Performance Evaluation of XOR-Based Cooperative Relays Using Finite Buffer and Batch Service Queue Models

Pedro M. R. Pereira
Inatel
Santa Rita do Sapucaí, Brazil
pedro.marcio@inatel.br

***Abstract*—This paper investigates the performance of a cooperative wireless relay system utilizing XOR-based network coding under limited buffer constraints. A discrete-event simulation framework is developed to model the queue dynamics, considering both single and dual-packet transmissions enabled by opportunistic XOR operations. To validate and contextualize the simulation results, two classical queueing models—the finite-buffer single-server queue ($M/M/1/J$) and the partial batch-service queue ($M/M^2/1$)—are employed as analytical baselines. Key performance metrics, including average system time, average system occupancy, and blocking probability, are evaluated under varying arrival rates, service rates, and buffer sizes. Results reveal that the XOR relay achieves intermediate performance between the two reference models, offering advantages in both delay and throughput in moderate to high load conditions. The analysis highlights both the benefits and the inherent limitations of XOR-based queueing in practical cooperative relay deployments.**

***Index Terms*—Cooperative communication, XOR network coding, queueing theory, finite buffer, discrete-event simulation, batch service, wireless relay networks.**

I. Introduction

Cooperative communication systems frequently utilize relay stations (RS) to enhance network reliability and throughput by forwarding data between source and destination nodes over orthogonal channels. A key technique employed at the relay is XOR-based network coding, wherein two data packets, denoted A and B, are combined using a bitwise XOR operation to produce a coded packet $(A \oplus B)'$. This coded packet can then be multicast over a single subchannel, effectively improving spectral efficiency [1], [2]. Assuming linear channel coding and modulation schemes, the XOR operation holds after modulation and coding, i.e., $(A \oplus B)' = A' \oplus B'$. Upon reception, each destination node (e.g., mobile station (MS) or base station (BS)) can recover its intended data by performing an XOR operation between the received coded packet and its known packet.

In practical cooperative relay networks, the RS is equipped with a finite buffer queue capable of storing up to N packets, including any packet currently being transmitted. The relay station follows a simple service policy based on the current queue occupancy:

- When only one packet is present in the queue, the RS transmits it immediately without applying network coding.
- When two or more packets are present, the RS performs an XOR operation on the first two queued packets, transmits the resulting coded packet, and removes both original packets from the queue.

Packet arrivals at the relay are modeled as a Poisson process with arrival rate λ (packets per second). In contrast, the service process (packet transmission) follows an exponential distribution with service rate μ (packets per second).

The queueing behavior observed in such XOR-based relay systems exhibits similarities with bulk service queueing models, where customers (packets) are served in batches rather than individually. Bulk service queues have been widely studied in operations research, with applications in manufacturing systems, transportation networks, and elevator operations [3]. In particular, the $M/M^K/1$ partial batch service model has been proposed as an analytical approximation for XOR relay systems [4]. This model assumes Poisson arrivals, exponentially distributed service times, and a single server capable of serving up to K customers per batch. A key characteristic of this model is that if fewer than K customers are available, the server begins service with the current customers, while allowing any new arrivals during the service time to immediately join the ongoing batch and complete service simultaneously with earlier arrivals.

While the $M/M^K/1$ model provides mathematical tractability, it introduces a significant modeling inaccuracy when applied to practical XOR relay systems. Specifically, the assumption that newly arriving packets can join an ongoing service and complete transmission concurrently with previously arrived packets does not align with the operational constraints of XOR network coding. In coding systems, the decision of which packets to code together must be made before transmission initiation, and the service time is fixed for the chosen packet set.

Recent research efforts have sought to address more realistic queueing behaviors in coded relay systems. [5] analyzed a finite-buffer XOR two-way relay system, providing closed-form expressions for throughput and delay that account for lossy wireless channels and signaling overhead. [6] proposed

a threshold-based transmission policy for two-way relay systems, optimizing coding opportunities under phase-type arrival distributions. Additionally, [7] studied an ALOHA-based two-way relay network with XOR coding, deriving both stability and delay performance metrics. Queueing-theoretic techniques have also been applied by [8] to analyze multi-hop mesh networks that implement inter-flow XOR network coding. Moreover, the benefits of buffer-aided relaying for improving throughput and delay have been highlighted in [9], emphasizing the role of opportunistic transmission policies.

