Decoding Probability Analysis of Network-Coded Data Collection and Delivery by Relay Drones

Ioannis Chatzigeorgiou and Elena Manole
School of Computing and Communications
Lancaster University
Lancaster LA1 4WA, United Kingdom
Email: {i.chatzigeorgiou, e.manole}@lancaster.ac.uk

Abstract—**Relay drones in delay-tolerant applications are dispatched to remote locations in order to gather data transmitted by a source node. Collected data are stored on the drones and delivered to one or multiple bases. This paper considers two schemes for broadcasting data to drones when feedback channels are not available: a *data carousel* and *systematic random linear network coding* (RLNC). We propose a theoretical framework for the calculation of the probability that a base will fully or partially recover the transmitted data and the probability that all involved bases will successfully obtain the data, when the bases are either isolated or interconnected. Theoretical results are validated through simulations. Design considerations are also discussed, including the relationship among the field size used by RLNC, the number of relay drones and the requirement for full data recovery or the retrieval of at least part of the data.**

Index Terms—**Unmanned aerial vehicles, random linear network coding, data carousel, fading channel, mission success.**

I. INTRODUCTION

Although unmanned aerial vehicles, commonly referred to as *drones*, were initially considered for military applications, such as surveillance, reconnaissance and rescue operations, their use has been extended to commercial and civil applications, including environmental monitoring, maritime safety, remote sensing and communications. Drones serving as flying base stations or communication relays have the potential to enhance wireless networks and enable future services [1], [2].

The focus of this paper is on *relay drones*, which provide on-demand wireless connectivity between remote network nodes when the line-of-sight channel is blocked, e.g. due to physical obstacles, or is compromised, e.g. due to intentional jamming or unintentional interference. In essence, drones offer another way to "put antennas in the sky" and lower the demand for limited satellite resources [3]. In delay-tolerant applications, relay drones are not always required to establish an end-to-end communication path between two nodes. They can be dispatched from multiple bases to a field site with the objective of collecting information transmitted by an on-site data-gathering source node. For example, the source node could be a wireless sensor forwarding measurements for precision agriculture, or a reconnaissance team sending critical data for emergency or military operations. Collected information is stored on the drones, which carry and deliver it to their designated bases for further processing.

This work considers the broadcasting of information by a source node to clusters of drones in the absence of feedback. The sequential transmission and periodical repetition of data packets, known as a *data carousel* [4], is compared to *random linear network coding* (RLNC) at the source node [5], [6], in terms of the probability that a particular base or all of the bases will fully or partially recover the broadcast information. The motivation for this paper is to characterize the performance of a data carousel, bring together recent theoretical advances in RLNC, e.g. [7]–[9] and develop a framework tailored for broadcast communication aided by relay drones that collect, carry and deliver received packets. The theoretical framework can then be used to quantify performance trade-offs in the two transmission schemes.

The remainder of this paper has been organized as follows: Section II presents the system model, describes the system configurations and proposes performance metrics. Sections III and IV analyze the system configurations and present analytic expressions for the performance metrics. The theoretical framework is validated through simulations in Section V and performance comparisons are discussed. Key findings and conclusions are summarized in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a source node S attempting to deliver a message to N clusters $\mathcal{C}_1, \ldots, \mathcal{C}_N$ of drones, as shown in Fig. 1. The number of drones in cluster $\mathcal{C}_i$ is given by $L_i = |\mathcal{C}_i|$, for $i = 1, \ldots, N$, while the total number of receiving drones is $L = \sum_i L_i$. The j-th drone in cluster $\mathcal{C}_i$ is denoted by $\mathrm{D}_{i,j}$. Each drone in $\mathcal{C}_i$ transports and relays information about the source message to base B_i. The N bases can be either *isolated* or *interconnected*. In the former case, base B_i relies entirely on the information relayed by cluster $\mathcal{C}_i$ to reconstruct the source message. In the latter case, each base has access to the information received by the other $N - 1$ bases.

