Performance Analysis of Serve on the Move Wireless LANs

Veeraruna Kavitha and M. Venkateswara Rao K.
IEOR, Indian Institute of Technology Bombay, India

***Abstract*—Recently proposals are made to mount wireless trans-receivers on periodically moving vehicles. These vehicles were primarily meant to facilitate human transportation in places like large universities. The idea was to design economical data/voice networks using the already existing infrastructure. The salient feature of these networks is that the wireless server does not stop, rather provides the service to the users while on move and as long as the contact is available. Thus the service of various users waiting on such networks is interlinked. One cannot model this system with existing continuous/discrete polling models, as the later assume the server to stop and serve before resuming with its journey. We obtain the conditions for stability and then the workload analysis. We also discuss optimal scheduling policies.**

I. Introduction

In applications like Ferry Based Wireless LANs (FWLAN), service is offered by a moving server in distributed fashion, whenever it encounters a waiting customer on/near its path. The Ferries are originally meant for carrying human beings (or goods) in a cyclic path repeatedly and the idea was to construct economic communication networks exploiting its repeated journey. Wireless LANs are fitted into such ferries which serve the wireless users in the background. Thus the FWLAN offers the service while on move, where as in existing polling models ([1], [12] and references therein) the server stops to attend the encountered customer. In this paper, we consider the analysis of the former. Thus our system models the FWLANs realistically.

Of all the polling models, studied so far in literature, the one closest to the serve on the move (SoM) FWLAN would be the time limited polling systems. The time limited polling systems are studied under the assumptions of exponential visit times (e.g., [10], [5], [6], [8], [9] etc.,), where as the visit times of the SoM FWLAN hardly exhibit any memoryless property. In fact they result because of traversing more or less the same length of interval in all the cycles and hence the random variations in the visit times can be better modeled using either constant times, or using discrete random times or with some functions of uniform random variables. We consider an autonomous version of time limited system ([7], [11] etc.,), wherein the server visit times are independent of the workload conditions in the queue visited. And this is exactly how a visit to a queue occurs in the FWLAN.

The performance of SoM FWLAN majorly depends upon the service intervals of various queues and their intersections. We showed that the system is stable when the combined load into a set of the intersecting queues is less than the load drained out per cycle, while the server is traversing over the combined service interval. We would also require stability condition of any subset of the intersecting queues. We further obtain approximate workload analysis of such general time limited polling systems, and demonstrate the accuracy of the approximation using simulations.

A. Paper Organization

Stability analysis is obtained in section III while the workload analysis is considered in later sections. In Appendix a g-time limited polling system is described and analyzed.

II. System Model

A Ferry Based Wireless LAN (FWLAN) is moving along a closed tour $\mathcal{C}$, of length $|\mathcal{C}|$, repeatedly with constant speed α and is serving static users that are waiting along/near its path. It can detect the waiting users, i.e., senses the existence of their signal, once the user is sufficiently close. If a favorable scheduling decision is made, it starts serving the encountered user. It does not stop to serve, rather it serves the user while on move. Thus the received signal starts fading away, as the server moves away from the user and service would be stopped once out of visibility zone. Thus for every user there exists an interval on the path, which we refer to as *service interval*, moving along which the server can potentially serve the user. Let L_p represent the *service interval* of point p on the server path, whose length equals $l_p := |L_p|$. The users in the visible vicinity of the path can be projected on to the path.

The server attends the detected user, either when the previous user completes its service or when it moves out of the visibility zone of the previous user. It may also switch over to the new user (the next visible user) either if the signal strength is significantly higher or if the load is higher, depending upon the *scheduling policy.* Thus there is a possibility that the service of a user might be postponed to the next cycle. In all, server attends: a) some customers from previous cycles; b) some from the present cycle.

Arrivals: Every arrival requests for job requirements of size B and arrives at a random position Q. We have N queues, and the probability that an arrival lands in queue $\mathcal{Q}_i$ is given by:

$$Prob(Q \in \mathcal{Q}_i) = q_i \text{ with } \sum_{i=1}^{N} q_i = 1. \qquad (1)$$

The service requests arrive at rate λ per cycle and these arrivals are modeled either by a Poisson process or as Bernoulli arrivals. In either case, the rate of arrival (per cycle) into a queue $\mathcal{Q}_i$ equals $\lambda_i = \lambda q_i$. In case of Bernoulli arrivals, at maximum one customer arrives in the queue in any cycle and this happens with probability λ_i. Alternatively, the number of arrivals per cycle into queue $\mathcal{Q}_i$ can be Poisson distributed with parameter λ_i. If $\mathcal{C}$ is the length of the path and if the FWLAN traverses the route at speed α and $\bar{\lambda}$ is rate of the Poisson arrivals, then rate per cycle is given by $\lambda = \bar{\lambda}|\mathcal{C}|/\alpha$. Most of our results can handle random speeds, however we consider fixed speeds for ease of explanation. For example IID (identically and independently distributed) speed variations across cycles, can be modeled as IID service intervals/visit times.

This represents a very commonly prevalent scenario. For example, the location of points with clustered arrivals can represent a queue. If there is a set of nearby (almost inseparable) points with high arrivals, one can again group them as a queue. Because of 'Serve on the Move' (SoM) nature, the arrivals elsewhere in the path would not alter the performance of the entire system. Either they are served immediately (in the same cycle) or the left overs are completed in the subsequent cycles. This is because, for sparse arrivals the probability of arrivals at interfering points (with intersecting service intervals) is negligible. Thus we concentrate only on clustered arrivals. We discretize the path (parts of the path with clustered arrivals) into finite intervals, model each interval as one queue (see [2] for similar details), which results in arrivals as given by (1).