A. Contributions

To address the limitations of existing queueing models, this paper makes the following contributions:

- We develop a detailed and realistic system model for a finite-buffer XOR relay station that eliminates the impractical assumption that new arrivals can join ongoing transmissions.
- We conduct a comprehensive performance analysis of the proposed system and compare it with two widely referenced benchmark models: the finite-buffer single-server queue ($M/M/1/J$) and the partial batch service model ($M/M^K/1$).

B. Paper Organization

The remainder of this paper is organized as follows: Section II reviews the theoretical background of the $M/M/1/J$ and partial batch $M/M^K/1$ queue models. Section III presents the detailed simulation model. Performance results and comparisons are discussed in Section IV. Finally, Section V concludes the paper.

II. Theoretical Queue Models

This section provides a brief overview of two classical queueing models commonly used to approximate the behavior of XOR-based relay systems: the finite-buffer single-server queue ($M/M/1/J$) and the partial batch service queue ($M/M^K/1$). Key performance metrics, including average queue length, blocking probability, and average waiting time, are summarized for each model.

Let λ denote the arrival rate (packets per second) and μ denote the service rate (packets per second). The traffic intensity is defined as $\rho = \lambda/\mu$. The total buffer capacity of the system is N packets, including the packet in service. Thus, the system can hold up to $J = N - 1$ waiting packets.

A. $M/M/1/J$ Queue

The $M/M/1/J$ queue represents a finite-capacity single-server system with Poisson arrivals and exponential service times [10]. The system blocks new arrivals when the buffer reaches its maximum capacity N.

The steady-state probability of blocking (also known as the loss probability) is given by

$$P_{\text{b}} = \frac{\rho^N(1-\rho)}{1-\rho^{N+1}}. \tag{1}$$

The expected number of packets in the system (including the one in service) is

$$E[Q] = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}. \tag{2}$$

Applying Little's Law for the effective arrival rate $\lambda_{\text{eff}} = \lambda(1 - P_{\text{b}})$, the average time a packet spends in the system (including service time) is

$$E[t_{\text{q}}] = \frac{E[Q]}{\lambda_{\text{eff}}} = \frac{E[Q]}{(1-P_{\text{b}})\lambda}. \tag{3}$$

B. Partial Batch Service $M/M^K/1$ Queue

The $M/M^K/1$ queue is a Markovian single-server queue with batch service, where the server can serve up to K packets simultaneously [10]. This model assumes Poisson arrivals, exponential batch service times (independent of the batch size), and allows the server to begin service as soon as at least one packet is available. Furthermore, in a partial batch setting, any new arrivals during service are allowed to join the ongoing batch and complete the service together with the earlier arrivals.

The characteristic equation for the partial batch $M/M^K/1$ queue is given by

$$\mu r^{K+1} - (\mu + \lambda)r + \lambda = 0, \tag{4}$$

where r represents the generating function root.

Let $\{r_0, r_1, \ldots, r_K\}$ denote the roots of the characteristic polynomial. Among these, r_0 is the unique root that satisfies $0 < r_0 < 1$.

The expected number of packets in the system (queue length plus the batch in service) is then [10]

$$E[Q] = \frac{r_0}{1-r_0}. \tag{5}$$

Similarly, the average time a packet spends in the system is

$$E[t_{\text{q}}] = \frac{E[Q]}{\lambda} = \frac{r_0}{\lambda(1-r_0)}. \tag{6}$$

C. Discussion

While the $M/M/1/J$ and $M/M^K/1$ models offer useful analytical benchmarks, both introduce simplifying assumptions that limit their applicability to practical XOR relay systems. The $M/M^K/1$ model, in particular, assumes that packets arriving during service can join the ongoing batch and complete service at the same time—a scenario that conflicts with the operational characteristics of XOR network coding, where coding decisions and batch composition must be determined before transmission starts.

III. Simulation Implementation

To complement the theoretical models described in Section II, a discrete-event simulation was developed to evaluate the performance of the cooperative relay system under realistic queueing constraints and XOR-based batch service. The simulation models a single-server queue with finite capacity and adopts a fixed batch size of $K = 2$ packets for coded transmissions, consistent with the partial batch service assumption.