Before transmission, the message at node S is segmented into k source packets, $u_1, \ldots, u_k$. Node S subsequently broadcasts a sequence of n_T packets, $x_1, \ldots, x_{n_\mathrm{T}}$. The broadcast sequence is composed of the k source packets followed by $n_\mathrm{T} - k$ packets, that is, $x_n = u_n$ for $n = 1, \ldots, k$, while the content of x_n for $n = k+1, \ldots, n_\mathrm{T}$ depends on the employed transmission method. In this paper, we assume that feedback channels between the drones and node S are not available. For
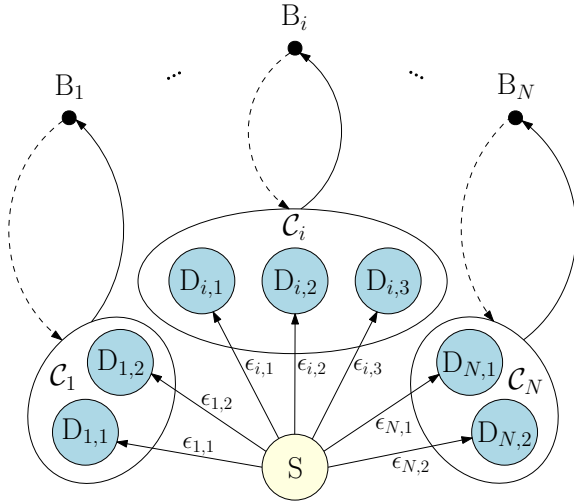
Figure 1. Depiction of a source node S transmitting packets to drone clusters $\mathcal{C}_1, \ldots, \mathcal{C}_N$ over a broadcast erasure channel. At the end of the transmission, each cluster $\mathcal{C}_i$ returns to base B_i to offload received packets.

instance, this could be the case when the adopted communication protocol does not support the reliable broadcasting of data, as in IEEE 802.11. Even when feedback mechanisms are supported, the feedback links of the power-constrained drones could be prone to high rates of packet loss due to interference or intentional jamming. A *data carousel* [4] is a conventional way of improving data reliability in unidirectional broadcast environments by repeatedly transmitting the source packets in a cyclic fashion. The relationship between the source packets and the transmitted packets can be expressed as:

$$x_n = u_{((n-1) \bmod k)+1} \ \text{ for } \ n = k+1, \ldots, n_T, \quad (1)$$

where $\bmod$ denotes the modulo operator. On the other hand, when *systematic RLNC* [7], [10] is employed, the k source packets are followed by $n_T - k$ coded packets, which are random linear combinations of the k source packets. A coded packet x_n can be obtained as follows:

$$x_n = \sum_{b=1}^{k} c_{n,b} \, u_b \ \text{ for } \ n = k+1, \ldots, n_T, \quad (2)$$

where each coefficient $c_{n,b}$ is chosen uniformly at random from a finite field of q elements, denoted by $\mathrm{GF}(q)$, for q a prime power.

The n_T packets are transmitted over a broadcast erasure channel, where $\epsilon_{i,j}$ is the packet erasure probability between the source node S and drone $D_{i,j}$, as depicted in Fig. 1. The packet erasure probability captures the channel conditions as well as the modulation and coding scheme (MCS) used by the transmitting source node and the receiving drones. In cases where the Nakagami-m fading model can accurately describe the channel between node S and drone $D_{i,j}$, e.g. as in [11], the packet erasure probability can be expressed in analytic form as

a function of MCS characteristics and channel parameters [12]. It can also be approximated by:

$$\epsilon_{i,j} \approx \left(\frac{m}{\overline{\gamma}_{i,j}} \right)^m \frac{w_m}{\Gamma(m)}, \quad (3)$$

where $m \geq 0.5$ is the shape factor of the Nakagami distribution, $\Gamma(\cdot)$ is the Gamma function, $\overline{\gamma}_{i,j}$ is the average signal-to-noise ratio (SNR) of the channel between node S and drone $D_{i,j}$, and w_m is an SNR threshold, which is specific to the employed MCS and can be computed using [12, eq. (13)].

After the source node has transmitted the n_T packets, the drones of cluster $\mathcal{C}_i$ do not attempt to reconstruct the source message but return to base B_i to deliver their collected packets. A base will reconstruct the source message, if the k source packets can be recovered from the packets delivered by the respective cluster of drones. The mission is deemed a success if all of the bases reconstruct the source message.

The following two sections present expressions for the probability of a base decoding a source message, either fully or partially, and the probability of mission success. The analysis considers a source node using either a data carousel or systematic RLNC to broadcast packets to clusters of drones, which transport and deliver received packets to isolated or interconnected bases.

III. ANALYSIS FOR ISOLATED BASES

Recall that a packet transmitted by the source node S will be received by drone $D_{i,j}$ with probability $1 - \epsilon_{i,j}$. Base B_i will eventually acquire all packets that have been collected by drones $D_{i,1}, \ldots, D_{i,L_i}$ in cluster $\mathcal{C}_i$. The probability that base B_i will not acquire a packet is the probability of that packet being 'erased' (i.e. not received) by all of the drones in cluster $\mathcal{C}_i$. We can thus write the *equivalent packet erasure probability* ϵ_i experienced by base B_i as:

$$\epsilon_i = \prod_{j=1}^{L_i} \epsilon_{i,j}. \quad (4)$$

Similarly, the probability that base B_i will obtain a particular packet is the probability that at least one of the drones in cluster $\mathcal{C}_i$ will receive it, given by $(1 - \epsilon_i) = 1 - \prod_{j=1}^{L_i} \epsilon_{i,j}$.