Service times: Let μ be the service rate per distance per time. If service is offered for l length while the server moves at α speed, then $(l/\alpha)\mu$ amount of job is completed. We assume without loss of generality that $\mu/\alpha = 1$, i.e., the amount of job completed in one cycle equals the length of the service interval. The job requirements B can depend upon the queue into which they arrive. The conditional moments (first and second) of the job requirements, given that the arrival belongs to $\mathcal{Q}_i$, are given by b_i and $b_i^{(2)}$.

Notations: The cycle indices are usually referred by subscripts k or n while the queue indices by subscripts i or j. The quantities that belong to queue i and for cycle k are referred by double subscripts $_{k,i}$ (e.g., $V_{k,i}$ for workload in $\mathcal{Q}_i$ at k-th departure epoch) while the corresponding stationary quantities are referred by single subscript $_i$ (e.g., V_i for stationary workload in $\mathcal{Q}_i$ at departure epoch).

III. Stability analysis

We call *the system is stable, if there exists a scheduling policy which renders all the queues stable.* We begin with introducing the required notations. Let $\mathcal{N}_{k,i}$ represent the number of arrivals into $\mathcal{Q}_i$ in its k-th cycle time (time duration between $(k-1)$-th and k-th departures of the same queue) and let $\xi_{k,i}$ be the corresponding workload:

$$\xi_{k,i} = \sum_{j=1}^{\mathcal{N}_{k,i}} B_{j,i},$$

where $B_{j,i}$ are IID service times of $\mathcal{Q}_i$ with b_i, $b_i^{(2)}$ as first and second moments. If there are no intersecting intervals, then the departure workload $V_{k,i}$ of any queue $\mathcal{Q}_i$ evolves independently of others according to (with $l_i := |L_i|$):

$$V_{k,i} = (V_{k-1,i} + \xi_{k,i} - l_i)^+ \text{ for all } i. \qquad (2)$$

Then the system is stable if and only if $E[\xi_i] < l_i$ for each i. This is a well known stability condition of the random walks on half line (see for example [3]). With Poisson arrivals, by Wald's lemma for any $n \geq 0$

$$P(\mathcal{N}_{k,i} = n) = e^{-\lambda_i} \frac{(\lambda_i)^n}{n!} \text{ and hence } E[\xi_i] = \lambda_i b_i.$$

With Bernoulli arrivals we again have $E[\xi_i] = \lambda_i b_i$, because:

$$P(\mathcal{N}_{k,i} = 1) = \lambda_i = 1 - P(\mathcal{N}_{k,i} = 0).$$

In all, the condition for stability with no intersecting intervals is given by:

$$\max_i \frac{\lambda_i b_i}{l_i} < 1. \tag{3}$$

We now consider two queues, say $\mathcal{Q}_1$ and $\mathcal{Q}_2$, where the service intervals L_1 and L_2 intersect for a length δ as in Figure 1. Recall that the system is stable, if there exists a scheduling policy which renders all the queues stable and the stability is ensured under the following conditions:

Theorem 1. *The two queues are stable if and only if $E[\xi_{k,1} + \xi_{k,2}] < l_1 + l_2 - \delta$ in addition to (3).*

Proof: If $E[\xi_i] > l_i$ for any $i = 1, 2$, clearly the queue is unstable ([3]) and hence the system is unstable. The two queues share the server capacity, only while the server is traversing δ length. Consider a combined system in which the server can serve any customer of the two queues, while it is traversing anywhere in the interval of the length $l_1 + l_2 - \delta$. The total workload process in the combined system, is clearly lower than the sum workload $V_{k,1} + V_{k,2}$ of the FWLAN[1]. And the combined system is unstable if $E[\xi_{k,1} + \xi_{k,2}] > l_1 + l_2 - \delta$, and hence so is our FWLAN.

When $c_e := E[\xi_{k,1} + \xi_{k,2}] - (l_1 + l_2 - \delta) < 0$, we consider a scheduling policy in which $\mathcal{Q}_1$ is served independently of $\mathcal{Q}_2$, after dividing the shared service interval as below:

$$\begin{aligned} l_i^\gamma &= E[\xi_i] + \gamma\delta \text{ for } i = 1, 2 \text{ where} \\ \gamma &:= \min\left\{\frac{-c_e}{2\delta}, \frac{l_1 - E[\xi_1]}{2\delta}, \frac{l_2 - E[\xi_2]}{2\delta}\right\}. \end{aligned}$$

The server attends $\mathcal{Q}_1$ while it is traversing interval of length l_1^γ starting from first end of L_1 and $\mathcal{Q}_2$ while traversing an interval of length l_2^γ whose end coincides with the end of L_2. Because of their definitions and by the given hypothesis, the serving intervals of the two queues do not intersect with each other and thus the two queues evolve independently. Further, both of them are stable again by the well known results of random walk on half line ([3]). ■

Thus the stability of two intersecting queues is guaranteed, once the rate of the combined arrivals into the two queues is less than the length of the joint service interval. This is true irrespective of the length (δ) of intersection. Of course it also demands for individual stability conditions. One can easily extend the above result to the following:

Theorem 2. *Consider M intersecting queues with δ_i, for any $i < M$, representing the length of intersection between the service intervals L_i and L_{i+1}. The M queues are jointly stable if and only if the following are satisfied along with (3):*

i) $E[\xi_i + \xi_{i+1}] < l_i + l_{i+1} - \delta_i$ for any $i < M$ and
ii) $E[\sum_{j=0}^2 \xi_{i+j}] < \sum_{j=0}^2 l_{i+j} - \sum_{j=0}^1 \delta_{i+j}$ for any $i < M - 1$ and so on till
iii) $E\left[\sum_{i \le M} \xi_i\right] < \sum_{j=0}^{M-1} l_{1+j} - \delta$ with $\delta := \sum_{i<M} \delta_i$. ■

Henceforth, the stability conditions are assumed to be satisfied.

IV. Queues without intersecting intervals

The workload analysis of any queue is independent of the others. For every i, the server drains out maximum possible workload of $\mathcal{Q}_i$ while traversing its service interval, L_i. The departure workload evolution is given by equation (2).