The discrete-event simulation procedure for the cooperative relay system is outlined in Algorithm 1.

The system state at any given time t is represented by the number of packets present in the system, denoted $Q(t)$, which includes both queued packets and those currently in service. The total system capacity is limited to N packets, including the packet(s) in service, such that $Q(t) \in \{0, 1, \ldots, N\}$. The queue length, denoted as $q(t)$, reflects the number of packets waiting for service, excluding those that are already being served. Additionally, the service state is tracked via $s(t)$, representing the number of packets currently undergoing transmission, which can assume values in $\{0, 1\}$ due to the possibility of single-packet or XOR-coded two-packet transmissions.

A. Arrival event

Packet arrivals follow a Poisson process with rate λ, resulting in exponentially distributed inter-arrival times with mean $1/\lambda$. Each arrival triggers an event that updates the system state. If $Q(t) < N$, the incoming packet is admitted to the system. If the server is idle, the arriving packet begins service immediately. If the server is busy, the packet is added to the queue. However, if $Q(t) = N$, indicating that the buffer is full, the packet is blocked, and the blocking counter is incremented. Figure 1 shows the flowchart of the arrival event.

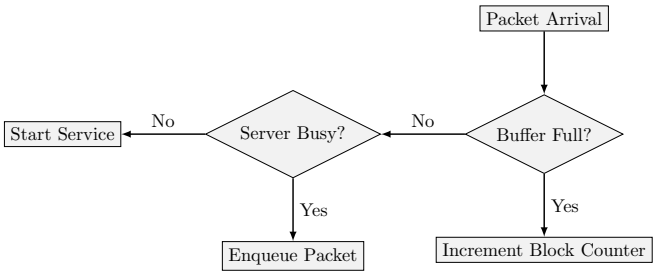

Fig. 1. Flowchart for arrival event handling in the simulation.

B. Departure event

Departures occur according to an exponential service process with rate μ. When a departure event occurs, the simulation determines the next service batch based on the current queue length and the batch size $K = 2$. If the queue is empty, the server becomes idle until the next arrival. If exactly one packet is waiting, the server serves that packet individually. If two or more packets are queued, and XOR coding is enabled, the server forms a coded batch consisting of the two oldest packets, effectively transmitting them as a single coded packet. Notably, each service event, whether handling one or two packets, is governed by an independent exponential service time with mean $1/\mu$, making the service time distribution agnostic to the batch size. Figure 2 shows a flowchart of the departure event.

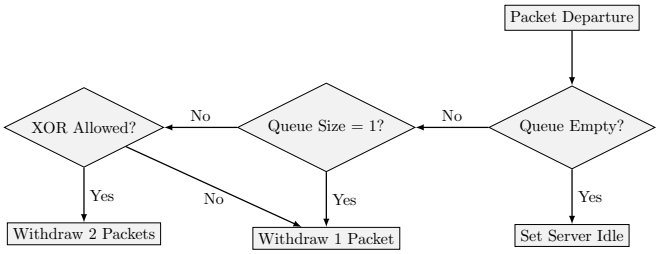

Fig. 2. Flowchart for departure event handling in the simulation

C. Performance estimators

The simulation employs a time-driven event scheduling mechanism, maintaining two dynamic event markers: the next arrival time and the next departure time. At each step, the global simulation clock advances to the time of the earliest scheduled event. The simulation continues until a predefined number of departures, $D_{\max}$, is reached, ensuring statistically stable results.

Performance metrics are obtained from the simulation using time-averaged and event-counting methods. Let A denote the total number of arrival attempts, B the number of blocked arrivals, and D the number of completed departures (transmissions) observed during the simulation period.