A. Transmission based on a Data Carousel

Let λ and ρ be the quotient and remainder, respectively, of the division of the number of transmitted packets n_T by the number of source packets k, that is, $n_T = \lambda k + \rho$ where $\lambda = \lfloor n_T/k \rfloor$ and $\rho = n_T \bmod k$. The sequence of n_T transmitted packets will thus consist of λ copies of the k source packets and one additional copy of the first ρ source packets. In total, $\lambda + 1$ copies of the first ρ source packets and λ copies of the remaining $k - \rho$ source packets will be broadcast.

The probability of base B_i reassembling the source message can be expressed as the product of two probabilities: (i) the probability that at least one copy, out of $\lambda + 1$ copies, of each of the first ρ source packets will be obtained, and (ii) the probability that at least one copy, out of λ copies, of each of

the remaining $k-\rho$ source packets will be retrieved. Therefore, we can write:

$$P_{\text{dc}}(\epsilon_i) = (1 - \epsilon_i^{\lambda+1})^\rho \, (1 - \epsilon_i^\lambda)^{k-\rho}, \qquad (5)$$

where 'data carousel' has been abbreviated to 'dc'.

Given that feedback links are not available, the reliability of the broadcast channel does not affect the order with which packets are being broadcast by the source node. When the N bases are isolated ('iso' for brevity), they will all reconstruct the source message, if and only if each one of them obtains the k source packets. Consequently, the probability of mission success can be expressed as:

$$P_{\text{iso-dc}}(\boldsymbol{\epsilon}) = \prod_{i=1}^{N} P_{\text{dc}}(\epsilon_i), \qquad (6)$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)$ is a vector containing the equivalent erasure probability experienced by each base.

B. Transmission based on Systematic RLNC

Base B_i will be able to reconstruct the k source packets of the message, if the mix of source and coded packets that have been collected by the drones in cluster $\mathcal{C}_i$ contain k linearly independent packets. Let $P_{\text{sr}}(k, n)$ denote the probability that a base will recover the k source packets when the drones of a cluster have deposited n received packets at that base, where $k \leq n \leq n_{\text{T}}$. Index 'sr' in $P_{\text{sr}}(k, n)$ is used as an abbreviation for 'systematic RLNC'. An expression for $P_{\text{sr}}(k, n)$ can be obtained from [7], [10] and can take the form:

$$P_{\text{sr}}(k, n) = \sum_{h=h_{\text{low}}}^{k} \frac{\binom{k}{h}\binom{n_{\text{T}}-k}{n-h}}{\binom{n_{\text{T}}}{n}} \prod_{w=0}^{k-h-1} \left(1 - q^{-n+h+w}\right), \quad (7)$$

where $h_{\text{low}} = \max(0, n - n_{\text{T}} + k)$. Note that (7) quantifies the probability that a base will obtain h source (hence, linearly independent) packets and $n - h$ coded packets, among which $k - h$ will be linearly independent, for all valid values of h. If a particular cluster $\mathcal{C}_i$ is considered, which experiences packet erasures with equivalent probability ϵ_i as defined in (4), the average probability of base B_i recovering the *full* message is given by [7]:

$$P_{\text{sr}}(\epsilon_i) = \sum_{n=k}^{n_{\text{T}}} \binom{n_{\text{T}}}{n} (1 - \epsilon_i)^n \, \epsilon_i^{n_{\text{T}}-n} \, P_{\text{sr}}(k, n). \qquad (8)$$

For RLNC over fields of large size, base B_i is highly likely to recover the k source packets as soon as k different packets are deposited at B_i. That is, $P_{\text{sr}}(k, n) \to 1$ for $q \to \infty$ [13], [14]. As a result, $P_{\text{sr}}(\epsilon_i)$ in (8) reduces to a complementary binomial cumulative distribution function, while the network-coded broadcast flow behaves like a collection of independent network-coded unicast flows. The probability of delivering the full message to the N isolated bases, which is the probability of mission success, can thus be approximated by:

$$P_{\text{iso-sr}}(\boldsymbol{\epsilon}) \approx \prod_{i=1}^{N} P_{\text{sr}}(\epsilon_i), \text{ for } q \to \infty \text{ and } P_{\text{sr}}(k, n) \to 1, \quad (9)$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)$. Approximation (9) can be extended to non-systematic RLNC but can be loose, especially when finite fields of small size are used, as explained in [9]. When systematic RLNC is employed, approximation (9) becomes a lower bound, that is [9, Theorem 2]:

$$P_{\text{iso-sr}}(\boldsymbol{\epsilon}) \geq \prod_{i=1}^{N} P_{\text{sr}}(\epsilon_i), \text{ for } \textit{any} \text{ valid } q. \qquad (10)$$

