Distributed Algorithms for Network Lifetime Maximization in Wireless Visual Sensor Networks

Yifeng He, Member, IEEE, Ivan Lee, Senior Member, IEEE,
and Ling Guan, Fellow, IEEE

Abstract

Network lifetime maximization is a critical issue in wireless sensor networks since each sensor has a limited energy supply. In contrast with conventional sensor networks, video sensor nodes compress the video before transmission. The encoding process demands a high power consumption, and thus raises a great challenge to the maintenance of a long network lifetime. In this paper, we examine a strategy for maximizing the network lifetime in wireless visual sensor networks by jointly optimizing the source rates, the encoding powers and the routing scheme. Fully distributed algorithms are developed using the Lagrangian duality to solve the lifetime maximization problem. We also examine the relationship between the collected video quality and the maximal network lifetime. Through extensive numerical simulations, we demonstrate that the proposed algorithm can achieve a much longer network lifetime compared to the scheme optimized for conventional wireless sensor networks.

Index Terms

Wireless visual sensor network, network lifetime maximization, power consumption, convex optimization, distributed algorithms

Response to Reviewers’ Comments is located in Section VIII of this manuscript

Yifeng He and Ling Guan are with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, M5B2K3 Canada. E-mail: (yhe@ee.ryerson.ca, lguan@ee.ryerson.ca.)

Ivan Lee is with School of Computer and Information Science, University of South Australia, Adelaide, Australia. E-mail: (Ivan.Lee@unisa.edu.au.)
I. INTRODUCTION

A wireless sensor network consists of geographically distributed sensors that communicate with each other over wireless channels [1]. A wireless visual sensor network is a special kind of WSN in that each sensor is equipped with video capture and processing components. WVSN facilitates a wide range of applications, such as video surveillance, emergency response, environmental tracking, and health monitoring [2].

An example setup of a WVSN is illustrated in Fig. 1. Each video sensor has a camera component to capture the video, and a processing component to compress the video. The video sensors construct a mesh network topology, and they communicate with each other within a limited transmission range. The video captured and encoded at each sensor is transmitted to a sink for further analysis and decision making.

Sensor nodes are typically battery powered, and battery replacement is infrequent or even impossible in many sensing applications. Hence, a tremendous amount of research efforts in wireless sensor networks has been focused on energy conservation. One aspect of this research is to maximize the network lifetime. In conventional wireless sensor networks, the data processing performed by the sensor node is assumed to be very simple. Thus, the energy consumption utilized for data collection and processing is often negligible. In contrast, video sensors in a WVSN need to compress the data prior to transmission. Efficient video compression algorithms are typically associated with high power consumptions.

It is quite challenging to prolong or maximize the network lifetime for a WVSN. First, the algorithms, which maximize the network lifetime for conventional wireless sensor networks, focuses on the allocation of transmission power and reception power. These algorithms cannot perform well when applied directly into WVSNs, since they neglect the power consumption on signal processing. Second, there is a tradeoff between the video quality and the network lifetime. A WVSN can extend its network lifetime by sacrificing the quality of the collected videos. To the best of our knowledge, such tradeoff has not been investigated in the literature.

This paper tackles the network lifetime maximization problem in wireless visual sensor networks. The contributions of this paper are twofold. First, we propose a distributed algorithm to maximize the network lifetime by jointly optimizing the source rates, the encoding powers, and the routing scheme. We investigate the network lifetime maximization problem in both large-
delay applications and small-delay applications, respectively. The proposed algorithm achieves a longer network lifetime compared to the network lifetime maximization algorithm for conventional wireless sensor networks. Second, we provide the relationship between the collected video quality and the maximal network lifetime in WVSNs. This relationship is very useful for the network design.

The rest of the paper is organized as follows. Related work is discussed in Section II. The system models for the WVSN are described in Section III. In Section IV, we study the achievable maximum network lifetime in the WVSN without transmission errors. The network lifetime maximization in the WVSN with transmission errors is investigated in Section V and Section VI. Section V targets large-delay WVSN applications, while Section VI targets small-delay WVSN applications. Finally, we summarize this paper in Section VII.

II. RELATED WORK

Over the past years, optimization techniques have been used to solve many problems raised in wireless and wired networks. Kelly et al investigated two classes of distributed rate control algorithms for communication networks [3]. Chiang et al applied convex optimization for network utility maximization [4]. Cross-layer optimization for wireless ad hoc networks was presented in [5][6]. Zhu et al proposed to jointly optimize the source rate and the routing scheme for multiple unicast video streams in wireless ad hoc networks [7]. Joint optimization of source coding, routing and resource allocation in wireless sensor networks was reported in [8].
Network lifetime maximization for conventional wireless sensor networks has been extensively studied in the past. Chang et al [9] developed a maximum lifetime routing scheme. Madan et al [10] solved the lifetime maximization problem with a distributed algorithm using the dual decomposition and the subgradient method. Lifetime maximization for interference-limited networks using a cross-layer approach was studied in [11]. The tradeoff between the source rate allocation and the network lifetime was investigated in [12][13]. However, these methods [9][10][11][12][13] cannot be applied directly to the wireless visual sensor networks, since they omit the processing power consumption at the sensor nodes.

Wireless visual sensor networks have recently become an active research area. In [14], the concept of accumulative visual information was introduced as a means for measuring the amount of visual information collected in a WVSN. Minimizing the video distortion by optimizing the power allocation in WVSNs was investigated in [2]. He et al investigated the resource-distortion optimization problem for video encoding and transmission over WVSNs [15]. A cooperative relaying architecture for delivering aggregated high-rate video data to the destination in wireless video sensor networks was proposed in [16]. In order to prolong the network lifetime for wireless sensor networks, an application-aware routing protocol, Distributed Activation based on Predetermined Routes (DAPR) [17], was proposed by avoiding the use of sensors in the sparsely deployed areas as routers. Soro et al extended DAPR in WVSNs [18] by introducing the total cost that combines the coverage and routing costs for each video sensor. However, DAPR cannot prolong the network lifetime for WVSNs to maximum since the encoding power at each sensor node has not been optimized.

III. SYSTEM MODELS

In this section, we describe the network graph, the channel error model, the power consumption model, and give the definition of network lifetime. These models and definition will be used to formulate the network lifetime maximization problem in the next sections.

A. Network Graph

A static wireless visual sensor network can be modeled as a directed graph $G = (N, L)$, where $N$ is the set of nodes and $L$ is the set of directed wireless links. Among the node set $N$,
one node belongs to the sink set $T$, while the other nodes belong to video sensor set $V$. Thus, $N = V \cup T$. In this work, we assume that there is only one sink. However, the setup can be easily extended to include multiple sinks. Two nodes $i$ and $j$ are connected by a link if they can directly communicate with each other.

The relationship between a WVSN node and its connected links is represented with a node-link incidence matrix $A$, whose elements are defined as

$$a_{il} = \begin{cases} 
1, & \text{if link } l \text{ is an outgoing link from node } i, \\
-1, & \text{if link } l \text{ is an incoming link into node } i, \\
0, & \text{otherwise}. 
\end{cases} \quad (1)$$

The relationship between a WVSN node and its outgoing links is represented with a matrix $A^+$, whose elements are given by

$$a_{il}^+ = \begin{cases} 
1, & \text{if link } l \text{ is an outgoing link from node } i, \\
0, & \text{otherwise}. 
\end{cases} \quad (2)$$

The relationship between a WVSN node and its incoming links is represented with a matrix $A^-$, whose elements are given by

$$a_{il}^- = \begin{cases} 
1, & \text{if link } l \text{ is an incoming link from node } i, \\
0, & \text{otherwise}. 
\end{cases} \quad (3)$$

Hence, $A = A^+ - A^-$. We assume a standard Medium Access Control (MAC) layer protocol is applied to resolve the link interference problem. Sensor node $h, \forall h \in V$, can capture and encode the video, and then generate data traffic with a source rate $R_h$ ($R_h = 0$ if sensor $h$ is not on the capture and encoding mode). We define session $h$ as the traffic flow originating from the sensor node $h$ to the sink. For each session, the flow conservation law holds at each node:

$$\sum_{l \in L} a_{il} x_{hl} = \eta_{hi}, \quad \forall h \in V, \forall i \in N, \quad (4)$$

where $x_{hl}$ is the flow rate at link $l$ for session $h$, and $\eta_{hi}$ is defined as

$$\eta_{hi} = \begin{cases} 
R_h, & \text{if } i \text{ is the source node of session } h, \\
-R_h, & \text{if } i \text{ is the sink of session } h, \\
0, & \text{otherwise}. 
\end{cases} \quad (5)$$
B. Channel Error Model

In wireless sensor networks, the channel at each link can be modeled as an Independent and Identically Distributed (IID) random bit error channel. Here, we employ a two-state Markov chain [19] to model the stochastic channel error pattern. The two states of the model are denoted as ”1” (GOOD) and ”0” (BAD). If the channel is in the GOOD state, the bit will be received correctly, and if the channel is in the BAD state, the bit will be received with error. The two-state Markov chain has been widely used to simulate the error patterns in wireless ad hoc networks [20] and wireless sensor networks [21].

Based on the Markov channel error model, the average bit error probability $p_b^l$ at link $l$ is then given by

$$ p_b^l = \frac{q_{10}^l}{q_{10}^l + q_{01}^l}, $$

(6)

where $q_{10}^l$ is the transition probability from a GOOD state to a BAD state, and $q_{01}^l$ is the transition probability from a BAD state to a GOOD state.