The empirical blocking probability, denoted $\hat{P}_{\mathrm{b}}$, is estimated as the ratio of blocked arrivals to total arrival attempts

$$\hat{P}_{\mathrm{b}} = \frac{B}{A}. \tag{7}$$

The average number of packets in the system, denoted $\hat{E}[Q]$, is obtained through time-averaged sampling of the instantaneous system occupancy $Q(t)$ over the total simulation duration T_{sim}

$$\hat{E}[Q] = \frac{1}{T_{\mathrm{sim}}} \int_0^{T_{\mathrm{sim}}} Q(t)\, dt. \tag{8}$$

The empirical average system time per packet, $\hat{E}[t_{\mathrm{q}}]$, is computed based on the observed sojourn times of all successfully transmitted packets. Let t_i^{arr} and t_i^{dep} represent the arrival and departure times, respectively, of the i-th transmitted packet. Then,

$$\hat{E}[t_{\mathrm{q}}] = \frac{1}{D} \sum_{i=1}^{D} \left(t_i^{\mathrm{dep}} - t_i^{\mathrm{arr}} \right). \tag{9}$$

IV. Results

This section presents a comprehensive performance evaluation of the cooperative relay system implementing the XOR queueing mechanism. The results were obtained through discrete-event simulation as described in Section III, with analytical models serving as theoretical benchmarks for validation. The evaluation focuses on three key performance indicators: average system time per packet, average system occupancy, and blocking probability.

Algorithm 1 Discrete-Event Simulation of the Cooperative Relay System

Initialize simulation clock: $t \leftarrow 0$
Initialize counters: total arrivals $A \leftarrow 0$, blocked arrivals $B \leftarrow 0$, departures $D \leftarrow 0$
Initialize system state: queue length $q \leftarrow 0$, server state (busy/idle)
Generate first arrival time: t_{arr} from exponential distribution with rate λ
Set next departure time: $t_{\text{dep}} \leftarrow \infty$
while $D < D_{\max}$ **do**
  if $t_{\text{arr}} < t_{\text{dep}}$ **then** ▷ Arrival event
    $t \leftarrow t_{\text{arr}}$
    $A \leftarrow A + 1$
    if server is busy **then**
      if $q < N - 1$ **then**
        Enqueue packet: $q \leftarrow q + 1$
      else
        $B \leftarrow B + 1$ ▷ Packet blocked
      end if
    else
      Start service immediately
      Schedule departure: t_{dep} from exponential with rate μ
    end if
    Generate next arrival: t_{arr} from exponential with rate λ
  else ▷ Departure event
    $t \leftarrow t_{\text{dep}}$
    $D \leftarrow D + 1$
    if $q \geq 1$ **then**
      if $q \geq 2$ **and** XOR mode enabled **then**
        Remove two packets: $q \leftarrow q - 2$
      else
        Remove one packet: $q \leftarrow q - 1$
      end if
      Schedule next departure: t_{dep} from exponential with rate μ
    else
      Set server to idle: $t_{\text{dep}} \leftarrow \infty$
    end if
  end if
  Record system occupancy: $Q(t) \leftarrow q +$ server state
end while
Compute performance estimates: $\hat{P}_{\text{b}}$, $\hat{E}[Q]$, $\hat{E}[t_{\text{q}}]$

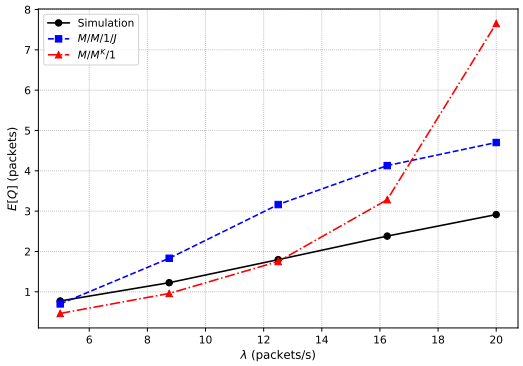

Fig. 3. Average system occupancy versus arrival rate λ.

For statistical reliability, each simulation run was executed until a total of 100,000 packet departures occurred, and the random number generator was initialized with a fixed seed to ensure reproducibility. The theoretical results for the $M/M/1/J$ and $M/M^2/1$ models were computed following the derivations in Section II. All code is publicly available[1].

To systematically assess the system behavior under varying load and resource conditions, each parameter of interest—arrival rate (λ), service rate (μ), and buffer size (N)—was varied independently, while the others were kept constant at their default values: $\lambda = 10$ packets/s, $\mu = 12$ packets/s, and $N = 5$ buffer positions (including the server).

A. Average Number of Packets in the System

Figs. 3–5 illustrate the relationship between the average number of packets in the system ($E[Q]$) and the varying system parameters.