The bound becomes tighter for increasing values of q and k, and decreasing values of N and ϵ_i. As reported in [9], the mean squared error (MSE) between the exact probability value – computed via simulations – and the lower bound is $9 \cdot 10^{-5}$ for $q = 2$, $k = 20$, $N = 6$ and $\boldsymbol{\epsilon} = (0.1, \ldots, 0.1)$. The MSE drops to $5 \cdot 10^{-6}$ when the packet erasure probability reduces to $\boldsymbol{\epsilon} = (0.01, \ldots, 0.01)$.

IV. ANALYSIS FOR INTERCONNECTED BASES

The probability of mission success can greatly improve if the N bases are interconnected, that is, each base can share all acquired packets with every other base at the expense of increased latency, which should not exceed a target value set by the system requirements. Even if the full source message may be reconstructed by the interconnected bases within an acceptable time frame, each base could have the additional requirement of recovering at least part of the message as soon as the drones of the respective cluster offload their received packets. In this section, in addition to the probability of mission success, we present expressions for the probability of a base recovering at least μ of the k source packets when it has obtained n of the n_T transmitted packets, where $\mu \leq k$ and $\mu \leq n \leq n_{\text{T}}$.

A. Transmission based on a Data Carousel

Let Y_1 and Y_2 be discrete random variables; Y_1 represents the number of source packets from the set $\{u_1, \ldots, u_\rho\}$ that have been retrieved by a base, whereas Y_2 represents the number of source packets from the set $\{u_{\rho+1}, \ldots, u_k\}$ that have been acquired by the same base. The joint probability mass function of Y_1 and Y_2, denoted by $p_{Y_1,Y_2}(y_1, y_2)$, is the probability that a base will recover *exactly* $Y_1 = y_1$ and $Y_2 = y_2$ source packets. Recall from Section III-A that the number of packets transmitted by the source node can be expressed as $n_{\text{T}} = \lambda k + \rho$.

Base B_i will acquire any y_1 of the first ρ source packets, i.e. $u_1, \ldots, u_\rho$, if at least one copy, out of $\lambda + 1$ copies, of each of these y_1 source packets reaches the base, while all of the copies of the remaining $\rho - y_1$ source packets are erased. The same base will obtain any y_2 of the last $k - \rho$ source packets, if at least one copy, out of λ copies, of each of these y_2 source packets is offloaded at the base, but no copies of the remaining $(k-\rho) - y_2$ source packets are received. Therefore, the joint probability mass function can be written as:

$$p_{Y_1,Y_2}(y_1, y_2) = \binom{\rho}{y_1} (1 - \epsilon_i^{\lambda+1})^{y_1} \, \epsilon_i^{(\lambda+1)(\rho-y_1)}$$
$$\cdot \binom{k-\rho}{y_2} (1 - \epsilon_i^\lambda)^{y_2} \, \epsilon_i^{\lambda[(k-\rho)-y_2]}, \quad (11)$$

which reduces to:

$$p_{Y_1,Y_2}(y_1, y_2) =$$
$$= \binom{\rho}{y_1}\binom{k-\rho}{y_2}(1-\epsilon_i^{\lambda+1})^{y_1}(1-\epsilon_i^\lambda)^{y_2}\,\epsilon_i^{\lambda(k-y_1-y_2)+\rho-y_1}. \quad (12)$$

The probability of base B_i recovering *at least* μ of the k source packets can be expressed in terms of the joint probability mass function as follows:

$$P_{\mathrm{dc}}^{(\mu/k)}(\epsilon_i) = \sum_{y=\mu}^{k}\sum_{y_1=y_{\mathrm{low}}}^{y_{\mathrm{hi}}} p_{Y_1,Y_2}(y_1, y-y_1), \quad (13)$$

where $y_{\mathrm{low}} = \max(0, y-k+\rho)$ and $y_{\mathrm{hi}} = \min(y, \rho)$. The inner sum in (13) computes the probability that base B_i will recover *exactly* y of the k source packets, for all valid values of y_1 and y_2 that give $y_1 + y_2 = y$. Note that y_2 in $p_{Y_1,Y_2}(y_1, y_2)$ has been written as $y_2 = y - y_1$. The outer sum aggregates these probabilities for y ranging from μ to k. For $\mu = k$, we obtain $y = k$, $y_1 = \rho$ and equation (13) collapses to (5), which gives the probability of base B_i acquiring all of the source packets.