FWLAN with N non intersecting queues can be modeled by N-queue g-time limited polling system of Appendix. The walking/switching times of the polling system are given by the physical time taken by the server to move from one queue to other, while the cycle times of all the queues are deterministic and equal $|\mathcal{C}|/\alpha$. The equation (2) resembles the evolution of waiting times in a fictitious GI/G/1 queue, wherein the inter-arrival times are deterministic, equal l_i and whose IID service times are given by $\{\xi_{k,i}\}_k$. This observation is the key element in the analysis of the g-time limited polling systems presented in Appendix. The stationary analysis of workload at departures (2) can be obtained once the first two moments of the idle period of the corresponding fictitious GI/G/1 queue is computed (see Appendix for more details).

An idle period (in fictitious queue) by definition is the time interval between the exit of the last customer and the next arrival, given there is a positive time difference between the two. In other words, an idle period is a fraction of those interval arrival periods, in which the exit of the last customer occurs before the next arrival. These result in two conditions:
a) when the newly arrived workload is finished before the next arrival (i.e., if $l_i - \xi_{k,i} > 0$) and when previous epoch workload, $V_{k-1,i} = 0$;
b) or because the newly arrived workload as well as the left over workload is finished before the next arrival, i.e., if $l_i - V_{k-1,i} - \xi_{k,i} > 0$.
Obviously the possibility of the former event is higher. When the load factor increases (by decreasing l_i) the busy period increases. However the nature of subsequent idle periods, once the queue is empty, remains the same. Hence as load factor increases, the fraction of idle periods resulting out of event (a) increases. Thus, our conjuncture is that the idle period, with sufficient load factor, can be well approximated

[1]One needs to use common cycle time for both the queues in this case.

by the events of the type[2] (a). So we approximate the stationary idle periods of fictitious queue by

$$\mathcal{I}_i = l_i - \xi_i \text{ conditioned that } l_i > \xi_i.$$

Thus with $p_i := Prob(\xi_i > 0) = Prob(\mathcal{N}_i > 0)$,

$$\begin{aligned} E[\mathcal{I}_i] &\approx \frac{l_i(1-p_i) + p_i E[(l_i - \xi_i); \xi_i \le l_i]}{(1-p_i) + p_i P(\xi_i \le l_i)}, \\ E[\mathcal{I}_i^2] &\approx \frac{l_i^2(1-p_i) + p_i E[(l_i - \xi_i)^2; \xi_i \le l_i]}{(1-p_i) + p_i P(\xi_i \le l_i)}. \end{aligned} \tag{4}$$

Substituting the above directly into equation (12) we obtain the stationary average of the departure epoch workload.

The arrivals in any queue of FWLAN occur while the server is traversing a long path of length $|\mathcal{C}|$, while they are served only when the server passes through the corresponding service interval, which is much smaller. Under stability conditions (3), the expected value of the arrived workload in one such cycle is smaller than the workload cleared while traversing the much smaller service intervals. Thus the load factors are usually large and hence idle period moments can be well approximated by (4). Further there is a positive probability $(1-p_i)$ of zero arrivals in any cycle and this probability can also be significant. If the probability of zero arrivals $(1-p_i)$ is significant, one can in fact neglect the second terms in the numerators of (4) and approximate idle period by inter-arrival time. We consider this approximation for the rest of the analysis which results in equation (13) of Appendix and corresponding formula for the stationary workload at departure epoch of $\mathcal{Q}_i$ in the FWLAN equals:

$$E[V_i] \approx \frac{E[\xi_i^2] - E[\xi_i] l_i}{2(l_i - E[\xi_i])} \quad \text{for any} \quad i. \tag{5}$$

The accuracy of this approximation is demonstrated via Monte-Carlo (MC) simulations in Table II.

Simulations*:* We consider a typical queue $\mathcal{Q}_i$ and generate a sufficiently long random sequence $\{\xi_{k,i}\}_{k \le N}$ (with $N > 5000000$). Using this, we generate the sample path of the departure workloads $\{V_{k,i}\}_{k \le N}$ as given by (2), and estimate the stationary average departure epoch workload using the sample mean $\dfrac{1}{N} \sum_{k \le N} V_{k,i}$.

We compare it with two approximate formulae: one obtained using (4) and the other given by (5). We consider both Bernoulli and Poisson arrivals and various terms related to the computations are in Table I. The Poisson arrivals with probability $P(\mathcal{N}_i > 1)$ is close to zero, will have similar terms as that for Bernoulli arrivals and we use the same approximation. We consider uniformly distributed service requirements $\{B_i\}$.

We notice from Table II that both the formulae (columns 3 and 4) approximate well the stationary workloads estimated via MC simulations (column 2). Majority of cases, the approximation error is well within 10%. When the probability of arrival p_i in a cycle is small (.05), the approximation is good even for small values of ρ (until 0.083 as seen from the first 4 rows). With moderate values of p_i we have good approximation for larger ρ: e.g., when (p_i, ρ) equals one of $(0.4, 0.5)$, $(0.6, 0.9)$, $(0.864, 0.95)$ for Poisson arrivals or the pair equals one of $(0.5, 0.5)$, $(0.9, 0.9)$ for Bernoulli arrivals.

In all, the formula (5) is a good approximation for stationary departure workloads (under significant load conditions).