Each packet has a fixed length of $G$ bits. A packet is regarded as lost when any bit error in that packet occurs. Then the Packet Loss Rate (PLR) at link $l$ is given by

$$ p_p^l = 1 - (1 - p_b^l)^G. $$

(7)

C. Power Consumption Model

Video sensor $h$ captures and encodes the video before it transmits the traffic to its downstream node. The distortion of the compressed video depends on the source rate $R_h$ and the encoding power consumption $P_{sh}$. According to the Power-Rate-Distortion (P-R-D) analytical model in [22], the encoding distortion is computed by

$$ d_{sh} = \sigma^2 e^{-\gamma \cdot R_h \cdot P_{sh}^{2/3}}, $$

(8)

where $\sigma^2$ is the average input variance, $\gamma$ is the encoding efficiency coefficient.

From the P-R-D model, we can see the following relationships: 1) At a fixed encoding power, the encoding distortion can be reduced by increasing the source rate; 2) At a fixed source rate, the encoding distortion can be reduced by increasing the encoding power. Thus, to achieve a given encoding distortion requirement for session $h$, we can either increase the encoding power or increase the source rate. However, increasing the encoding power may raise the power.
consumption of the source node. On the other hand, increasing the source rate may cause the downstream nodes to consume more power in order to relay extra traffic to the sink. Therefore, optimal allocation between the source rate and the encoding power is critical for maximizing network lifetime.

Based on a power consumption model in wireless sensor networks [23], the *transmission power consumption* at link $l$ can be formulated as:

$$P_{tl} = c^s_{l} y_l, \quad \text{and} \quad c^s_{l} = \alpha + \beta d^p_{l},$$

(9)

where $y_l$ is the aggregate rate transmitted through link $l$, $c^s_{l}$ is the transmission energy consumption cost of link $l$, $\alpha$ is the energy cost of transmit electronics, $\beta$ is a coefficient term corresponding to the energy cost of transmit amplifier, $d_l$ is the distance between the transmitter and the receiver along link $l$, and $n_p$ is the path-loss exponent [24].

The *reception power consumption* at a node $i$ can be formulated as:

$$P_{ri} = c^r \sum_{l \in L} a_{il} y_l,$$

(10)

where $c^r$ is the energy consumption cost of the radio receiver, and $\sum_{l \in L} a_{il} y_l$ represents the aggregate rate received at node $i$.

In the WVSN model, the node set $N$ consists of the sink set $T$ and the video sensor set $V$. If a node belongs to the sink set $T$, it consumes only the reception power. The encoding power and the transmission power at the sink node are both 0. If a node (node $h$) belongs to the video sensor set $V$, it is the source node for session $h$, generating the source bit stream $R_h$. Therefore node $h (h \in V)$ consumes the encoding power and the transmission power for session $h$. At the same time, if node $h$ serves as a relay node, relaying the bit streams for session $j (j \in V, j \neq h)$, node $h$ will also consume the transmission power and the reception power for session $j$. In general, the total power dissipation at node $i (i \in N)$ consists of the encoding power consumption, the transmission power consumption and the reception power consumption, and is given by

$$P_i = P_{si} + P_{ti} + P_{ri} = P_{si} + \sum_{l \in L} a_{il}^+(c^s_{l} y_l) + c^r \sum_{l \in L} a_{il} y_l,$$

(11)

where $P_{si} = 0$, if $i$ is not in the video sensor set $V$.

Equation (11) characterizes the total power dissipation for any node in WVSN. For the sink node $k (k \in T)$, the total power dissipation is actually $P_k = P_{rk} = c^r \sum_{l \in L} a_{kl} y_l$.
since \( P_{sk} = 0 \) and \( \sum_{l \in L} a_{hl}^t(c_l^t y_l) = 0 \). For a video sensor node \( h(h \in V) \), the total power dissipation is \( P_h = P_{sh} + \sum_{l \in L} a_{hl}^t(c_l^t y_l) + c^r \sum_{l \in L} a_{hl}^r y_l \), where the first term \( P_{sh} \) represents the encoding power for session \( h \), the second term \( \sum_{l \in L} a_{hl}^t(c_l^t y_l) \) represents the transmission power for transmitting not only the bit streams for session \( h \) but also the bit streams for the other sessions, and the third term \( c^r \sum_{l \in L} a_{hl}^r y_l \) represents the reception power for relaying the bit streams for the other sessions.

**D. Network Lifetime**

Generally, the network lifetime is defined as the time period from the start time of the network until the time when the whole network fails due to the energy exhaustion of a set of sensors \( K \), where \( K \subseteq N \). In some mission-critical applications, each sensor node is critical to the operation of the network operation. The exhaustion of energy of any node will cause the failure of the whole network. For example, each access is monitored by a visual sensor in a security-monitoring application. If any of the visual sensors fails due to energy exhaustion, the intruder can break into the monitored area. In that case, the whole security-monitoring system loses its function even though most of the visual sensors are working well. The network lifetime in such applications is defined as the minimum node lifetime [9][10][12]. In a WVSN, sensor node \( i \) has an initial energy \( B_i \), and the lifetime of node \( i \) is given by \( T_i = B_i/P_i, \forall i \in N \). Then the network lifetime is given by \( T_{net} = \min_{i \in N} \{ T_i \} = \min_{i \in N} \{ B_i/P_i \} \). In the following, we will use the minimum node lifetime as the network lifetime.

**IV. Achievable Maximum Network Lifetime**

In wireless visual sensor networks, the sink collects each video from the corresponding sensor. The distortion of each video consists of an encoding distortion and a transmission distortion (distortion due to transmission errors). There is a tradeoff between the video quality and the maximum network lifetime. If a high-quality video is desired, the maximum network lifetime will be compromised. On the other hand, WVSN can survive a longer time if it has a lower quality (larger distortion) requirement.

Given a distortion requirement at the sink, the maximum network lifetime depends on the network status. In a WVSN in the presence of transmission errors, sensor nodes can use various
techniques, such as retransmissions, or Forward Error Correction (FEC), to combat the transmission errors. However, these methods introduce additional power consumption and thus reduce the network lifetime. On the other hand, if sensor nodes do not apply retransmission or FEC to recover from transmission errors, the transmission distortion will be introduced. Subject to the total distortion requirement, video sensors need to encode the video with a smaller encoding distortion, which correspondingly requires a higher encoding power or a larger source rate, thus reducing the network lifetime.

The achievable maximum network lifetime in a WVSN is obtained when there is no transmission error. In this section, we will investigate the achievable maximum network lifetime. The maximum network lifetime in a WVSN with transmission errors will be examined in the next two sections: Section V and Section VI.

A. Problem Formulation

In a WVSN that has no transmission error, the total distortion is equal to the encoding distortion since the transmission distortion is zero. In this case, the received video at the sink is measured by the encoding distortion in Mean Squared Error (MSE).

We state the problem of the achievable maximum network lifetime as: given the topology of a static WVSN, and the initial energy at each node, to maximize the network lifetime by jointly optimizing the source rate and the encoding power at each video sensor, and the link rate of each session, subject to the requirement of the collected video quality. Mathematically, the problem can be formulated as follows.

\[
\begin{align*}
\text{maximize}_{(R, x, P_s)} & \quad T_{\text{net}} \\
\text{subject to} & \quad \sum_{l \in L} a_{hl} x_{hl} = \eta_{hi}, \\
& \quad \sum_{h \in V} x_{hl} = y_l, \\
& \quad \sigma^2 e^{-\frac{R_h P_{sh}^{2/3}}{3}} \leq D_h, \\
& \quad T_{\text{net}} = \min_{i \in N} \{T_i\} = \min_{i \in N} \left\{ \frac{B_i}{P_{si} + \sum_{l \in L} a_{il} (c_l y_l) + c_r \sum_{l \in L} a_{il} y_l} \right\}, \\
& \quad x_{hl} \geq 0, \\
& \quad R_h \geq 0, \\
& \quad P_{sh} > 0, \\
& \quad \forall h \in V, \forall i \in N, \\
& \quad \forall l \in L, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
& \quad \forall h \in V, \\
\end{align*}
\]
where $B_i$ is the initial energy at node $i$, $P_i$ is the total power consumption at node $i$, $x_{hl}$ is the link rate at link $l$ for session $h$, $y_l$ is the aggregate flow rate at link $l$, $R_h$ is the source rate of session $h$, $P_{sh}$ is the encoding power at the source node of session $h$, and $D_h$ is the upper bound of the encoding distortion for session $h$ in MSE. The first constraint $\sum_{l \in L} a_{il} x_{hl} = \eta_{hi}$ represents the flow conservation at each node for each session, the second constraint $\sum_{h \in V} x_{hl} = y_l$ represents that the aggregate flow rate $y_l$ at a link is the summation of the link rates of all the sessions at this link, and the third constraint $\sigma^2 e^{-\gamma R_h} P_{sh}^{2/3} \leq D_h$ represents that the encoding distortion for session $h$ is required to be no larger than the corresponding upper bound $D_h$.

