the SA term permits the estimation propagation over the network. The value of α_k could be chosen from the metropolis weight described in [10].

D. Estimation Results

We conducted Ns3 simulations to study the impact of network properties on the quality of our estimation with the Metropolis weight. The considered networks have ten mobile nodes (when moving, node speed is between 5 and 7 m/s). The routing protocol is AODV. Nodes are either source or server, each source has a constant (1 Mb/s) flow towards one of the servers, packet 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 figure 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 clarity purpose, 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 source	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

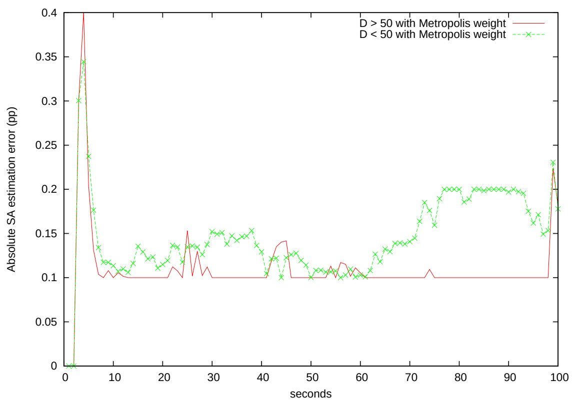


Fig. 1. Average error of the estimated SA depending on network profiles.

brutally when sources are started. However, the error decreases over time. The algorithm better performs under a very dynamic topology. We identified two reasons: (1) when the network is too dynamic AODV performance decreases and all nodes tends 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 analysis approach is to build system properties from events and understand their impact on the SA.

A. Observation and Event Definitions

An observation can be seen as tri-dimensional point, it is the perception of an event by an agent at a specific time. An event is a *perceptible* modification of the system state, it is a N-dimensional point which can be represented by a frame where the first field is the event type (*eType*) which determines the validity and the meaning of the following ones. 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.

Regarding time distance, timestamps difference is the natural way to proceed but it could be meaningful to use difference

(a) Example of Observations Table

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 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 EVENT IN DYNAMIC WIRELESS NETWORKS

between hour of day or day of week. Distance between agents, could be geographical, logical (number of hop) 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 figure 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.

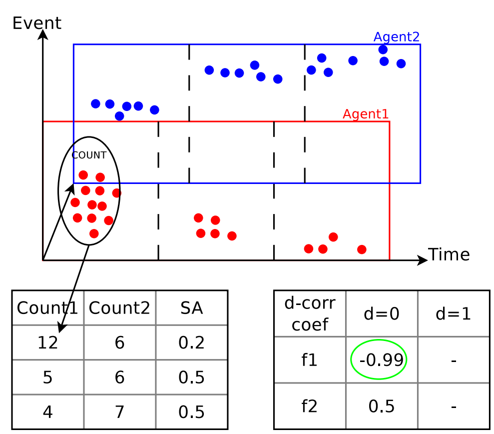


Fig. 2. Features and System Assessment Delayed Correlation

C. Analysis of a Dynamic Wireless Networks Assessment

For understanding purpose, we lead an analysis of the scenario given in figure 3 and give some of the most relevant

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

constructed features for this study in table III. Each feature is related to a node, the # stands for "*number of*".

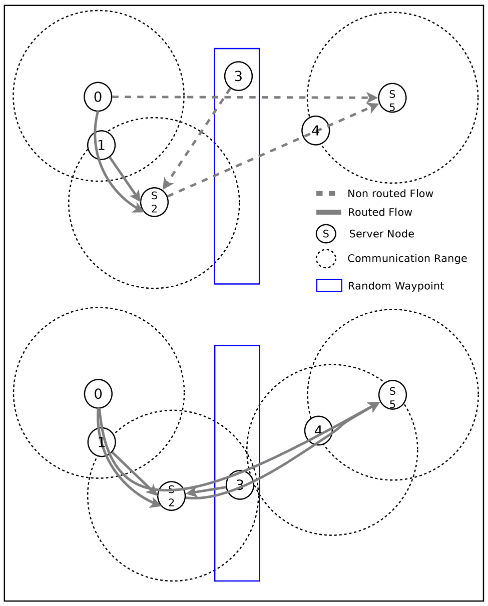
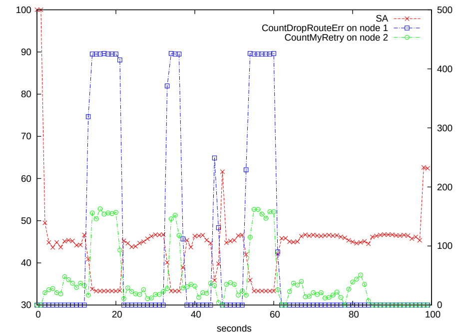
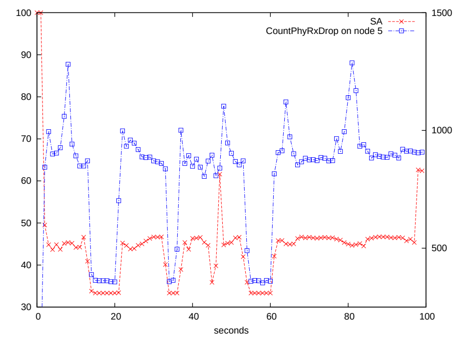


Fig. 3. Topology of the Studied Network

In this scenario, 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 figure 4. *CountDropRouteErr* on node 1, *CountPhyRxDrop* on node 5 and *CountMyRetry* on node 2 scores are respectively -0.92, 0.79, -0.88. In figure 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 figure 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. Since the number of drops is much greater than the number of retries, it might come from the fact that node 5 could still be in the carrier range of node 2. Figure 4(b) confirms that node 5 does not drop packets for low SA, since node 2 does not send them because of 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 complexities forces the management task to be mainly delegated to networks. In doing so, networks need to evaluate themselves and understand the way they behave. 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 load and nodes mobility. Future work could be lead on this sensitivity and on the reduction of the estimation error. 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 diagnosis 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 others multi-agent systems (wired networks, farm of servers or social networks) for their assessment and analysis.

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.