As shown in Fig. 3, the average system occupancy increases with the arrival rate for all models. The XOR relay closely follows the $M/M/1/J$ model at low arrival rates, where the probability of having multiple packets in the system is small, resulting in behavior dominated by single-packet transmissions. However, as λ increases, the XOR relay begins to leverage its batch-service capability through XOR combinations, resulting in a lower average occupancy compared to the $M/M/1/J$, but still higher than the $M/M^2/1$ due to buffer limitations.

Fig. 4 shows that increasing μ reduces the system occupancy in all models. The XOR relay exhibits a more gradual decrease compared to the $M/M^2/1$ model, which benefits from idealized, always-available batch processing. In contrast, the XOR relay relies on the stochastic accumulation of packets.

In Fig. 5, as the buffer size increases, the occupancy for both the simulation and $M/M/1/J$ rises. At the same time, the $M/M^2/1$ remains insensitive to N, as expected due to its infinite-buffer assumption.

B. Average System Time

Figs. 6–8 depict the average time each packet spends in the system ($E[t_{\text{q}}]$) under different configurations.

[1] https://github.com/ErosloursViv/TP547/tree/main/Trabalho%20Final%203

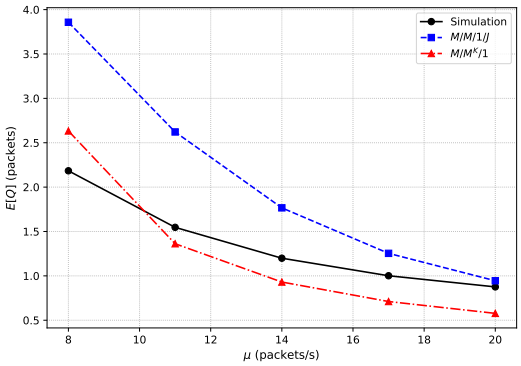

Fig. 4. Average system occupancy versus service rate μ.

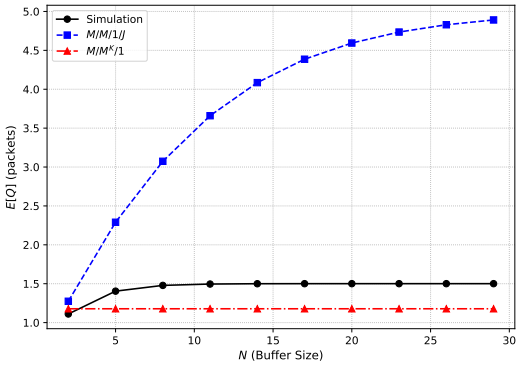

Fig. 5. Average system occupancy versus buffer size N.

Fig. 6 shows that the average system time grows rapidly as λ approaches the service rate μ, especially for the $M/M/1/J$ and simulation results. The XOR relay benefits from reduced system time at higher loads due to occasional batch departures, though the probabilistic nature of XOR pairing limits this gain. The $M/M^2/1$ model predicts consistently lower system times, driven by its assumption of uninterrupted two-packet batch servicing.

As observed in Fig. 7, increasing μ leads to shorter system times for all models. The discrepancy between the XOR relay and the $M/M^2/1$ model narrows for higher μ values, since faster service rates reduce queueing effects.

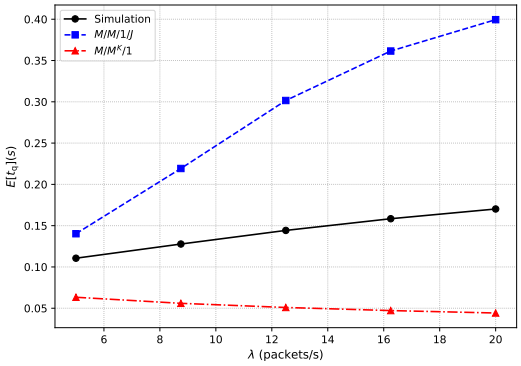

Fig. 6. Average system time versus arrival rate λ.