We replace the variable $T_{net}$ using $q = 1/T_{net}$. Since $T_{net} \leq B_i/(P_{si} + \sum_{l \in L} a_{il}^+(c_i^s y_l) + c^r \sum_{l \in L} a_{il}^-(c_i^s y_l), \forall i \in N$, we have $q B_i \geq P_{si} + \sum_{l \in L} a_{il}^+(c_i^s y_l) + c^r \sum_{l \in L} a_{il}^-(y_l), \forall i \in N$. Then the problem (12) is converted to an equivalent formulation as follows.

\[
\begin{align*}
\text{minimize}_{(R,x,P)} \quad & q \\
\text{subject to} \quad & \sum_{l \in L} a_{il} x_{hl} = \eta_{hi}, \quad \forall h \in V, \forall i \in N, \\
& \sum_{h \in V} x_{hl} = y_l, \quad \forall l \in L, \\
& \log(\sigma^2/D_h)/(\gamma P_{sh}^{2/3}) \leq R_h, \quad \forall h \in V, \\
& P_{si} + \sum_{l \in L} a_{il}^+(c_i^s y_l) + c^r \sum_{l \in L} a_{il}^-(y_l) \leq q B_i, \quad \forall i \in N, \\
& x_{hl} \geq 0, \quad \forall h \in V, \forall l \in L, \\
& R_h \geq 0, \quad \forall h \in V, \\
& P_{sh} \geq 0, \quad \forall h \in V.
\end{align*}
\]

(13)

In problem (13), we minimize the variable $q$ by jointly optimizing the source rate and the encoding power at each video sensor, and the link rate at each link for each session. However, the optimization problem (13) cannot be solved in a fully distributed manner, because the value of $q$ needs to be broadcast to each node. In order to develop a fully distributed algorithm, we introduce an auxiliary variable $q_i, \forall i \in N$ for node $i$. In problem (13), each node maintains a common $q$, which is equivalent to the case that node $i$ maintains an individual $q_i$ while $q_i = q_j, \forall i,j \in N$. The equality constraint $q_i = q_j, \forall i,j \in N$ can be represented in another way $\sum_{i \in N} a_{il} q_i = 0, \forall l \in L$.

Since $q = (1/T_{net}) > 0$, the objective that minimizes $q$ is equivalent to the one that minimizes $|N| q^2$, where $|N|$ is the number of nodes in the WVSN. \textbf{By using auxiliary variable $q_i (\forall i \in N)$ to replace the common $q$, the objective function $|N| q^2$ is equal to $\sum_{i \in N} q_i^2$ under the equality constraint $q_i = q_j, \forall i \in N$, which can be expressed in another way $\sum_{i \in N} a_{il} q_i = 0, \forall l \in L$.}
0, ∀l ∈ L. Therefore, the optimization problem (13) is converted to the following equivalent formulation:

\[
\begin{align*}
\text{minimize}_{t \in \mathcal{R}, x, \mathbf{P}, \mathbf{q}} & \quad \sum_{i \in \mathcal{N}} q_i^2 \\
\text{subject to} & \quad \sum_{l \in \mathcal{L}} a_{il} x_{hl} = \eta_{hi}, \quad \forall h \in \mathcal{V}, \forall i \in \mathcal{N}, \\
& \quad \log(\frac{\sigma^2}{D_h})/(\gamma_p^{2/3}) \leq R_h, \quad \forall h \in \mathcal{V}, \\
& \quad P_s + \sum_{l \in \mathcal{L}} a_{il}^2 (c_i^s \sum_{h \in \mathcal{V}} x_{hl}) + c^r \sum_{l \in \mathcal{L}} a_{il}^2 \sum_{h \in \mathcal{V}} x_{hl} \leq q_i B_i, \quad \forall i \in \mathcal{N}, \\
& \quad \sum_{i \in \mathcal{N}} a_{il} q_i = 0, \quad \forall l \in \mathcal{L}, \\
& \quad x_{hl} \geq 0, \quad \forall h \in \mathcal{V}, \forall l \in \mathcal{L}, \\
& \quad q_i > 0, \quad \forall i \in \mathcal{N}, \\
& \quad R_h \geq 0, \quad \forall h \in \mathcal{V}, \\
& \quad P_{sh} > 0, \quad \forall h \in \mathcal{V},
\end{align*}
\]

(14)

We will use primal-dual method [26] to develop a distributed algorithm for the optimization problem (14). However, the objective function in problem (14) is not strictly convex with respect to variables \((R, x)\). Therefore, the corresponding dual function is non-differentiable, and the optimal values of \((R, x, \mathbf{P}, \mathbf{q})\) are not immediately available. We add a quadratic regularization term for each link rate variable and each source rate variable to make the objective function strictly convex. Then the optimization problem (14) is approximated to the following:

\[
\begin{align*}
\text{minimize}_{t \in \mathcal{R}, x, \mathbf{P}, \mathbf{q}} & \quad \sum_{i \in \mathcal{N}} q_i^2 + \sum_{h \in \mathcal{V}} \sum_{l \in \mathcal{L}} \delta x_{hl}^2 + \sum_{h \in \mathcal{V}} \delta R_h^2 \\
\text{subject to} & \quad \text{the same constraints as in (14)}
\end{align*}
\]

where \(\delta (\delta > 0)\) is the regularization factor. When the regularization factor \(\delta\) is close to 0, the objective value in problem (15) will be close to the objective value in problem (14).

Let us denote by \((R^*, x^*, \mathbf{P}_s^*, \mathbf{q}^*)\) the optimal solution to the optimization problem (14), and \((\tilde{R}, \tilde{x}, \tilde{P}_s, \tilde{q})\) the optimal solution to the optimization problem (15). Based on the optimization problem (14), we have \(\sum_{i \in \mathcal{N}} (q_i^*)^2 \leq \sum_{i \in \mathcal{N}} \tilde{q}_i^2\) and \(q_i^* = \tilde{q}_i^* = q^*, \forall i \in \mathcal{N}\). Based on the optimization problem (15), we have \(\sum_{i \in \mathcal{N}} \tilde{q}_i^2 + \sum_{h \in \mathcal{V}} \sum_{l \in \mathcal{L}} \delta x_{hl}^2 + \sum_{h \in \mathcal{V}} \delta R_h^2 \leq \sum_{i \in \mathcal{N}} (q_i^*)^2 + \sum_{h \in \mathcal{V}} \sum_{l \in \mathcal{L}} \delta (x_{hl}^*)^2 + \sum_{h \in \mathcal{V}} \delta (R_h^*)^2\) and \(\tilde{q}_i = \tilde{q}_j = \tilde{q}^*, \forall i \in \mathcal{N}\). From the above relationships, we then have the following inequalities: \((q^*)^2 \leq \tilde{q}^2, \sum_{h \in \mathcal{V}} \sum_{l \in \mathcal{L}} \delta (x_{hl}^*)^2 + \sum_{h \in \mathcal{V}} \delta (R_h^*)^2 \geq \sum_{h \in \mathcal{V}} \sum_{l \in \mathcal{L}} \delta x_{hl}^2 + \sum_{h \in \mathcal{V}} \delta R_h^2\) and \(|N| \tilde{q}^2 + \sum_{h \in \mathcal{V}} \sum_{l \in \mathcal{L}} \delta x_{hl}^2 + \sum_{h \in \mathcal{V}} \delta R_h^2 \leq |N| (q^*)^2 + \sum_{h \in \mathcal{V}} \sum_{l \in \mathcal{L}} \delta (x_{hl}^*)^2 + \sum_{h \in \mathcal{V}} \delta (R_h^*)^2\). Then we get the range of \(\tilde{q}^2\): \((q^*)^2 \leq \tilde{q}^2 \leq |N| (q^*)^2 + \sum_{h \in \mathcal{V}} \sum_{l \in \mathcal{L}} \delta (x_{hl}^*)^2 + \sum_{h \in \mathcal{V}} \delta (R_h^*)^2\).
formulate the Lagrangian corresponding to primal problem (15) as below:

\[ (q^*)^2 + \frac{\delta}{|N|} (\sum_{h \in V} \sum_{l \in L} (x^*_l)^2 + \sum_{h \in V} (R^*_h)^2 - \sum_{h \in V} \sum_{l \in L} x^*_l \delta - \sum_{h \in V} \tilde{R}^*_h)^2 \]  

Since \( q^* = \frac{1}{T_{net}^*} \) where \( T_{net}^* \) is the maximum network lifetime obtained from the optimization problem (14), and \( \tilde{q} = \frac{1}{\tilde{T}_{net}} \) where \( \tilde{T}_{net} \) is the maximum network lifetime obtained from the optimization problem (15), we get the range of \( \tilde{T}_{net}^2 \) as below:

\[ \frac{1}{(\tilde{T}_{net})^2} \leq \frac{1}{(T_{net})^2} \leq \frac{1}{T_{net}^*} + \frac{\delta}{|N|} (\sum_{h \in V} \sum_{l \in L} (x^*_l)^2 + \sum_{h \in V} (R^*_h)^2 - \sum_{h \in V} \sum_{l \in L} x^*_l \delta - \sum_{h \in V} \tilde{R}^*_h). \]

In summary, the maximum network lifetime obtained from the optimization problem (15) is smaller than that obtained from the optimization problem (14). However, the loss of the network lifetime is small when the regularization factor \( \delta \) is a small number.

B. Distributed Algorithm

In the problem (15), the objective function is strictly convex, the inequality functions are convex, and the equality functions are linear. Therefore, it is a convex optimization problem [25]. In addition, there exists a strictly feasible solution that satisfies all the constraints in the problem (15). In other words, the Slater’s condition is satisfied, and the strong duality holds [25]. Thus, we can obtain the optimal solutions indirectly by first solving the corresponding dual problem [4][25]. The dual-based approach leads to an efficient distributed algorithm [26].