Configuration	Terms related to $\mathcal{Q}_i$
Bernoulli	$E[\xi_i] = p_i b_i,\ E[\xi_i^2] = p_i b_i^{(2)}$
Poisson	$E[\xi_i] = p_i b_i,\ E[\xi_i^2] = \lambda_i b_i^{(2)} + \lambda_i^2 (b_i)^2$
Bernoulli Uniform($[0, b]$)	$E[(l_i - \xi_i); 0 < \xi_i \le l_i] = p_i \frac{l_i^2}{2b}$, for $b > l_i$ and $E[(l_i - \xi_i)^2; 0 < \xi_i \le l_i] = p_i \frac{l_i^3}{3b}$

TABLE I
SOME TERMS RELATED TO THE FORMULAE

l_i, λ_i, P/B	MC	$E[V_i]$=(5)	$E[V_i]$ =(4) in (12)	ρ	p_i
.3, .05, P	16.12	16.54	16.54	.83	.05
.3, .05, B	15.76	15.92	15.92	.83	.05
3, .05, P	.191	.178	.186	.083	.05
3, .05, B	.179	.167	.175	.083	.05
5, .5, B	1.11	.833	1.17	.5	.5
3.5, .5, P	7.18	7.08	7.2	.714	.39
3.5, .5, B	4.16	3.96	4.13	.714	.5
4, .7, P	21.76	21.58	21.81	.875	.5
4, .7, B	9.73	9.33	9.76	.875	.7
5, .9, P	28.25	27.75	28.2	.9	.59
5, .9, B	8.20	7.5	8.65	.9	.9
10.5, 2, P	63.5	61.67	64.35	.95	.864

TABLE II
MONTE CARLO SIMULATIONS WITH $B_i \sim$ UNIFORM ([0,10])

V. QUEUES WITH INTERSECTING SERVICE INTERVALS

We begin with the scenario of Figure 1, that of two queues with intersecting service intervals. The evolution of the two queues depends upon the scheduling policies used. The two queues evolve as below:

$$\begin{aligned} V_{k,1} &= (V_{k-1,1} + \xi_{k,1} - G_{k,1})^+ \\ V_{k,2} &= (V_{k-1,2} + \xi_{k,2} - G_{k,2})^+, \end{aligned} \tag{6}$$

[2]This conjuncture needs explicit proof and we are working towards it.

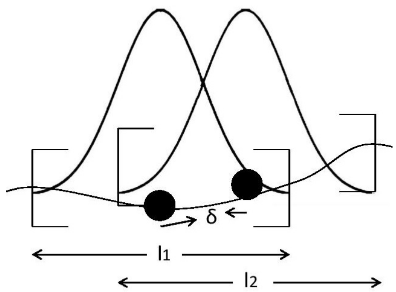

Fig. 1. Two intersecting queues

where the sequences $\{G_{k,1}\}_k$, $\{G_{k,2}\}_k$ are determined by the scheduling policy used. By *scheduling decision, we mean the allocation of the shared interval to the individual queues, either dynamic or static.* Note that $G_{k,1}$, $G_{k,2}$ and $G_{k,1}$+ $G_{k,2}$ will be upper bounded respectively by l_1, l_2 and l_1+ l_2 - δ, at any time k.

In *a naturally used policy,* the server first senses the users of the first queue in the orbit, say $\mathcal{Q}_1$, completes their work requirements to the maximum extent possible (while traversing L_1). It then looks for the users of $\mathcal{Q}_2$. Thus the workload evolution at $\mathcal{Q}_1$ is independent of the second one (as in the case with no intersecting intervals) while that of the second one depends upon the left overs of the first one:

$$\begin{aligned} G_{k,1} &= l_1 \text{ for all } k \text{ and} \\ G_{k,2} &= (l_2-\delta)+\min\left\{\delta,\ (l_1-(V_{k-1,1}+\xi_{k,1}))^+\right\}. \end{aligned} \quad (7)$$

With natural scheduling policy the first queue in the orbit is the favored one, while the second one can suffer, more so if its load requirements are larger, i.e., if $E[\xi_2]/l_2 > E[\xi_1]/l_1$. We thus study two more sets of scheduling policies. We continue with the independent policies of Theorem 2 and obtain optimal shared interval division. That is, we set

$$G_{k,1} = l_1 - \gamma\delta \text{ and } G_{k,2} = l_2 - (1-\gamma)\delta \text{ for all } k, \quad (8)$$

and chose a γ^* that minimizes the sum of the workloads. We compare the optimal policy with the natural policy.

If the server can detect the existence of workload in $\mathcal{Q}_2$, while serving $\mathcal{Q}_1$, one can do better. We call these as *coupled policies,* which attain the better of the two policies discussed above and obtain optimal division of service intervals. The results of Appendix can be used as before, if the scheduling sequences $\{G_{k,1}\}_k$, $\{G_{k,2}\}_k$ are IID. The study of non IID scheduling policies require the results of Markovian queueing models, and would be a topic of future research.

We are interested in choosing the optimal policies that minimize the stationary average workload given by equation (15) of Appendix. And as observed in the Appendix the only controllable part of this workload is the stationary average workloads of the various queues at the departure instances, $E[V_1]$ and $E[V_2]$ and hence work directly with these.

A. Natural Policy (7)

In a natural policy the server attends the users, as and when it senses one and keeps attending them as long as they are in its visibility zone. Hence it first senses the users of the first queue in orbit and attends that of the next one only when the users of first queue are away. Because of preference, the shared service interval is used by the first queue whenever it has load. This kind of a policy could be good as long as the load factor of the first queue is high. On the contrary if the second queue has much higher load factor, it might be better to dedicate the shared interval to the second one. The partial load of the first one, if required, can be postponed to the next cycle to be served utilizing the service interval unavailable to the second queue. This helps drain out the higher load-second queue at faster rate, while the lower load of the first queue is drained out sufficiently faster, just using the private service interval.