To compute the probability of mission success, the group of N interconnected bases can be viewed as a single base for packet offloading, while the union of the N clusters can be treated as a superset of L drones. The system model reduces to a source node broadcasting packets to a single cluster of L drones, which offload received packets at a single base experiencing equivalent erasure probability:

$$\epsilon = \prod_{i=1}^{L}\epsilon_i. \quad (14)$$

Hence, the probability that the N interconnected bases will obtain the k source packets and reconstruct the source message can be calculated using:

$$P_{\mathrm{int\text{-}dc}}(\epsilon) = P_{\mathrm{dc}}(\epsilon), \quad (15)$$

where 'int' is short for 'interconnected' and $P_{\mathrm{dc}}(\cdot)$ is given by (5).

B. Transmission based on Systematic RLNC

Literature on RLNC is primarily concerned with the probability of a receiver recovering all of the k source packets and, hence, reassembling the full transmitted message. The probability of obtaining a fraction of the message was investigated in [8] and can be extended to the system model under consideration. In particular, the probability that a base will recover at least μ source packets from n packets offloaded by a cluster of drones, can be obtained from [8, Proposition 2]:

$$P_{\mathrm{sr}}(\mu, k, n) = \frac{1}{\binom{n_{\mathrm{T}}}{n}} \cdot$$
$$\cdot \sum_{r=\mu}^{\min(n,k)}\sum_{h=h_{\mathrm{low}}}^{r}\left(\binom{k}{h}\binom{n_{\mathrm{T}}-k}{n-h}q^{-(n-h)(k-r)}\prod_{w=0}^{r-h-1}(1-q^{-n+h+w})\right.$$
$$\left.\cdot\sum_{b=b_{\mathrm{low}}}^{r-h}\binom{k-h}{b}\sum_{\ell=0}^{k-h-b}(-1)^\ell\binom{k-h-b}{\ell}\begin{bmatrix}k-h-b-\ell\\r-h-b-\ell\end{bmatrix}_q\right) \quad (16)$$

where $h_{\mathrm{low}} = \max(0, n-n_{\mathrm{T}}+k)$, $b_{\mathrm{low}} = \max(0, \mu-h)$ and $\begin{bmatrix}b\\\ell\end{bmatrix}_q$ is the *Gaussian binomial coefficient* defined by [15, p. 125]:

$$\begin{bmatrix}b\\\ell\end{bmatrix}_q = \begin{cases}\displaystyle\prod_{i=0}^{\ell-1}\frac{(q^b-q^i)}{(q^\ell-q^i)}, & \text{for } \ell \leq b,\\ 0, & \text{for } \ell > b.\end{cases} \quad (17)$$

Expression (16) enumerates all combinations of $r \geq \mu$ linearly independent received packets, which consist of h source packets and $r - h$ coded packets. The coded packets can be decoded into at least $\mu - h$ source packets. If we substitute (16) into (8), the probability that base B_i will recover at least μ of the k source packets and will fully (for $\mu = k$) or partially (for $0 < \mu < k$) reconstruct the source message, is given by:

$$P_{\mathrm{sr}}^{(\mu/k)}(\epsilon_i) = \sum_{n=k}^{n_{\mathrm{T}}}\binom{n_{\mathrm{T}}}{n}(1-\epsilon_i)^n\,\epsilon_i^{n_{\mathrm{T}}-n}\,P_{\mathrm{sr}}(\mu, k, n). \quad (18)$$

Note that for $\mu = k$, then $P_{\mathrm{sr}}(\mu, k, n)$ in (16) reduces to $P_{\mathrm{sr}}(k, n)$ in (7). Therefore, (16) and (18) can be viewed as generalisations of (7) and (8), respectively.

Similar to the data carousel, if the N bases are interconnected, the system reduces to a source node broadcasting packets to a single cluster of L drones, which receive, transport and deliver packets to a single base. Therefore, the probability that the N interconnected bases will fully reconstruct the source message and the mission will be successfully completed is simply:

$$P_{\mathrm{int\text{-}sr}}(\epsilon) = P_{\mathrm{sr}}(\epsilon), \quad (19)$$

where $P_{\mathrm{sr}}(\cdot)$ is given by (8) and ϵ is the equivalent average packet erasure probability experienced by the union of all clusters, defined in (14).