We introduce dual variables \( u_{hi}, \forall h \in V, \forall i \in N; v_{hi}, \forall h \in V; \lambda_i, \forall i \in N; w_l, \forall l \in L \) to formulate the Lagrangian corresponding to primal problem (15) as below:

\[
L(R, x, P_s, q, u, v, \lambda, w) = \sum_{i \in N} q_i^2 + \sum_{h \in V} \sum_{l \in L} \delta x_{hl}^2 + \sum_{h \in V} \delta R_h^2 + \sum_{h \in V} \sum_{i \in N} u_{hi} (\sum_{l \in L} a_{il} x_{hl} - \eta_{hi}) + \sum_{h \in V} v_{hi} (\log(\sigma^2/D_h)/(\gamma P_{sh}^{2/3}) - R_h) + \sum_{i \in N} \lambda_i (P_{si} + \sum_{l \in L} a_{il}^+ (c_{il}^2 \sum_{h \in V} x_{hl}) + c^r \sum_{l \in L} a_{il}^- \sum_{h \in V} x_{hl} - q_i B_i) + \sum_{l \in L} w_i \sum_{i \in N} a_{il} q_i \\
= \sum_{i \in N} (q_i^2 + q_i (\sum_{l \in L} a_{il} w_l - \lambda_i B_i)) + \sum_{h \in V} (v_{hi} \log(\sigma^2/D_h)/(\gamma P_{sh}^{2/3}) + \lambda_h P_{sh}) + \sum_{h \in V} \sum_{l \in L} (\delta x_{hl}^2 + x_{hl} (c_{il}^2 \sum_{i \in N} \lambda_i a_{il}^+ + c^r \sum_{i \in N} \lambda_i a_{il}^- + \sum_{i \in N} u_{hi} a_{il})) + \sum_{h \in V} (\delta R_h^2 - v_{hi} R_h - \sum_{i \in N} u_{hi} \eta_{hi}).
\]

The Lagrange dual function \( G(u, v, \lambda, w) \) is the minimum value of the Lagrangian over primal variables \( R, x, P_s, q, u, v, \lambda, w \), and it is given by

\[
G(u, v, \lambda, w) = \min \{ L(R, x, P_s, q, u, v, \lambda, w) \}. \]  

(17)
The Lagrange dual problem corresponding to the primal problem (15) is then given by

\[
\begin{align*}
&\text{maximize}_{(u,v,\lambda,w)} \quad G(u, v, \lambda, w) \\
&\text{subject to} \quad v \geq 0, \quad \lambda \geq 0.
\end{align*}
\]  
(18)

The objective function in the Lagrange dual problem is a concave and differentiable function. Therefore we can use subgradient method [27] to find the maximum of the objective function. If the step size \( \theta^{(k)} \) at the \( k^{th} \) iteration follows a non-summable diminishing rule:

\[
\lim_{k \to \infty} \theta^{(k)} = 0, \quad \sum_{k=1}^{\infty} \theta^{(k)} = \infty,
\]  
(19)

the subgradient method is guaranteed to converge to the optimal value [27].

With subgradient method, the dual variable at the \( (k+1)^{th} \) iteration is updated by

\[
u^{(k+1)}_h = \max \{0, v^{(k)}_h - \theta^{(k)}(R^{(k)}_h - \log(\sigma^2/D_h)/(\gamma(P^{(k)}_{sh})^{2/3}))\}, \quad \forall h \in V,
\]  
(21)

\[
\begin{align*}
\lambda^{(k+1)}_i &= \max \{0, \lambda^{(k)}_i - \theta^{(k)}(q^{(k)}_i)B_i - \sum_{l \in L} a^{+}_{il} c^i l \sum_{h \in V} x^{(k)}_{hl} - c^i \sum_{l \in L} a^{+}_{il} \sum_{h \in V} x^{(k)}_{hl} - P^{(k)}_{si}\}, \quad \forall i \in N,
\end{align*}
\]  
(22)

\[
w^{(k+1)}_l = w^{(k)}_l + \theta^{(k)} \sum_{i \in N} a_{il} q^{(k)}_i, \quad \forall l \in L,
\]  
(23)

The step size we use in our algorithm is: \( \theta^{(k)} = \omega/\sqrt{k} \), where \( \omega > 0 \). It follows non-summable diminishing rule.

Given the dual variables at the \( k^{th} \) iteration, we calculate the primal variables as follows:

1) \( q_i \) at node \( i \):

\[
q^{(k)}_i = \arg \min_{q_i > 0} (q^{2}_i + q_i \left( \sum_{l \in L} a_{il} w^{(k)}_l - \lambda^{(k)}_i B_i \right)), \forall i \in N.
\]  
(24)

2) The encoding power \( P_{sh} \) at video sensor \( h \):

\[
P^{(k)}_{sh} = \arg \min_{P_{sh} > 0} \left( v^{(k)}_h \log(\sigma^2/D_h)/(\gamma P^{2/3}_{sh}) + \lambda^{(k)}_h P_{sh} \right), \forall h \in V.
\]  
(25)

3) The source rate \( R_h \) at video sensor \( h \):

\[
R^{(k)}_h = \arg \min_{R_h > 0} (\delta R^2_h - v^{(k)}_h R_h - \sum_{i \in N} u^{(k)}_{hi} \eta_{hi}), \forall h \in V.
\]  
(26)
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
<td>Average input variance of the video in MSE</td>
<td>3500</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Encoding efficiency coefficient</td>
<td>55.54 W$^{3/2}$/Mbps</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Energy cost of transmit electronics</td>
<td>0.5 J/Mb</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Coefficient term of the transmit amplifier</td>
<td>$1.3 \times 10^{-8}$ J/Mb/m$^4$</td>
</tr>
<tr>
<td>$n_p$</td>
<td>Path-loss exponent</td>
<td>4</td>
</tr>
<tr>
<td>$c'$</td>
<td>Energy consumption cost of radio receiver</td>
<td>0.5 J/Mb</td>
</tr>
<tr>
<td>$B_i$</td>
<td>The initial energy at node $i$</td>
<td>5.0 MJ</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Regularization factor</td>
<td>0.2</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Step size parameter</td>
<td>0.15</td>
</tr>
</tbody>
</table>

**TABLE I**

Configuration of model parameters in a wireless visual sensor network

4) The link rate $x_{hl}$ at link $l$ for session $h$:

$$x_{hl}^{(k)} = \arg \min_{x_{hl} \geq 0} (\delta x_{hl}^2 + x_{hl}(c_i \sum_{i \in \mathcal{N}} \lambda_i^{(k)} a_{il}^+ + c' \sum_{i \in \mathcal{N}} \lambda_i^{(k)} a_{il}^- + \sum_{i \in \mathcal{N}} u_{hi}^{(k)} a_{il})), \forall h \in \mathcal{V}, \forall l \in \mathcal{L}. \tag{27}$$

The above algorithm is fully distributed. Each node computes the primal variables: 1) the auxiliary variable $q_i$, 2) the encoding power $P_{sh}$, 3) the source rate $R_h$, and 4) the outgoing link rate $x_{hl}$ from this node, using the dual variables of itself and its neighboring nodes. When the dual variables converge, the primal variables also converge to their optimal values. The message exchange is limited within the one-hop neighbors, thus the communication overhead is reduced greatly.

**C. Simulation Results**

In this subsection, we evaluate the proposed distributed solution for the network lifetime maximization problem in the lossless scenario. We consider a static WVSN with 10 nodes randomly located in a square region of 50m-by-50m. Node 10 is the sink, and the other nodes are video sensors. Each node has a maximum transmission range of 30m. CIF sequence "Foreman" is used in the simulations. The values of the model parameters are listed in Table I. All 9 video sensors encode the videos and transmit them to the sink. If not specified particularly,
the upper bound of the encoding distortion $D_h$ in MSE is set to 100, corresponding to a Peak Signal-to-Noise Ratio (PSNR) of 28.13 dB.

The convergence of the proposed algorithm is shown in Fig. 2. With a convergence threshold of $10^{-5}$, the dual function converges after 390 iterations, as shown in Fig. 2(a). The iteration of the auxiliary variables $q_i$ is illustrated in Fig. 2(b). $q_i$ is computed at each individual node, and they converge to a common $q$. The maximum network lifetime is then given by $T_{net} = 1/q$. The convergence of the encoding power and the source rate is shown in Fig. 2(c) and Fig. 2(d), respectively. We observe that both the encoding power and the source rate converge to the
optimal values when the dual function converges to the maximum. Each node optimally allocates
the encoding power and source rate. The nodes close to the sink have heavy duty in relaying
the traffic. Therefore these nodes encode the video with a lower encoding power and save the
power for transmission and reception. On the other hand, the sensor nodes far away from the
sink encode the video with a higher encoding power but a lower source rate.

The truly maximal network lifetime obtained by solving the optimization problem (14)
using the centralized algorithm is $1.03 \times 10^7$ seconds, shown in Fig. 3(b) at a regularization
factor of 0. We introduce regularization factor in order to develop a distributed algorithm.
The proposed distributed algorithm for WVSN shares the computation burden among all
the nodes at the price of a small performance loss compared to the centralized solution.
As shown in Fig. 3(b), the proposed distributed algorithm sacrifices the maximum network
lifetime by 5.4% at a regularization factor of 0.1 compared to the centralized solution. In
Fig. 3, we can also see the tradeoff between the suboptimality and the computation complexity for
different regularization factor. The proposed distributed algorithm with a smaller regularization
factor can obtain a longer network lifetime at the cost of a more expensive computation. When
the regularization factor is increased from 0.1 to 0.5, the number of iterations for convergence
is reduced from 801 to 277, and the maximum network lifetime is reduced from $9.74 \times 10^6$
seconds to $8.69 \times 10^6$ seconds.
The proposed algorithm jointly optimizes both the source (e.g., the source rate and the encoding power) and the routing scheme. We compare the proposed algorithm to two other schemes: 1) the Routing-Optimized Scheme (ROS) proposed in [10], in which the routing scheme is optimized, while the encoding power at each sensor node is fixed at the same value; and 2) Video-based DAPR (V-DAPR) presented in [18], in which a single route is pre-determined for each session based on the cost metric. In order for fair comparison, the total encoding power of all the sensor nodes in ROS or V-DAPR is equal to that in the proposed scheme. The comparison of the power consumption at each sensor node is illustrated in Fig. 4. In the proposed scheme, the power consumption including the encoding power and the transmission/reception power at each sensor node (represented in bar C) converges to the same level, meaning that each sensor node will exhaust its energy almost at the same time. In ROS (represented in bar B) or V-DAPR (represented in bar A), the power consumption at different sensor node is uneven, thus some nodes will die before other nodes. The network lifetime is determined by the lifetime of the node who has the highest power consumption. The highest power consumption in the proposed algorithm is 0.53 W, smaller by 0.06 W compared to that in ROS, and smaller by 0.16 W compared to that in V-DAPR, respectively.