We *analyze the policy (7) under an extra assumption that* $\xi_1 \le l_1$ *almost surely.* Under this assumption, the workload in $\mathcal{Q}_1$ is never carried forward to the next cycle, i.e., $V_{k,1} = 0$ almost surely and hence with $\tilde{l}_i := l_i - \delta$ for $i = 1, 2$

$$\begin{aligned} G_{k,2} &= l_2 - \delta + \min\{\delta, (l_1 - \xi_{k,1})^+\} \\ &= l_2 1_{\{\xi_{k,1} \le \tilde{l}_1\}} + (\tilde{l}_2 + (l_1 - \xi_{k,1})) 1_{\{\xi_{k,1} > \tilde{l}_1\}}. \end{aligned}$$

Thus the first scheduling sequence $\{G_{k,1}\}_k$ is a constant sequence while $\{G_{k,2}\}_k$ is an IID sequence and hence the results of Appendix can be applied. It is easy to see that $E[G_1] = l_1$, $E[G_1^2] = l_1^2$ while with $\tilde{l}_i := l_i - \delta$

$$\begin{aligned} E[G_2] &= l_2 P(\xi_1 \le \tilde{l}_1) + E[\tilde{l}_2 + l_1 - \xi_1; \xi_1 > \tilde{l}_1] \\ E[G_2^2] &= l_2^2 P(\xi_1 \le \tilde{l}_1) + E[(\tilde{l}_2 + l_1 - \xi_1)^2; \xi_1 > \tilde{l}_1]. \end{aligned}$$

We have $E[V_1] = 0$ and the stationary workload of the second queue at the departure epoch $E[V_2]$ is obtained by substituting the above into equation (13) of Appendix.

This assumption is satisfied by Bernoulli arrivals (at maximum one arrival) with $B_{k,1} \le l_1$ almost surely for all k. It is also satisfied approximately by Poisson arrivals when the arrival rate λ_1, is small.

B. Independent policies (8)

From (8), we have constant scheduling sequences which satisfy the assumptions of Appendix. For any fraction γ, with $l_1^\gamma := l_1 - \gamma\delta$ and $l_2^\gamma := l_2 - (1-\gamma)\delta)$

$$E[G_i] = l_i^\gamma \text{ and } E[G_i^2] = (l_i^\gamma)^2, \text{ for } i = 1, 2.$$

The stationary workloads $E[V_1]$, $E[V_2]$ can be computed using (13) as before, and exactly resemble equation (5) with l_i replaced by l_i^γ. Let $\bar{V}^I(\gamma) = E^I[V_1] + E^I[V_2]$ represent the total workload and let $\sigma_{\xi_i}^2 := E[\xi_i^2] - (E[\xi_i])^2$. We have the following result.

Lemma 1. *The total departure epochs workload $\bar{V}^I(\gamma)$ is optimized by the division threshold*

$$\begin{aligned}\gamma^* &= \max\{0, \min\{1, \tilde{\gamma}\}\} \text{ where} \\ \tilde{\gamma} &:= \frac{l_1\sigma_{\xi_2} - l_2\sigma_{\xi_1} - E[\xi_1]\sigma_{\xi_2} + E[\xi_2]\sigma_{\xi_1} + \delta\sigma_{\xi_1}}{\delta(\sigma_{\xi_2} + \sigma_{\xi_1})}.\end{aligned}$$

Proof: Let $g(\gamma)$ represent the derivative of $\bar{V}^I(\gamma)$:

$$\begin{aligned}g(\gamma) &= \delta\left(\frac{E[\xi_1^2] - (E[\xi_1])^2}{2(l_1^{\gamma^*} - E[\xi_1])^2} - \frac{E[\xi_2^2] - E[\xi_2])^2}{2(l_2^{\gamma^*} - E[\xi_2])^2}\right) \\ &= \delta\left(\frac{\sigma_{\xi_1}^2}{2(l_1^{\gamma^*} - E[\xi_1])^2} - \frac{\sigma_{\xi_2}^2}{2(l_2^{\gamma^*} - E[\xi_2])^2}\right)\end{aligned}$$

If $\tilde{\gamma} < 0$, $\sigma_{\xi_2}(l_1 - E[\xi_1]) - \sigma_{\xi_2}(l_2\sigma_{\xi_1} - \delta - E[\xi_1]) < 0$, and this implies $g(0) > 0$. The derivative g is an increasing function of γ, when $l_1^\gamma > 0$ and $l_2^\gamma > 0$. These γ render both the queues stable and hence one needs chose optimal among these gamma. Thus $g(0) > 0$ implies $g(\gamma) > 0$ for all γ. Hence $\bar{V}^I(\gamma)$ is increasing with $\gamma \uparrow$ and therefore $\gamma^* = 0$. Similarly if $\tilde{\gamma} > 1$, then $g(1) < 0$ as well as $g(0) < 0$. Thus $\bar{V}^I(\gamma)$ decreases with $\gamma \uparrow$ and hence $\gamma^* = 1$.

When $0 < \tilde{\gamma} < 1$, then $g(0) > 0$ and $g(1) < 0$ and $g(\tilde{\gamma}) = 0$. By differentiating g with respect to γ, one can see that the second derivative of workload $\bar{V}^I(\gamma)$ is positive at $\tilde{\gamma}$. Thus we have the minimizer at the interior $\tilde{\gamma}$. ■

M intersecting queues: Consider M intersecting queues in cascade, as in Theorem 2, with δ_i representing the intersecting length between L_i and L_{i+1} for any $i < M$. For each i, we divide the common service interval δ_i optimally between $\mathcal{Q}_i$ and $\mathcal{Q}_{i+1}$ and consider independent processing as before. Let $\Gamma = [\gamma_1, \cdots, \gamma_{M-1}]$ represent the corresponding fractions. The lengths of the service intervals dedicated for individual queues are:

$$\begin{aligned}l_1^\Gamma &:= l_1 - \gamma_1\delta_1, \qquad l_M^\Gamma = l_M - (1-\gamma_{M-1})\delta_{M-1} \\ l_i^\Gamma &:= l_i - (1-\gamma_{i-1})\delta_{i-1} - \gamma_i\delta_i, \text{ for all } 1 < i < M.\end{aligned}$$

The stationary workload for queue $\mathcal{Q}_i$ is given again by (5) after replacing l_i by l_i^Γ. Computing the derivative and equating it to zero, as done in previous section for case with interior optimizer, we obtain the optimal division ratios as $\Gamma^* = D^{-1}\mathbf{e}$, where the matrix D and the vector $\mathbf{e}$ are defined in the equation (9) given at the top of the next page. The boundary optimal points can be obtained as done in Lemma 1.