V. RESULTS AND DISCUSSION

In order to demonstrate the exactness of the theoretical framework, the model of a system consisting of one source node, two clusters of drones and two bases was developed in MATLAB. Cluster C_1 contained $L_1 = 3$ drones, while cluster C_2 comprised $L_2 = 2$ drones. The packet erasure probabilities at the five relay drones were set to $\epsilon_{1,1} = 0.45$, $\epsilon_{1,2} = 0.55$, $\epsilon_{1,3} = 0.65$, $\epsilon_{2,1} = 0.3$ and $\epsilon_{2,2} = 0.4$. A message of $k = 20$ packets was broadcast by the source node using a data carousel or systematic RLNC over $\mathrm{GF}(q)$, for $q \in \{2, 8\}$. For each transmission method, 50000 experiments involving the transmission of n_{T} packets by the source node to the two bases, via the two clusters of drones, were carried out for $n_{\mathrm{T}} = 20, \ldots, 35$. Successful missions and instances of a base recovering at least some or all of the source packets were counted and averaged over all experiments. Plots of the measured probabilities, obtained through simulations, and the computed probabilities, obtained from the analytic expressions in Section III and Section IV, are presented in Fig. 2.

As can be seen in Fig. 2a, when the two bases are isolated and a data carousel is used for broadcasting, an increase in the number of transmitted packets causes only a small improvement in the probability of mission success, denoted
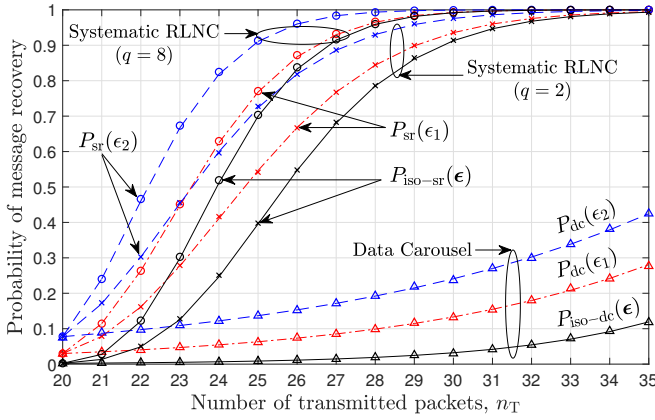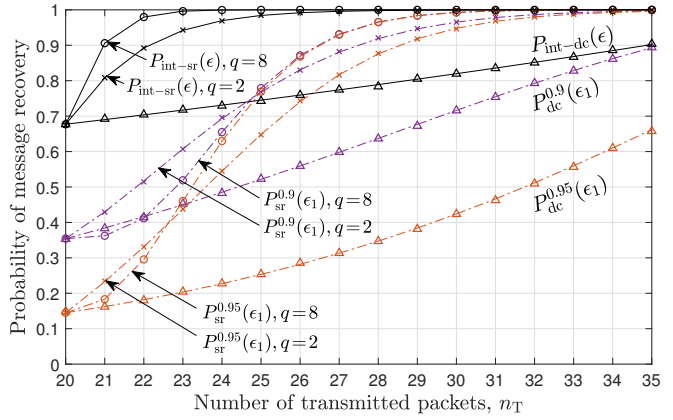
Figure 2. Comparison of theoretical values (curves) and simulation results (markers on curves) for the probability of mission success and the probability of a base, either B_1 or B_2, recovering a source message of $k = 20$ packets for an increasing value of transmitted packets (n_T). The probability of mission success for a data carousel or systematic RLNC is shown in (a) for isolated bases, given by $P_{\text{iso-dc}}(\epsilon)$ and $P_{\text{iso-sr}}(\epsilon)$, respectively, and (b) for interconnected bases, given by $P_{\text{int-dc}}(\epsilon)$ and $P_{\text{int-sr}}(\epsilon)$, respectively. Curves labeled as $P_z(\epsilon_i)$ or $P_z^\alpha(\epsilon_i)$ for $z \in \{\text{dc, sr}\}$ depict the probability that base B_i, for $i \in \{1, 2\}$, will recover all of the source packets or at least αk of the source packets, irrespective of whether the two bases are isolated or interconnected.

by $P_{\text{iso-dc}}(\epsilon)$ and computed using (6). As expected, base B_2 is more likely to reconstruct the source message than base B_1, as the relationship between the equivalent packet erasure probabilities ϵ_1 and ϵ_2, obtained from (4), is $\epsilon_2 < \epsilon_1$, therefore $P_{\text{dc}}(\epsilon_2) > P_{\text{dc}}(\epsilon_1)$. For systematic RLNC over GF(2), the value of n_T has a notable impact on the probability of mission success, i.e. $P_{\text{iso-sr}}(\epsilon)$, which is bounded by (10). As is the case with RLNC, this impact becomes more pronounced when GF(8) is used. Similar observations can be made for the probability that base B_i, for $i \in \{1, 2\}$, will retrieve all of the source packets, given by $P_{\text{sr}}(\epsilon_i)$ in (8). In all cases, simulation results are in agreement with analytic expressions.