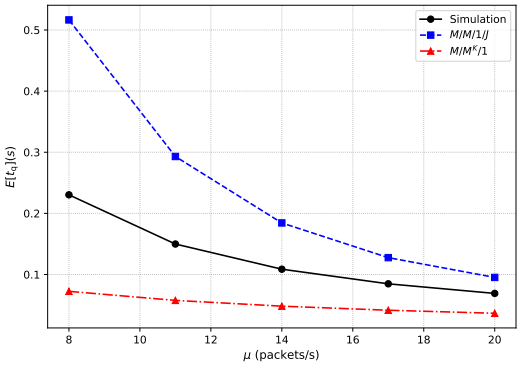

Fig. 7. Average system time versus service rate μ.

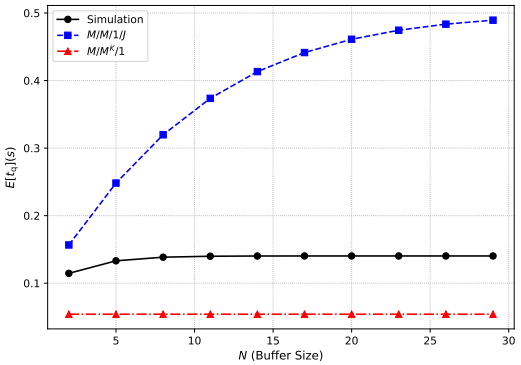

Fig. 8. Average system time versus buffer size N.

Fig. 8 demonstrates that larger buffer sizes allow longer queueing times, increasing $E[t_q]$ for both the XOR relay and $M/M/1/J$, while the $M/M^2/1$ remains constant.

C. Blocking Probability

The blocking performance, shown in Figs. 9–11, is critical in characterizing system reliability under constrained buffer conditions.

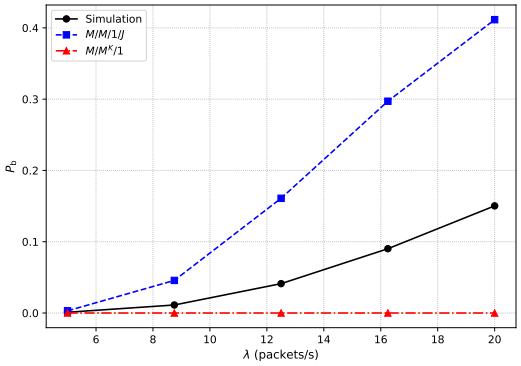

Fig. 9. Blocking probability versus arrival rate λ.

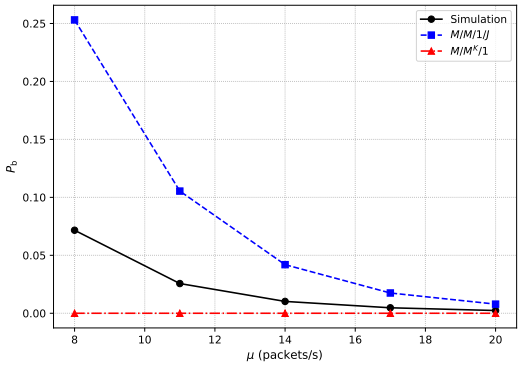

Fig. 10. Blocking probability versus service rate μ.

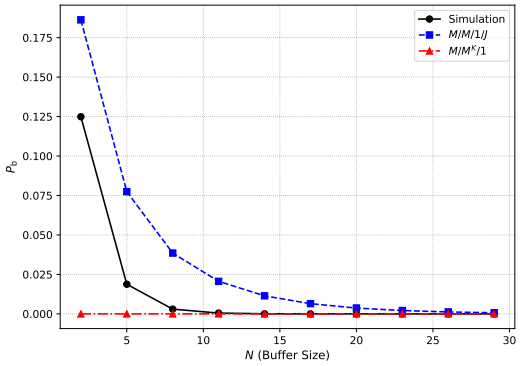

Fig. 11. Blocking probability versus buffer size N.

In Fig. 9, the blocking probability remains negligible at low arrival rates for all models. As λ increases, the blocking probability in the XOR relay closely tracks the $M/M/1/J$ prediction, diverging only at higher loads where batch departures alleviate buffer occupancy.

Fig. 10 illustrates that lower service rates yield higher blocking probabilities due to increased queue buildup. Again, the XOR relay demonstrates slightly lower blocking compared to the $M/M/1/J$ model at moderate loads, highlighting the effectiveness of XOR-based batch processing in relieving buffer congestion.