The aggregate link rates are depicted in Fig. 5. The thickness of an edge is proportional to the amount of aggregate flow at the corresponding link. The traffic is transported via multiple paths.
helping to prolong the network lifetime. The nodes close to the sink (e.g., node 5, 7, 8, and 9) relay the traffic from the nodes far away from the sink, therefore these nodes will consume a higher power in the transmission and reception, which can be observed in Fig. 4.

For the simulation results in Fig. 4 and Fig. 5, the upper bound of the encoding distortion $D_h(\forall h \in V)$ for session $h$ is set to the same value 100, and each sensor node has the same initial energy. Subject to the same encoding distortion upper bound, the proposed algorithm optimally allocate the source rate and the encoding power at each video sensor in order to maximize the network lifetime. For the video sensor far away from the sink, the video is encoded at a lower source rate with a higher encoding power. A lower source rate at the far node helps to reduce the transmission and reception power consumption at the relay nodes. For example, node 1 is far away from the sink, it encodes the video at 0.105 Mbps with the encoding power of 0.48 W, leading to an encoding distortion of 99.96 in MSE. We can observe from Fig. 4 (bar C) that the encoding power at node 1 takes a major part of the total power consumption. On the other hand, the video sensor close to the sink encodes the video at a higher source rate with a lower encoding power. The reasons are: 1) the video sensor close to the sink needs to save more power on relaying the streams from the other video sensors; 2) the encoded stream from the video sensor
close to the sink requires less relays to reach the sink. For example, node 8 is close to the sink, it encodes the video at 0.140 Mbps with an encoding power of 0.31 W, as shown in Fig. 4 (bar C). The encoding distortion at node 8 is 99.58 in MSE.

In some sensing applications, the visual information close to the sink is more important than that far away from the sink. For such applications, we can set a lower upper bound of the encoding distortion $D_h$ for the video sensor close to the sink, and a higher $D_h$ for the one far away from the sink. Fig. 6 shows the power allocation and the PSNR at each node when the upper bound of the encoding distortion $D_h (\forall h \in V)$ is proportional to the distance between the sensor node and the sink. For example, node 1 is the farthest node from the sink, and it is allocated with $D_1 = 150.0$. Node 9 is the closest node, which is allocated with $D_9 = 42.0$. With such prioritized allocation of $D_h$, the sink collects the video stream from node 1 (the farthest node) at an average PSNR of 26.37 dB, while it collects the video stream from node 9 (the closest node) at an average PSNR of 31.89 dB, as shown in Fig. 6(b).

There is a tradeoff between the collected video quality (represented in PSNR) and the achievable maximum network lifetime. If a visual sensor network desires a high-quality video, it will have to sacrifice its network lifetime. On the other hand, a sensor network expecting a
longer lifetime has to lower the quality of the collected video. As shown in Fig. 7, the proposed algorithm supports a longer network lifetime for different quality requirements compared to ROS or V-DAPR, because the proposed algorithm optimizes the allocation of the encoding power, the transmission power and the reception power at each node.

We show the reconstructed picture of frame 1 in Foreman CIF sequence in Fig. 8. If all the video sensors are required with an upper bound of encoding distortion $D_h, \forall h \in V$: (a) $D_h = 300.0$, (b) $D_h = 100.0$, and (c) $D_h = 10.0$

Fig. 7. Tradeoff between the PSNR requirement and the achievable maximum network lifetime in lossless transmission

Fig. 8. Comparison of the visual quality at frame 1 in Foreman CIF sequence with different distortion requirement $D_h, \forall h \in V$: (a) $D_h = 300.0$, (b) $D_h = 100.0$, and (c) $D_h = 10.0$
is increased to 38.38 dB as shown in Fig. 8(c), indicating that the sink collects the videos at an excellent quality. To obtain such excellent quality, each video sensor has to consume more power on encoding and transmitting the bit streams with a high rate, thus shortening the network lifetime to $7.08 \times 10^6$ seconds.

We study the convergence behavior of the proposed distributed algorithm under the dynamic change of the video content in Fig. 9(a). We characterize the video content with the average input variance. Initially the average input variance is 3500 in MSE. At iteration 100, the average input variance is reduced to 2500. The value of the dual function adapts itself to the content change, and quickly evolves to another steady state after 18 iterations. At iteration 200, the average input variance is changed from 2500 to 5500. The value of the dual function transits from the previous steady state to a new steady state after 39 iterations. The adaptation of the proposed algorithm to the dynamic changes of the network topology is shown in Fig. 9(b). At iteration 100, node 7 leaves the network, which causes the transition of the dual function. After 57 iterations, the dual function reaches a new steady state. At iteration 300, a new node joins the network. The dual function adapts itself to the topology change, and converges to another steady state after 131 iterations. The results in Fig. 9 demonstrate that the proposed algorithm can quickly re-converge to
a steady state under dynamic conditions.

We vary the size of the network, and compare the network lifetime between the proposed algorithm and ROS in Fig. 10. We randomly place 10 nodes in a 50m-by-50m square area, 20 nodes in a 100m-by-100m square area, 30 nodes in a 150m-by-150m square area, respectively. In three scenarios, one node is the sink, all the other nodes are video sensors, which capture and transmit the video to the sink. As shown in Fig. 10, when the network size is increased, more sessions are generated in the WVSN, thus power dissipation at each node is increased, leading to a shorter network lifetime. The proposed distributed algorithm archives a longer network lifetime compared to ROS regardless of the variation of the network size and the network topology.

V. Maximum Network Lifetime for Large-Delay Applications

The WVSN applications can be classified into two categories according to their delay requirement [2]. The first category is large-delay WVSN application, in which there is no stringent delay requirement. It only requires that the data be successfully delivered to the sink for future analysis. Environmental data collection belongs to this category. The second category is small-delay WVSN application, in which the video data is required to be transmitted to the sink, over the sensor networks, with a small delay for fast response and decision making. A real-time traffic monitoring system belongs to the second category.
In Section IV, we studied the achievable maximum network lifetime without transmission errors. If transmission errors exist, there is a reduction in the maximum network lifetime compared to the achievable maximum network lifetime. In this section, we investigate this reduction for the large-delay WVSN applications. We will examine the network lifetime reduction for small-delay WVSN applications in the next section.

A. Problem Formulation and Solution

Large-delay WVSN applications, such as video surveillance, typically maintain the video quality at a high priority. Corrupted packets can be retransmitted since delay is permitted. Retransmissions improve the overall video quality at the receiver. However, retransmissions also consume more energy which will reduce the maximum network lifetime. Thus, we analyze the tradeoff between the power consumption and the reliability, and the impact on the maximum network lifetime.

At link $l$, the transmitter sends a packet to the receiver. If the packet is received correctly, the receiver will not notify to the transmitter. On the other hand, if the packet is received with errors, the receiver will send back a Negative Acknowledge (NACK) to the transmitter to request retransmission. In this work, the NACK is kept simple and its energy consumption is assumed negligible. We also assume that NACK is always received correctly, and the maximum number of the retransmission $N_{max}$ is sufficiently large. Using the Markov channel error model, the average number of transmissions $\bar{N}_l$ required over link $l$ for successfully transmitting a packet is given by

$$\bar{N}_l = \sum_{k=0}^{N_{max}} (k+1)(p^p_l)^k(1-p^p_l) \approx \frac{1}{1-p^p_l}, \quad \forall l \in L,$$

where $p^p_l$ is the packet loss rate at link $l$, as defined in equation (7).

In the retransmission scenario, the power consumption at node $i$ is modified from equation (11) to the following:

$$P_i = P_{si} + P_{ti} + P_{ri} = P_{si} + \sum_{l \in L} a_{il} \bar{N}_l (c^s_l \sum_{h \in V} x_{hl} + c^r \sum_{l \in L} \bar{N}_l (a_{il} \sum_{h \in V} x_{hl})), \quad \forall i \in N.$$

The network lifetime maximization problem with retransmissions for large-delay WVSN
applications is modified from optimization problem (15) into the following form:

\[
\begin{align*}
\text{minimize}_{(R,x,P_s,q)} & \quad \sum_{i \in N} q_i^2 + \sum_{h \in V} \sum_{l \in L} \delta x_{hl}^2 + \sum_{h \in V} \delta R_h^2 \\
\text{subject to} & \quad \sum_{l \in L} a_{il} x_{hl} = \eta_{hi}, \quad \forall h \in V, \forall i \in N, \\
& \quad \log(\sigma^2 / D_h) / (\gamma P_{sh}^{2/3}) \leq R_h, \quad \forall h \in V, \\
& \quad P_{si} + \sum_{l \in L} a_{il}^+ N_i (c_i^s \sum_{h \in V} x_{hl}) + c_r^+ \sum_{l \in L} N_l (a_{il}^- \sum_{h \in V} x_{hl}) \\
& \quad \leq q_i B_i, \quad \forall i \in N, \\
& \quad \sum_{i \in N} a_{il} q_i = 0, \\
& \quad x_{hl} \geq 0, \quad \forall h \in V, \forall l \in L, \\
& \quad q_i > 0, \quad \forall i \in N, \\
& \quad R_h \geq 0, \quad \forall h \in V, \\
& \quad P_{sh} > 0, \quad \forall h \in V. \\
\end{align*}
\]

(30)

The network lifetime maximization problem in (30) is essentially the same as the achievable maximum network lifetime problem, except that the power consumption constraint in (30) integrates the retransmission power consumption. We can use the primal-dual method proposed in Section IV to obtain a fully distributed solution. Simulation results for the retransmission case for large-delay WVSN applications are presented below.

\[\text{B. Simulation Results}\]

We use the same setup for the video sensor network and the same model parameters as in Section IV-C. In the simulation, we set the packet size to 512 bits. By varying the transition probability from a GOOD state to a BAD state $q_{10}^l$ and the transition probability from a BAD state to a GOOD state $q_{01}^l$, we obtain different average PLRs.