If the two bases are interconnected, the probability of mission success improves significantly, as shown in Fig. 2b. In the case of a data carousel, $P_{\text{int-dc}}(\epsilon)$ is close to 90% for $n_T = 35$. Systematic RLNC requires fewer packet transmissions to achieve $P_{\text{int-sr}}(\epsilon) \approx 90\%$, i.e. $n_T = 22$ for $q = 2$, while $n_T = 21$ for $q = 8$. Recall that $P_{\text{int-dc}}(\epsilon)$ and $P_{\text{int-sr}}(\epsilon)$ can be computed using (15) and (19), respectively. Fig. 2b also depicts the probability that base B_1 will recover at least $\mu = 18$ or $\mu = 19$ of the $k = 20$ source packets, thus μ/k is set to 0.9 and 0.95, respectively, in (13) and (18). Both $P_{\text{dc}}^{\mu/k}(\epsilon_1)$ and $P_{\text{sr}}^{\mu/k}(\epsilon_1)$ are computed before any information is exchanged between bases. Therefore, the respective curves in Fig. 2b would remain the same if the bases were isolated. A noteworthy observation is that RNLC over large finite fields yields a high probability of full message recovery but offers a low probability of partial message recovery for a small number of transmitted packets. For example, base B_1 stands a higher chance of recovering at least 18 of the 20 source packets ($\mu/k = 0.9$) if GF(2) is chosen over GF(8) when $21 \leq n_T \leq 24$, as shown in Fig. 2b.

Having validated the proposed theoretical framework, we illustrate with examples how it can contribute to system design. In order to further investigate the observation made in Fig. 2b,

we concentrate on systematic RLNC over GF(q) and look into the impact of the field size q on the probability that a base will partially or fully recover the source message. Let us focus on a base, e.g. B_1, which collects the packets received by a cluster of L_1 drones, where $L_1 \in \{2, 4, 8\}$. For simplicity, we have set $\epsilon_{1,1} = \ldots = \epsilon_{1,L_1} = \epsilon$, where the packet erasure probability ϵ takes values in the range $[0.05, 0.95]$. Fig. 3 depicts the probability that base B_1 will recover at least $\mu = 24$ of the $k = 30$ source packets ($\mu/k = 0.8$) or all of the source packets ($\mu/k = 1$), when $q \in \{2, 4, 8\}$ and $n_T = 36$. The figure shows that, for a given L_1, the probability of full message recovery improves for an increasing value of q, but the trend is reversed when recovery of even a large part of the message is desirable. If the base can afford to dispatch a large cluster of drones, the increased receive diversity will make the system resilient to packet erasures but performance degradation will be abrupt beyond an erasure probability value, e.g. $\epsilon \approx 0.7$ for $L_1 = 8$.

When a single base ($N = 1$) is required to collect data from a remote source node, which needs to keep packet transmissions as low as possible due to energy or time constraints, the optimal number of dispatched drones depends on the expected packet error probability and the chosen transmission method. As in the previous example, we assume that the links between the source node and the relay drones are characterized by the same average packet erasure probability, denoted by ϵ. Fig. 4 depicts the impact of the number of drones L on the number of transmitted packets n_T for $\epsilon \in \{0.1, 0.4, 0.7\}$, when the target probability of mission success is 99%. For a given value of ϵ, we observe that there is a minimum value of L for which the number of transmitted packets becomes equal to the number of source packets, which has been set to $k = 30$. For instance, when $\epsilon = 0.4$, no more than $L = 9$ drones are required to reduce the equivalent erasure probability at the base, given by (4), to a value that minimises the number of transmitted packets to $n_T = 30$. Furthermore, if $L = 9$
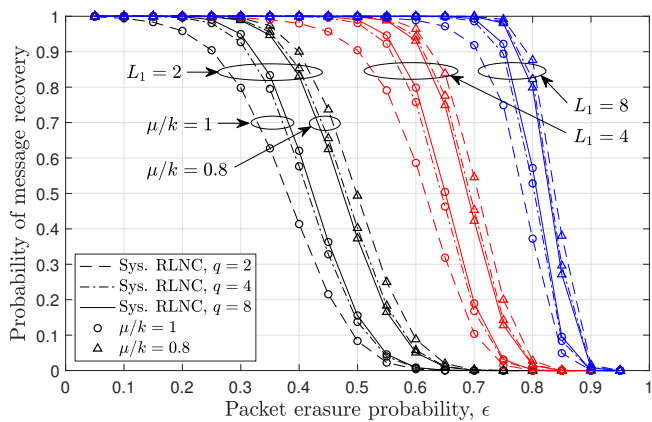
Figure 3. Probability of base B_1 partially ($\mu/k = 0.8$) or fully ($\mu/k = 1$) recovering a message of $k = 30$ source packets. Systematic RLNC over GF(q) is used to broadcast $n_T = 36$ packets to L_1 drones, where $q \in \{2, 4, 8\}$ and $L_1 \in \{2, 4, 8\}$.
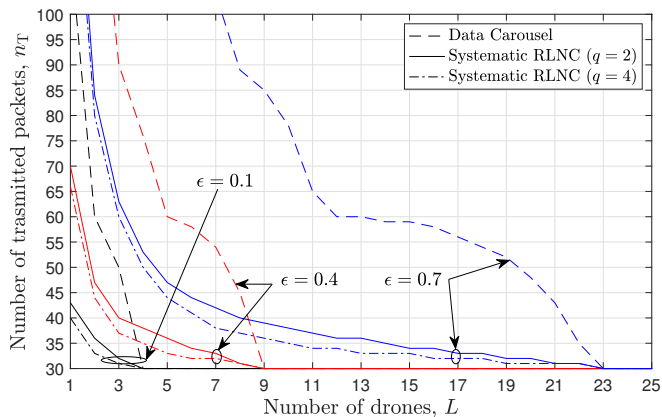