$\lambda_1, \lambda_2, l_1 = l_2$	γ^*	$E^N[V]$, $E^I[V]$	ρ_1, ρ_2
0.97, 0.1, 0.12	0	0.41, 0.43,	.97, .56
0.9, 0.1, 0.12	0.116	0.39, 0.41	.93, .57
0.8, 0.16, 0.2	0.217	0.29, 0.32	.86, 52
0.2, 0.23, 0.15	1	1.17 , 0.86	.29, .77
0.24, 0.28, 0.18	1	1.30, 0.87	.34, .78

TABLE III
NATURAL POLICY VERSUS OPTIMAL INDEPENDENT POLICY

C. Comparison of the two policies

We now compare the two policies, natural and optimal independent policy. The sum workloads are computed for the example scenario with $B_2 \sim$ Uniform([0,1]) and $B_1 \equiv l_1$ and the results are tabulated in Table III. We consider Bernoulli arrivals at both the queues. When load factor of the first queue is larger than that of the second, the natural policy performs better (first three rows of Table III). Under the converse conditions, the independent policy performs better (last two rows of Table III). In fact the performance of the natural policy is significantly inferior (33% in the last row), calling in the need for better scheduling policies.

D. Coupled Policies

We have seen that the natural policy can be inefficient, especially when the queues later visited by the server, have higher load factor. This calls for a policy which forces the server to stop serving the customers of the closer queues, even before it moves out of the visibility zone. We found the optimal switching threshold, while using independent policies. It would be advantageous to consider dynamic switching (based on the workload status of the second queue), in case there is a mechanism to sense the workload of the second, a priori. In the following we present few preliminary discussions, while a detailed analysis would be considered in future. We consider the policies, that depend only upon the new workload arrived in the latest cycle:

$$\begin{aligned}G_{k,1} &= l_1 - \delta + \delta 1_{\{\xi_{k,1}>0\}} 1_{\{\xi_{k,2}=0\}} \\ &\quad + \gamma\delta\left(1_{\{\xi_{k,1}=0\}}1_{\{\xi_{k,2}=0\}} + 1_{\{\xi_{k,1}>0\}}1_{\{\xi_{k,2}>0\}}\right) \\ G_{k,2} &= l_2 - \delta + \delta 1_{\{\xi_{k,1}=0\}} 1_{\{\xi_{k,2}>0\}} \\ &\quad + (1-\gamma)\delta\left(1_{\{\xi_{k,1}=0\}}1_{\{\xi_{k,2}=0\}} + 1_{\{\xi_{k,1}>0\}}1_{\{\xi_{k,2}>0\}}\right).\end{aligned}$$

The corresponding moments are given by:

$$\begin{aligned}E[G_1] &= l_1 - \delta + \delta p_1(1-p_2) + \gamma\delta((1-p_1)(1-p_2) + p_1p_2), \\ E[G_2] &= l_2 - \delta + \delta p_2(1-p_1) + \delta(1-\gamma)(1-p_1-p_2+2p_1p_2).\end{aligned}$$

$$D = \begin{bmatrix} -\delta_1(\sigma_{\xi_2}+\sigma_{\xi_1}) & \delta_2\sigma_{\xi_2} & 0 & \cdots & 0 \\ \delta_1\sigma_{\xi_3} & -\delta_2(\sigma_{\xi_2}+\sigma_{\xi_3}) & \delta_3\sigma_{\xi_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \delta_{M-3}\sigma_{\xi_{M-1}} & -\delta_{\xi_{M-2}}(\sigma_{\xi_{M-1}}+\sigma_{\xi_{M-2}}) & \delta_{M-1}\sigma_{\xi_{M-1}} \\ 0 & 0 & \cdots & \delta_{M-2}\sigma_{\xi_M} & -\delta_{M-1}(\sigma_{\xi_M}+\sigma_{\xi_{M-1}}) \end{bmatrix}$$

$$\mathbf{e} = \begin{pmatrix} l_2\sigma_{\xi_1} - l_1\sigma_{\xi_2} - \delta_1\sigma_{\xi_1} + E[\xi_1]\sigma_{\xi_2} - E[\xi_2]\sigma_{\xi_1} \\ l_3\sigma_{\xi_2} - l_2\sigma_{\xi_3} + \delta_1\sigma_{\xi_3} - \delta_2\sigma_{\xi_2} + E[\xi_2]\sigma_{\xi_3} - E[\xi_3]\sigma_{\xi_2} \\ \vdots \\ l_{M-1}\sigma_{\xi_{M-2}} - l_{M-2}\sigma_{\xi_{M-1}} + \delta_{M-3}\sigma_{\xi_{M-1}} - \delta_{M-2}\sigma_{\xi_{M-2}} + E[\xi_{M-2}]\sigma_{\xi_{M-1}} - E[\xi_{M-1}]\sigma_{\xi_{M-2}} \\ l_M\sigma_{\xi_{M-1}} - l_{M-1}\sigma_{\xi_M} + \delta_{M-2}\sigma_{\xi_M} - \delta_{M-1}\sigma_{\xi_{M-1}} + E[\xi_{M-1}]\sigma_{\xi_M} - E[\xi_M]\sigma_{\xi_{M-1}} \end{pmatrix}. \tag{9}$$