In the error scenarios, each node needs to retransmit the corrupted packets, thus introducing more power consumption in transmission and reception. Extra power consumption leads to a reduction of network lifetime, which is shown in Fig. 11. Compared to the lossless transmission, the maximum network lifetime is reduced by averagely 4.9% when the average PLR is 7.2%, 8.0% when the average PLR is 13.4% and 15.0% when the average PLR is 26.4%, respectively.

We compare the proposed scheme with ROS. In Fig. 12(a), we compare the maximum network lifetime at different quality requirements while the average PLR is 13.4%. In Fig. 12(b), we set the PSNR requirement to 28.13 dB, and compare the maximum network lifetime by varying the

September 25, 2008 DRAFT
IV. MAXIMUM NETWORK LIFETIME FOR SMALL-DELAY APPLICATIONS

For large-delay WVSN applications, wireless video sensors are allowed to retransmit the packets if they are not received correctly. However, for small-delay WVSN applications, due to the stringent PSNR requirement and the very small delay tolerance, retransmission is not feasible. Thus, in small-delay WVSN applications, we focus on the router network lifetime.

Fig. 11. Reduction of maximum network lifetime in large-delay WVSN applications

Fig. 12. Comparison of maximum network lifetime in large-delay WVSN applications: (a) with different PSNR requirement, and (b) with different average PLR.

average PLR. In both cases, the proposed algorithm optimizes the source coding and the routing scheme simultaneously, thus achieving a longer network lifetime compared to ROS.

VI. MAXIMUM NETWORK LIFETIME FOR SMALL-DELAY APPLICATIONS

For large-delay WVSN applications, wireless video sensors are allowed to retransmit the packets if they are not received correctly. However, for small-delay WVSN applications, due to the stringent PSNR requirement and the very small delay tolerance, retransmission is not feasible. Thus, in small-delay WVSN applications, we focus on the router network lifetime.
to the stringent delay requirement, packet retransmissions are infeasible even if the packets are received with errors.

In this section, we study the maximum network lifetime with and without FEC for small-delay WSN applications.

A. With FEC

In wireless sensor networks, there are two popular methods to address the issues of corrupted or lost packets: Automatic Repeat Requests (ARQ) and FEC. ARQ requires packet retransmissions, which are typically applied for large-delay applications. FEC embeds additional bits to detect and recover from corrupted or lost data. This is suitable for small-delay sensing applications [21][28].

The redundant bits introduced by FEC consume transmission power and reception power. Moreover, the encoding and decoding process of FEC also consumes additional power. Therefore, FEC reduces the maximum network lifetime compared to the achievable maximum network lifetime.

1) Problem Formulation and Solution: In this work, we use a kind of FEC, Reed-Solomon (RS) code [29] to recover the corrupted information in erroneous packets. RS code is widely used for image or video communications. Let \( RS(n, k) \) be the code for transmission, where \( n \) is the block size in number of packets, \( k (k < n) \) is the number of information packets, \( m = n - k \) is the number of the redundancy packets. Any combination of \( t_c = \lceil (n - k)/2 \rceil \) erroneous packets out of \( n \) can be recovered [21].

We use RS code at each link. Each node receives the RS blocks from its upstream nodes. The receiving node decodes the RS blocks and then re-encodes them before transmitting the RS codes to its downstream nodes. We apply Real-time Transport Protocol/Real-time Transport Control Protocol (RTP/RTCP) [30] at each node. Thus, each node can estimate the PLRs at its outgoing links. The packets are RS encoded adaptively according to the estimated PLR values.

We use two-state Markov chain to model the transmission channel, and we assume that the channel error at each link is independent to the amount of the traffic. RS coding is applied at each link. At link \( l \), the packet loss rate is \( p_l^p \), the information bit rate is \( x_{hl} \), and the RS code bit rate is denoted by \( z_{hl} \). The packet length is \( G \) bits/packet. Thus, the information rate in packets is \( \lceil x_{hl}/G \rceil \) packets/second, and the rate of the RS code is \( \lceil z_{hl}/G \rceil \) packets/second. In order to
correct the erroneous packets, the number of erroneous packets needs to be less than or equal to the correction capacity of the RS code, which is expressed by

\[ p^p_i \left\lfloor \frac{z_{hl}}{G} \right\rfloor \leq \frac{\left\lfloor \frac{z_{hl}}{G} \right\rfloor - \left\lceil \frac{x_{hl}}{G} \right\rceil}{2}. \]  

(31)

We define a slack factor \( \kappa (\kappa > 1) \). The bit rate of the RS code at link \( l \) is then given by

\[ z_{hl} = \frac{x_{hl}}{1 - 2\kappa p^p_l}. \]

The slack factor is a given system parameter. A larger \( \kappa \) means a stronger protection for the information bits.

We use the power consumption model of RS codec proposed in [31]. The power consumption cost in RS encoding is \( c^{RSE} \), and the power consumption cost in RS decoding is \( c^{RSD} \). The power consumed by RS encoding at node \( i \) is given by

\[ p^RSE_i = c^{RSE} \sum_{l \in L} (a^+_{il} \sum_{h \in V} x_{hl}), \forall i \in N. \]  

(32)

The power consumed by RS decoding at node \( i \) is given by

\[ p^RSD_i = c^{RSD} \sum_{l \in L} (a^-_{il} \sum_{h \in V} x_{hl}), \forall i \in N. \]  

(33)

The total power consumption at node \( i \) consists of video encoding power consumption, RS encoding power consumption, transmission power consumption, RS decoding power consumption, and reception power consumption, such that

\[ P_i = P_{si} + P^{RSE}_i + P_{ti} + P^{RSD}_i + P_{ri} \]

\[ = P_{si} + c^{RSE} \sum_{l \in L} (a^+_{il} \sum_{h \in V} x_{hl}) + \sum_{l \in L} a^+_{il} (c^+ \sum_{h \in V} z_{hl}) + c^{RSD} \sum_{l \in L} (a^-_{il} \sum_{h \in V} x_{hl}) + c^+ \sum_{l \in L} (a^+_{il} \sum_{h \in V} \frac{x_{hl}}{1 - 2\kappa p^p_l}) + \]

\[ + c^{RSD} \sum_{l \in L} (a^-_{il} \sum_{h \in V} x_{hl}) + c^+ \sum_{l \in L} (a^-_{il} \sum_{h \in V} \frac{x_{hl}}{1 - 2\kappa p^p_l}). \]  

(34)

The network lifetime maximization problem with RS code for small-delay WVSN applications is then modified from the optimization problem (15) into:
minimize$_{(R, x, P_s, q)}$  

\[
\sum_{i \in N} q_i^2 + \sum_{h \in V} \sum_{l \in L} \delta x_{hl}^2 + \sum_{h \in V} \delta R_h^2 
\]

subject to 

\[
\sum_{l \in L} a_{il} x_{hl} = \eta_{hi}, \quad \forall h \in V, \forall i \in N,
\]

\[
\log\left(\frac{\sigma^2}{D_h} / \left(\gamma P_{sh}^{2/3}\right)\right) \leq R_h, \quad \forall h \in V,
\]

\[
P_{si} + c^{RSE} \sum_{l \in L} (a_{il}^+ \sum_{h \in V} x_{hl}) + \sum_{l \in L} a_{il}^+ (c^p \sum_{h \in V} \frac{x_{hl}}{1-2\kappa_p}) + c^{RSD} \sum_{l \in L} (a_{il}^- \sum_{h \in V} x_{hl}) + c^r \sum_{l \in L} (a_{il}^- \sum_{h \in V} \frac{x_{hl}}{1-2\kappa_p}) \leq q_i B_i, \quad \forall i \in N,
\]

\[
\sum_{i \in N} a_{il} q_i = 0,
\]

\[
x_{hl} \geq 0,
\]

\[
q_i > 0,
\]

\[
R_h \geq 0,
\]

\[
P_{sh} > 0,
\]

\[(35)\]

The network lifetime maximization problem with RS code in (35) represents a convex problem which can be solved using a fully distributed algorithm via dual decomposition. The simulation results are presented below.

2) Simulation Results: We use the same model parameters shown in Table I and the error scenarios described in Section V-B. The RS encoding power consumption cost is \(c^{RSE} = 8 \times 10^{-5} J/Mb\), and the RS decoding power consumption cost is \(c^{RSD} = 2.1 \times 10^{-4} J/Mb\). The slack factor \(\kappa\) is set to 1.1.
Fig. 13 shows the reduction of network lifetime due to the extra power consumption at the RS coding and decoding. At the PSNR requirement of 28.13 dB, the maximum network lifetime in the lossless transmission is $9.41 \times 10^6$ seconds. It is reduced by 9.6% at the average PLR of 7.2%, and 33.0% at the average PLR of 26.4%.

In the lossy scenarios with FEC, the proposed scheme performs better than ROS. The comparison is demonstrated in Fig. 14. The proposed scheme improves the network lifetime by 14.8% -15.1% over ROS, when the PSNR requirement is varied from 23.36 dB to 38.13 dB. In Fig. 14(b), we vary the PLR from 0 to 26.4%, the proposed scheme prolongs the network lifetime by 13.1%-16.1% over ROS. The improvement comes from the optimal power allocation in the source encoding, the RS encoding and decoding, the transmission and the reception.