Figure 4. Required number of packet transmissions (n_T) as a function of the number of drones (L) for a single base (i.e., $N = 1$) to fully decode a message of $k = 30$ source packets with probability 99%, when the packet erasure probability is $\epsilon \in \{0.1, 0.4, 0.7\}$.

drones are dispatched, the simple data carousel can be used. The additional encoding and decoding complexity introduced by systematic RLNC can be justified when the available drones are $L = 8$ or fewer, as the value of n_T does not increase as sharply as in the case of the data carousel. If a low value of n_T is essential, an increase in the size of the field over which RLNC is performed, from $q = 2$ to $q = 4$, can marginally decrease the number of transmitted packets, as shown in Fig. 4.

VI. CONCLUSIONS

This paper considered a source node that broadcasts packets to clusters of drones using either a data carousel or systematic RLNC. Successfully received packets are stored, carried and

delivered to bases, which can be either isolated or interconnected. A theoretical framework was developed for the calculation of (i) the probability that a particular base will retrieve some or all of the source packets, and (ii) the probability that all bases will obtain all of the source packets. The framework was validated through simulations and performance trade-offs were identified. Findings established that systematic RLNC over large finite fields should be used when full message recovery is desirable, whereas RLNC over small finite fields could be employed when partial message recovery is essential. Even though systematic RLNC offers a clear advantage over a data carousel, in terms of data reliability, the latter transmission method is still a viable solution if simplicity in the communication process takes priority over the cost of having a large number of drones in each cluster.

REFERENCES

[1] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[2] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, third quarter 2019.

[3] D. F. Carr, "Communications relay grows with expansion of UAV missions," *Defence Systems*, Aug. 2009.

[4] B. Bulut, A. Doufexi, and A. Nix, "Carousel-based wireless multicast transmission over WLANs employing raptor codes," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, London, UK, Sep. 2013.

[5] J. W. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A digital fountain approach to reliable distribution of bulk data," in *Proc. ACM SIGCOMM*, Vancouver, Canada, Oct. 1998.

[6] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Info. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct 2006.

[7] A. L. Jones, I. Chatzigeorgiou, and A. Tassi, "Binary systematic network coding for progressive packet decoding," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, London, UK, Jun. 2015.

[8] J. Claridge and I. Chatzigeorgiou, "Probability of partially decoding network-coded messages," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 1945–1948, Sep. 2017.

[9] E. Tsimbalo, A. Tassi, and R. J. Piechocki, "Reliability of multicast under random linear network coding," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2547–2559, Jun. 2018.

[10] B. Shrader and N. M. Jones, "Systematic wireless network coding," in *Proc. IEEE Military Comm. Conf. (MILCOM)*, Boston, USA, Oct. 2009.

[11] E. Yanmaz, R. Kuschnig, and C. Bettstetter, "Achieving air-ground communications in 802.11 networks with three-dimensional aerial mobility," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013.

[12] Y. Xi, A. Burr, J. Wei, and D. Grace, "A general upper bound to evaluate packet error rate over quasi-static fading channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1373–1377, May 2011.

[13] A. Eryilmaz, A. Ozdaglar, and M. Médard, "On delay performance gains from network coding," in *Proc. 40th Conf. on Inform. Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2006.

[14] D. E. Lucani, M. Médard, and M. Stojanovic, "Broadcasting in time-division duplexing: A random linear network coding approach," in *Proc. Workshop on Network Coding, Theory and Applications (NetCod)*, Lausanne, Switzerland, Jun. 2009.

[15] P. J. Cameron, *Combinatorics: Topics, techniques, algorithms.* Cambridge University Press, 1994.