The stationary workload is obtained by substituting the above into (13). Computing the derivative and equating it to zero, we obtain the optimal division ratio for two intersecting queues, when the optimizer is an interior point:

$$\begin{aligned} \gamma^* &= \frac{-l_1\sigma_{\xi_2} + l_2\sigma_{\xi_1} + \delta\sigma_{\xi_2} - \delta p_1(\sigma_{\xi_2}+\sigma_{\xi_1})}{(\sigma_{\xi_2}+\sigma_{\xi_1})(1-p_1-p_2+2p_1p_2)} \\ &+ \frac{\delta p_1 p_2(\sigma_{\xi_2}+\sigma_{\xi_1}) + E[\xi_1]\sigma_{\xi_2} - E[\xi_2]\sigma_{\xi_1}}{(\sigma_{\xi_2}+\sigma_{\xi_1})(1-p_1-p_2+2p_1p_2)}. \end{aligned}$$

CONCLUSIONS

Ferry based WLANs offer the service while on move, where as the existing polling models do not consider this feature. We obtain approximate analysis of such general time limited polling systems, and demonstrate that the approximation is good under high load conditions using simulations. We show that the controllable part of the stationary average workload is given by the stationary average workload at departure epochs. The departure epoch stationary workload is obtained using the waiting time analysis of a fictitious GI/G/1 queue. The performance of SoM FWLAN majorly depends upon the service intervals of various queues and their intersections. We obtained the stability conditions, considering the intersecting service intervals. We also discussed some scheduling policies, which allocate/divide the shared service interval among the various intersecting queues. We considered static and few dynamic policies. We also discussed optimal policy in a given family of scheduling policies.

As of now, we considered the policies that make independent decisions across cycles. We basically considered policies that depend upon the new workload arrived in the cycle just before the visit. While a more realistic dynamic policy would exhibit correlations, depends both upon the new workload as well as the workload left over at the previous departure epoch. The study of such policies would be a topic of future interest.

REFERENCES

[1] Bin Tariq, Muhammad Mukarram, Mostafa Ammar, and Ellen Zegura. Message ferry route design for sparse ad hoc networks with mobile nodes., In Proc. of ACM MobiHoc, Florence, Italy, 2225 May 2006

[2] Veeraruna Kavitha, and Eitan Altman. Queueing in space: design of message ferry routes in sensor networks. Proc. ITC09 (2009).

[3] Sean P. Meyn, and Richard L. Tweedie. Markov chains and stochastic stability. Cambridge University Press, 2009.

[4] Blanc, J.P.C. Queueing Models: analytical and numerical methods, Tilburg University, 2004-2009.

[5] Al Hanbali, A., de Haan, R., Boucherie, R.J. and van Ommeren, J.K. Time-limited polling systems with batch arrivals and phase-type service times (2012).

[6] Imed Frigui, Attalhiru-Sule Alfa, Analysis of a time-limited polling system, Computer Communications, Volume 21, 1998.

[7] Ahmad Al Hanbali, Roland Haan, Richard J. Boucherie, and Jan-Kees Ommeren. A Tandem Queueing Model for Delay Analysis in Disconnected Ad Hoc Networks (2008).

[8] Leung, K.K., "Cyclic-service systems with nonpreemptive, time-limited service(1994).

[9] Edmundo de Souza e Silva, H. Richard Gail, and Richard R. Muntz. Polling systems with server timeouts and their application to token passing networks (1995).

[10] Eliazar, I., Yechiali, U. Polling under the randomly timed gated regime. Stoch. Models (1998).

[11] De Haan, R., Boucherie, R. J., & van Ommeren, J. K. (2009). A polling model with an autonomous server. Queueing Systems, 62(3), 279-308.

[12] Veeraruna Kavitha, and Richard Combes. Mixed polling with rerouting and applications. Performance Evaluation (2013): 1001-1027.

APPENDIX: G-TIME LIMITED POLLING SYSTEMS

We consider a time-limited finite polling system, in which the visit time at any queue is pre-decided and is independent of the awaiting workload. So far in the literature, these systems are studied only under exponentially distributed visit times. While we require the time limits/visit times with any arbitrary distribution. Hence we call them as g-time limited polling systems, to indicate that the distribution of the visit times is general.

We have N-queues, the server visits them periodically in the same order, every cycle. In the k-cycle it spends time $G_{k,i}$ at queue $\mathcal{Q}_i$, irrespective of whether or not the later

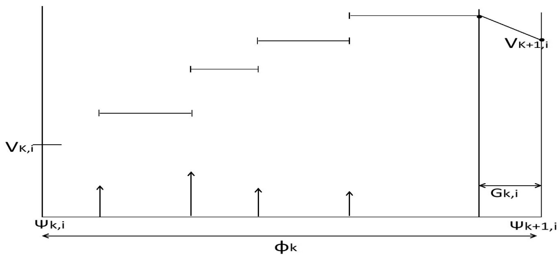

Fig. 2. Workload process in a cycle

is empty. This system is similar to the autonomous system, studied previously in the context of exponential visit times. The service of any interrupted customer is resumed exactly at the point left, when the server next visits its queue. We consider IID visit times $\{G_{k,i}\}_k$, for all the queues. The server spends time $S_{k,i}$ to walk between the queues $\mathcal{Q}_i$ and $\mathcal{Q}_{i+1}$ (modulo addition) during the k-th cycle.

In all the cases we assume that the cycle time is independent of the (fractions of the) times spent in individual queues. For any queue (say $\mathcal{Q}_i$) the sequence of workloads $\{\xi_{i,k}\}_k$, arrived in successive cycle times, is assumed to be an independent sequence and this is true for each i. Here we consider queue specific cycle times, the time period between successive exits by the server, of the same queue.