**B. Without FEC**

For small-delay WVSN applications, if no FEC is used, transmission errors will cause decoding errors for the reconstructed video. This error may propagate over subsequent video frames due to drifting error, which exists in popular video coding techniques using block-based motion compensation. The video distortion caused by the transmission errors is called transmission distortion.
1) Transmission Distortion: The transmission distortion of each session depends on the end-to-end PLR. First we formulate the end-to-end PLR for each session based on the Markov channel error model described in Section III-B.

For session \( h \), suppose there are \( J_h \) paths, labeled with \( j_h = 1, \ldots, J_h \), from the video sensor \( h \) to the sink. Let \( F(j_h) \) denote the set of the links in path \( j_h \). Path \( j_h \) transports a normalized portion \( b_{j_h} \) of the source traffic \( R_h \) originating from the video sensor \( h \). Hence, \( b_1 + b_2 + \ldots, + b_{J_h} = 1 \).

We define a binary variable for each link \( l \):

\[
Q^j_{lh} = \begin{cases} 
1, & \text{if } l \in F(j_h), \\
0, & \text{otherwise.}
\end{cases}
\]  

(36)

For path \( j_h \), the end-to-end PLR can be computed as

\[
p_{j_h}^E = 1 - \prod_{l \in L} (1 - Q^j_{lh} p_l^p).
\]  

(37)

Typically, the PLR \( p_l^p \) at each link is very small (e.g. \( p_l^p \ll 1 \)). Thus, the end-to-end packet loss rate at path \( j_h \) can be approximated as \( p_{j_h}^E \approx \sum_{l \in L} (Q^j_{lh} p_l^p) \) [32].

With multi-path routing, the traffic of session \( h \) is disseminated over \( J_h \) paths. Therefore, the end-to-end PLR for session \( h \) is given by

\[
p_{h}^E = \sum_{j_h} b_{j_h} p_{j_h}^E \approx \sum_{j_h} b_{j_h} \sum_{l \in L} (Q^j_{lh} p_l^p) = \sum_{l \in L} p_l^p \sum_{j_h} (b_{j_h} Q^j_{lh}) = \frac{1}{R_h} \sum_{l \in L} p_l^p x_{hl}.
\]  

(38)

We use the transmission distortion model presented in [33]. The transmission distortion \( d_{th} \) of session \( h \) is given by

\[
d_{th} = \psi_h \frac{p_h^E}{1 - p_h^E} \overline{F_h} = \psi_h \frac{1}{R_h} \sum_{l \in L} p_l^p x_{hl} \overline{F_h} = \psi_h \overline{F_h} \frac{1}{R_h - \sum_{l \in L} p_l^p x_{hl}},
\]  

(39)

where \( \psi_h \) is the model parameter, and \( \overline{F_h} \) is the time average of the frame difference [33].

The total video distortion \( d_h \) of session \( h \) consists of the encoding distortion \( d_{sh} \) and the transmission distortion \( d_{th} \). The former study concluded that the encoding distortion and transmission distortion are uncorrelated [33]. Therefore \( d_h = d_{sh} + d_{th} \), where \( d_{sh} \) is given by equation (8), and \( d_{th} \) is given by equation (39).

2) Tradeoff Between Maximum Network Lifetime and Video Quality: We apply the distributed algorithm presented in Section IV to the WVSN with transmission errors. The total distortion consists of the encoding distortion and the transmission distortion. Therefore, for the same
network lifetime, the total video distortion with transmission errors is larger than that without transmission errors.

We present a simulation to evaluate the tradeoffs between the maximum network lifetime and the video quality. The same model parameters in Table I are used in this simulation. The parameters of transmission distortion model are set as: $\psi_h = 0.8$ and $F_h = 200$. As illustrated in Fig. 15, the maximum network lifetime is increased at the price of the video quality. At the same maximum lifetime, the packet errors introduce transmission distortion causing a worse video quality. At the network lifetime of $9.41 \times 10^6$ seconds, we can receive a video at 28.13 dB if there is no transmission error. The packet loss introduces transmission distortion, which degrades the received video quality. If the PLR is 7.2%, the PSNR of the received video will be 27.20 dB under the same network lifetime $9.41 \times 10^6$ seconds. The PSNR is decreased to 23.92 dB if the PLR is increased to 26.4%.

In Fig. 16, we compare the tradeoff between the network lifetime and the video quality among the proposed scheme and ROS. We can observe that the tradeoff trace in the proposed scheme is upper than that in ROS. For the same network lifetime, the proposed scheme enables the sink to collect the video at a higher quality. In other words, the proposed scheme extends the network lifetime subject to the same quality requirement.
VII. CONCLUSIONS

In this paper, we have studied the network lifetime maximization problem in wireless visual sensor networks. Firstly, we investigated the achievable maximum network lifetime in WVSN without transmission errors. Then, we further examined the maximum network lifetime considering transmission errors. We investigated the error remedy techniques in both large-delay WVSN applications and small-delay WVSN applications, and studied their impacts on maximum network lifetime. We have derived distributed algorithms using Lagrangian duality to maximize the network lifetime by jointly optimizing the source rates, the encoding powers, and the routing scheme. Through extensive numerical simulations, we demonstrated that the proposed algorithm can support a much longer network lifetime compared to the scheme optimized for the conventional wireless sensor networks.

VIII. RESPONSE TO REVIEWERS’ COMMENTS

We would like to thank all the reviewers for their valuable and important comments. We have significantly improved and revised the paper to clarify and address the various issues that the reviewers pointed out. In this section, we have tried to respond to each reviewers comments, and describe the corresponding changes in the manuscript.


A. Response to Reviewer 1’s Comments

The paper is well written.

Comment 1:
The objective function in (12) is confusing. What is the "min" operation for and on what variables?

Response:
The objective function in the optimization problem (12) is the network lifetime $T_{net}$, which is defined as the minimum among the node lifetimes $T_i$, $\forall i \in N$. That is: $T_{net} = \min_{i \in N} \{ T_i \} = \min_{i \in N} \{ \frac{B_i}{T_i} \} = \min_{i \in N} \{ \frac{B_i}{T_i} + \sum_{l \in L} a_{li} (c_l y_l) + c^r \sum_{l \in L} a_{li} y_l \}.$

We clarified this expression in the optimization problem (12) in the revised manuscript.

Comment 2:
The formulation in (12) and (13) is questionable, since one video session only has one source node which needs power for video encoding while the rest nodes just need transmission energy. So, you have differentiate source nodes from those relaying nodes. Also, a relay node can be a source node for another session and vice versa.

Response:
In general, the total power dissipation at node $i (i \in N)$ consists of the encoding power consumption, the transmission power consumption and the reception power consumption, and is given by

$$P_i = P_{si} + P_{ti} + P_{ri} = P_{si} + \sum_{l \in L} a_{il}^+ (c_l y_l) + c^r \sum_{l \in L} a_{il}^{-} y_l,$$

(40)

where $P_{si} = 0$, if $i$ is not in the video sensor set $V$.

Equation (40) characterizes the total power dissipation for any node in WVSN. For the sink node $k (k \in T)$, the total power dissipation is actually $P_k = P_{rk} = c^r \sum_{l \in L} a_{kl}^{-} y_l$ since $P_{sk} = 0$ and $\sum_{l \in L} a_{kl}^+ (c_l y_l) = 0$. For a video sensor node $h (h \in V)$, the total power dissipation is $P_h = P_{sh} + \sum_{l \in L} a_{hl}^+ (c_l y_l) + c^r \sum_{l \in L} a_{hl}^{-} y_l$, where the first term $P_{sh}$ represents the encoding power for session $h$, the second term $\sum_{l \in L} a_{hl}^+ (c_l y_l)$ represents the transmission power for transmitting not only the bit streams for session $h$ but also the bit streams for the other sessions, and the third term $c^r \sum_{l \in L} a_{hl}^{-} y_l$ represents the reception power for relaying the bit streams for the other sessions.
In the optimization problem (12) and (13), we use Equation (40) to represent the total power dissipation at a node. Equation (40) integrates both the power consumption as a role of the source node role and the power consumption as a role of the relay node.

The detailed revision is shown in the last two paragraphs of Section III-C in the revised manuscript.

**Comment 3:**

Please use $R$ for bit rate instead of $S_h$, in (13), the source rate $s_h$ is also a global variable since the source coding bit rate of a video stream will affect the scheduling of all link on its path. This will affect the distributed algorithm.

**Response:**

We agree with the reviewer on this issue. We used $R_h$ to replace $s_h$ in the revised manuscript.

In the optimization problem (13), the adjustment of the source rate $s_h$ will affect all link rates on its path. Therefore we minimize the objective function by jointly optimizing the source rate and the encoding power at each video sensor node, and the link rate at each link for each session, via a distributed algorithm.

The revision is shown in the fourth paragraph of Section IV-A in the revised manuscript.

**Comment 4:**

Need to justify, either theoretically or experimentally that the distributed algorithm achieves the optimum obtained by centralized solution. How large the performance loss?

**Response:**

The truly maximal network lifetime is obtained by solving the optimization problem (14) using the centralized solution. However centralized solution is non-scalable, thus not appropriate for wireless sensor networks. Our work aims for distributed solution by introducing a small regularization factor. The price for the proposed distributed algorithm is a loss of maximum network lifetime compared to the centralized solution. As shown in Fig. 3(b), the proposed distributed algorithm sacrifices the maximum network lifetime by 5.4% at a regularization factor of 0.1 compared to the centralized solution.

The detailed revision is shown in Fig. 3 and the third paragraph of Section IV-C in the revised manuscript.
Comment 5:
The paper needs to show results on different topologies.