In Fig. 11, increasing N reduces the blocking probability across both the simulation and $M/M/1/J$, with the XOR relay consistently achieving marginally better performance at intermediate buffer sizes. The $M/M^2/1$ model does not report blocking, reflecting its infinite-buffer assumption.

D. Overall Analysis

The results confirm that the XOR relay exhibits performance characteristics that fall between the two analytical bounds established by the $M/M/1/J$ and $M/M^2/1$ models. At low system loads and small buffer sizes, the XOR relay behaves similarly to the $M/M/1/J$ due to limited packet accumulation. As the load increases, the XOR mechanism introduces opportunities for implicit batching, reducing both system time and occupancy relative to the single-packet service baseline.

However, the XOR relay never fully reaches the performance of the $M/M^2/1$ model due to its dependency on random packet arrivals and the finite buffer constraint. The inability to store arbitrary-sized batches limits the full realization of batch service gains predicted by the theoretical $M/M^2/1$ formulation.

V. Conclusion

This paper presented a comprehensive performance analysis of a cooperative relay system employing XOR-based queueing. Through a combination of discrete-event simulation and analytical modeling, we evaluated the system's behavior under varying traffic loads, service rates, and buffer sizes. Three key performance metrics were studied: average system time, average system occupancy, and blocking probability.

The simulation results demonstrated that the XOR relay achieves a balanced performance that lies between two classical queueing models: the single-packet service $M/M/1/J$ and the idealized batch-service $M/M^2/1$. At low traffic loads and small buffer sizes, the XOR relay closely follows the $M/M/1/J$ behavior, as packet arrivals are sparse and opportunities for XOR combination are limited. However, under moderate to high traffic conditions, the XOR mechanism enables occasional two-packet batch departures, resulting in noticeable improvements in both delay and blocking performance.

The results also highlighted that while the XOR relay benefits from implicit batching, its performance is constrained by the stochastic nature of packet arrivals and the finite buffer capacity. Unlike the $M/M^2/1$ model, which assumes unlimited queueing space and immediate batch availability, the XOR relay must operate under realistic resource constraints.

Overall, the study confirms that XOR queueing offers a practical compromise between system complexity and performance improvement. These findings provide valuable insights for designing efficient buffer management and scheduling policies in cooperative wireless relay networks. Future work may explore adaptive queue management techniques and investigate the impact of traffic burstiness or priority-based scheduling on the performance of XOR relays.

References

[1] H. Xu and B. Li, "An optimization framework for xor-assisted cooperative relaying in cellular networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 5, pp. 979–991, 2014.

[2] S. Chieochan and E. Hossain, "Cooperative relaying in wi-fi networks with network coding," *IEEE Wireless Communications*, vol. 19, no. 2, pp. 57–65, 2012.

[3] S. S and I. Kandaiyan, "Bulk service queueing models - a survey," *International Journal of Pure and Applied Mathematics*, vol. 106, pp. 43–56, 01 2016.

[4] J. Brito, "A simplified queueing model to analyze cooperative communication with network coding," in *Proceedings of the Thirteenth International Conference on Networks (ICN 2014)*, T. Gyires, C. B. Westphall, and G. Kálmán, Eds. Nice, France: IARIA, February 2014, pp. 119–123.

[5] S. Shi, S. Li, and J. Tian, "Performance analysis for practical xor two-way relay under finite relay buffer," *Int. J. Wireless Inf. Networks*, vol. 23, p. 112–121, 2016.

[6] C. Gunasekara, "Analytical model for optimal data transmission policy in two-way relay network coding under phase-type arrivals," *IET Commun.*, vol. 13, p. 2697–2704, 2019.

[7] A. Mahdavi Javid, M. Setayesh, F. Farhadi, and F. Ashtiani, "Analysis of network coding in a slotted aloha-based two-way relay network," in *ArXiv*, 2016.
[8] S. Kafaie, M. H. Ahmed, Y. Chen, and O. A. Dobre, "Throughput analysis of network coding in multi-hop wireless mesh networks using queueing theory," *ArXiv*, 2018.
[9] V. Jamali, N. Zlatanov, and R. Schober, "Buffer-aided relay networks with fixed-rate transmission—delay-constrained case," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, p. 1339–1355, 2015.
[10] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*, 4th ed. Wiley, 2018.