Each queue is processed independently of the other queues and hence can be analyzed independently. Let $\Phi_{k,i}$ represent the cycle time (time between two consecutive departures) with respect to queue $\mathcal{Q}_i$:

$$\Phi_{k,i} = \sum_{j=1}^{i-1}(G_{k,j} + S_{k,j}) + \sum_{j \geq i}^{N}(G_{k+1,j} + S_{k+1,j}).$$

The cycle times $\{\Phi_{k,i}\}_k$ form IID sequence for each i. The remaining details of the polling system are similar to that of the FWLAN. With $V_{k,i}$ representing the total workload left at $\mathcal{Q}_i$ when the server leaves the queue for the k-th time, the system evolution can be written using the following equation:

$$V_{k,i} = (V_{k-1,i} + \xi_{k,i} - G_{k,i})^+. \tag{10}$$

Analysis of workload using GI/G/1 queue

The waiting time of the k-th customer of a GI/G/1 queue can be written in terms of the service time $\mathcal{B}_k$ and the inter arrival times $\mathcal{A}_k$ as below ([4]):

$$\mathcal{W}_k = (\mathcal{W}_{k-1} + \mathcal{B}_k - \mathcal{A}_k)^+.$$

From [4, equation (6.105)] the average waiting time equals:

$$E[\mathcal{W}] = \frac{E[\mathcal{B}^2] - 2E[\mathcal{B}]E[\mathcal{A}] + E[\mathcal{A}^2]}{2(E[\mathcal{A}] - E[\mathcal{B}])} - \frac{E[\mathcal{I}^2]}{2E[\mathcal{I}]}, \tag{11}$$

where $\mathcal{I}$ is the idle period and the quantities in the above equation are stationary averages.

Equation (10) resembles the evolution of GI/G/1 queue, and hence the stationary average of the departure-workload process $\{V_k\}$ can be obtained using the equation (11). Further for large load factors (as in the case of FWLAN), as discussed before, we approximate the idle times with inter arrival times. Thus the stationary average of the workload at departure epochs of queue $\mathcal{Q}_i$ is given by (see (10)-(11))

$$\begin{aligned} E[V_i] &= \frac{E[\xi_i^2]-2E[G_i]E[\xi_i]+E[{G_i}^2]}{2(E[G_i]-E[\xi_i])} - \frac{E[\mathcal{I}_i^2]}{2E[\mathcal{I}_i]} && (12)\\ &\approx \frac{E[\xi_i^2]-2E[G_i]E[\xi_i]+E[{G_i}^2]}{2(E[G_i]-E[\xi_i])} - \frac{E[G_i^2]}{2E[G_i]} \\ &= \frac{E[\xi_i^2]E[G_i]-2E[\xi_i](E[G_i])^2+E[{G_i}^2]E[\xi_i]}{2E[G_i](E[G_i]-E[\xi_i])}. && (13)\end{aligned}$$

Time average of workload process

Let $V_i(t)$ represent the workload of queue i at time t. We already obtained the stationary analysis of the workload at departure instances, $\{V_{k,i}\}$ as given by (10) , where $V_{k,i} = V_i(\Psi_{k,i})$ with the cycle end instances $\Psi_{k,i} := \sum_{n \leq k} \Phi_{k,i}$. This is a Markov process, in particular a random walk in half line, which is stable and for which the law of large numbers [3, Proposition 17.6.1, pp. 447] is applicable, as the time limits $\{G_{k,i}\}$ are bounded with probability one.

The area under the workload process during k-th cycle majorly depends upon $V_{k,i}$, new arrivals $\xi_{k+1,i}$ and their arrival instances $\{U_n\}_{n=1}^{\mathcal{N}_{k+1,i}}$ as shown in the Figure 2. For Poisson arrivals the arrival instances $\{U_n\}$, conditioned on the cycle length $\Phi_{k+1,i}$, are uniformly distributed over $\Phi_{k+1,i}$, while for Bernoulli arrivals these are assumed to be uniformly distributed.

The computations below can be inaccurate when one or more arrivals occur while the server is at the queue. Further the workload decreases linearly when the server is at the queue, however we neglect this effect too. Given that the cycle times $\Phi_{k,i}$ are large compared to visit times $G_{k,i}$, these inaccuracies lead to negligible errors. Barring these inaccuracies, the required area in k-th cycle equals:

$$\begin{aligned} \bar{V}_{k,i} &:= \int_{\Psi_{k,i}}^{\Psi_{k+1,i}} V_i(t)dt \\ &\approx V_{k,i}\Phi_{k,i} + \sum_{n-1}^{\mathcal{N}_{k+1,i}} B_{n,i}(\Phi_{k,i} - U_n). && (14)\end{aligned}$$

The above area of the workload, is a function of workload at departures (10) and new arriving workload ξ_{k+1} and their arrival instances. The time average of the workload process can be obtained by considering the following limit:

$$\bar{v}_i := \lim_{k\to\infty} \frac{1}{\Psi_{k,i}} \int_0^{\Psi_{k,i}} V_i(t)dt = \lim_{k\to\infty} \frac{\sum_{n=0}^{k-1} \bar{V}_{k,i}}{\Psi_{k,i}}.$$

Using law of large numbers, [3, Proposition 17.6.1, pp 447], and Wald's lemma the above limit is almost surely a constant given by:

$$\begin{aligned} \bar{v}_i &= \frac{E[V_i]E[\Phi_i] + \lambda_i b_i (E[\Phi_i] - E[\Phi_i/2])}{E[\Phi_i]} \\ &= E[V_i] + \frac{\lambda_i b_i}{2} \text{ a.s.} \end{aligned} \tag{15}$$

The above is obtained by computing the stationary average of the terms on the right hand side of (14), which in turn is obtained by conditioning on the cycle lengths $\Phi_{k,i}$ and the number of arrivals $\mathcal{N}_{k,i}$. In case of Poisson arrivals, by PASTA property $\bar{v}_i$ represents the stationary average workload of the system in queue i. We notice from (15) that optimizing the stationary average workload $\bar{v}_i$ is equivalent to optimizing the stationary average workload at departure instances $E[V_i]$ (given by (12)).

The total stationary average workload of the polling system equals:

$$\bar{v} = \sum_i \bar{v}_i = \sum_i \left(E[V_i] + \frac{\lambda_i b_i}{2} \right).$$