Response:
We vary the size of the network, and compare the network lifetime between the proposed algorithm and Routing-Optimized Scheme (ROS) in Fig. 10. We randomly place 10 nodes in a 50m-by-50m square area, 20 nodes in a 100m-by-100m square area, 30 nodes in a 150m-by-150m square area, respectively. In three scenarios, one node is the sink, all the other nodes are video sensors, which capture and transmit the video to the sink. As shown in Fig. 10, when the network size is increased, more sessions are generated in the WVSN, thus power dissipation at each node is increased, leading to a shorter network lifetime. The proposed distributed algorithm archives a longer network lifetime compared to ROS regardless of the variation of the network size and the network topology.

The detailed revision is shown in Fig. 10 and the last paragraph of Section IV-C in the revised manuscript.

Comment 6:
Also it is expected that the paper provides comparison with existing methods.

Response:
We compare the proposed algorithm to two existing schemes: 1) the Routing-Optimized Scheme (ROS) proposed in [10], in which the routing scheme is optimized, while the encoding power at each sensor node is fixed at the same value; and 2) Video-based DAPR (V-DAPR) presented in [18], in which a single route is pre-determined for each session based on the cost metric. In order for fair comparison, the total encoding power of all the sensor nodes in ROS or V-DAPR is equal to that in the proposed scheme. The comparison of the power consumption at each sensor node is illustrated in Fig. 4. In the proposed scheme, the power consumption at each sensor node (represented in bar C) converges to the same level, meaning that each sensor node will exhaust the energy almost at the same time. In ROS (represented in bar B) or V-DAPR (represented in bar A), the power consumption at different sensor nodes is uneven, thus some nodes will die before other nodes. The network lifetime is determined by the lifetime of the node who has the highest power consumption. The highest power consumption in the proposed algorithm is 0.53 W, smaller by 0.06 W compared to that in ROS, and smaller by 0.16 W
compared to that in V-DAPR, respectively.

We also compare the maximum network lifetime with different PSNR requirement for the collected videos in Fig. 7. The proposed algorithm supports a longer network lifetime for different quality requirements compared to ROS or V-DAPR, because the proposed algorithm optimizes the allocation of the encoding power, the transmission power and the reception power at each node.

The detailed revision is shown in Fig. 4, the fourth paragraph of Section IV-C, Fig. 7, and the eighth paragraph of Section IV-C in the revised manuscript.


**Comment 7:**

Need to demonstrate how the convergence behavior is affected by the dynamic changes of network conditions, such as video content change and network topology change.

**Response:**

We study the convergence behavior of the proposed distributed algorithm under the dynamic change of the video content in Fig. 9(a). We characterize the video content with the average input variance. Initially the average input variance is 3500 in MSE. At iteration 100, the average input variance is reduced to 2500. The value of the dual function adapts itself to the content change, and quickly evolves to another steady state after 18 iterations. At iteration 200, the average input variance is changed from 2500 to 5500. The value of the dual function transits from the previous steady state to a new steady state after 39 iterations. The adaptation of the proposed algorithm to the dynamic changes of the network topology is shown in Fig. 9(b). At iteration 100, node 7 leaves the network, which causes the transition of the dual function. After 57 iterations, the dual function reaches a new steady state. At iteration 300, a new node joins the network. The dual function adapts itself to the topology change, and converges to another steady state after 131 iterations. The results in Fig. 9 demonstrate that the proposed algorithm
can quickly re-converge to a steady state under dynamic conditions.

The detailed revision is shown in Fig. 9 and the second to the last paragraph of Section IV-C in the revised manuscript.

B. Response to Reviewer 2’s Comments

The paper is well written.

Comment 1:
There is a typo error on P5 where the two ‘sesion’ in (5) should be ‘session’.

Response:
We thank the reviewer for pointing out this typo. We corrected it in the revised manuscript.

Comment 2:
From Fig. 4 and 5 as well as the analysis therein, the nodes close to the sink consumes less power in compression and yields lower bit rate streams, which will lead to a lower reconstructed video quality according to the Power-Rate-Distortion model in (8). On the other hand, it is often assumed that the information captured by video sensors close to the sink is much more important than that by those far away from the sink. So, have you ever considered this problem and how would handle it?

Response:
According to Power-Rate-Distortion model \( d_{sh} = \sigma^2 e^{-\gamma R_h P_{sh}^{2/3}} \), the encoding distortion depends on the source rate \( R_h \) and the encoding power \( P_{sh} \). For a distortion requirement, the video sensor can either encode the video at a lower source rate with a higher encoding power, or encode the video at a higher source rate with a lower encoding power.

For the simulation results in Fig. 4 and Fig. 5, the upper bound of the encoding distortion \( D_h (\forall h \in V) \) for session \( h \) is set to the same value 100, and each sensor node has the same initial energy. Subject to the same encoding distortion upper bound, the proposed algorithm optimally allocate the source rate and the encoding power at each video sensor in order to maximize the network lifetime. For the video sensor far away from the sink, the video is encoded at a lower source rate with a higher encoding power. A lower source rate at the far node helps to reduce the transmission and reception power consumption at the relay nodes. For example, node 1 is
far away from the sink, it encodes the video at 0.105 Mbps with the encoding power of 0.48 W, leading to an encoding distortion of 99.96 in MSE. We can observe from Fig. 4 (bar C) that the encoding power at node 1 takes a major part of the total power consumption. On the other hand, the video sensor close to the sink encodes the video at a higher source rate with a lower encoding power. The reasons are: 1) the video sensor close to the sink needs to save more power on relaying the streams from the other video sensors; 2) the encoded stream from the video sensor close to the sink requires less relays to reach the sink. For example, node 8 is close to the sink, it encodes the video at 0.140 Mbps with an encoding power of 0.31 W, as shown in Fig. 4 (bar C). The encoding distortion at node 8 is 99.58 in MSE. Node 8 and node 1 have the same video quality.

I agree with the reviewer that the visual information close to the sink is more important than that far away from the sink in some sensing applications. For such applications, we can set a lower upper bound of the encoding distortion $D_h$ for the video sensor close to the sink, and a higher $D_h$ for the one far away from the sink. Fig. 6 shows the power allocation and the PSNR at each node when the upper bound of the encoding distortion $D_h (\forall h \in V)$ is proportional to the distance between the sensor node and the sink. For example, node 1 is the farthest node from the sink, and it is allocated with $D_1 = 150.0$. Node 9 is the closest node, which is allocated with $D_9 = 42.0$. With such prioritized allocation of $D_h$, the sink collects the video stream from node 1 (the farthest node) at an average PSNR of 26.37 dB, while it collects the video stream from node 9 (the closest node) at an average PSNR of 31.89 dB, as shown in Fig. 6(b).

The detailed revision is shown in Fig. 4, Fig. 5, Fig. 6, and Paragraph 4 to 7 of Section IV-C in the revised manuscript.

C. Response to Reviewer 3’s Comments

In this paper, authors consider the network lifetime maximization problem for wireless visual sensor networks (WVSN), which is critical when each sensor has a limited energy supply. The proposed strategy maximizes the network lifetime by jointly optimizing the power consumption of video encoding, transmission and reception for each node.

Comment 1:
The authors circumvented the lack of strict convexity by proximal optimization, e.g. adding quadratic terms onto the objective function. Generally, the subsequent dual algorithm would exist an oscillation problem. Could authors provide a proof of the proposed algorithm on this issue?

Response:

The objective function in the Lagrange dual problem (18) is a concave and differentiable function. Therefore we can use subgradient method [27] to find the maximum of the objective function. If the step size $\theta^{(k)}(\theta^{(k)} > 0)$ at the $k^{th}$ iteration follows a non-summable diminishing rule:

$$\lim_{k \to \infty} \theta^{(k)} = 0, \quad \sum_{k=1}^{\infty} \theta^{(k)} = \infty,$$

(41)

the subgradient method is guaranteed to converge to the optimal value, which has been proved in [27].

The revision is shown in the fourth paragraph of Section IV-B in the revised manuscript.


Comment 2:

Also, how could the quadratic regularization term for each link rate variable and source rate variable (in a minimized way) be interpreted in terms of video distortion? Those are monotonous functions without tradeoff involved in optimization. Could authors show the simulated video picture?

Response:

We add the quadratic regularization term for each link rate variable and source rate variable to make the objective function strictly convex, such that we can develop distributed algorithm using dual decomposition.

According to Power-Rate-Distortion model $d_{sh} = \sigma^2 e^{-\gamma R_h P_{sh}^{2/3}}$, the encoding distortion depends on the source rate $R_h$ and the encoding power $P_{sh}$. For a distortion requirement, the video sensor can either encode the video at a lower source rate with a higher encoding power, or encode the video at a higher source rate with a lower encoding power. The proposed algorithm optimally allocates the source rate and the encoding power at each sensor node in order to minimize the objective function $f = \sum_{i \in N} q_i^2 + \sum_{h \in V} \sum_{l \in L} \delta x_{hl}^2 + \sum_{h \in V} \delta R_h^2$ subject to the
constraints in the optimization problem (15). If the source rate $R_h$ at video sensor $h$ is increased, $x_{hl}$ will be increased based on flow conservation law, leading to a higher transmission and the reception power consumption at this node and the relay nodes. Subject to a fixed encoding distortion requirement, video sensor $h$ can reduce the encoding power $P_{sh}$ if the source rate $R_h$ is increased. In the optimization problem (15), the constraint $\sum_{i \in N} q_i = 0, \forall l \in L$ requires all $q_i(\forall i \in N)$ are equal when the algorithm converges, and $q_i \geq \frac{P_i}{B_i}$ where $P_i$ is the total power dissipation at node $i$. Therefore, the optimal solution to the optimization problem (15) is to optimally allocate the source rate and the encoding power at each sensor node to enable each sensor node have an equal $P_i$. In a word, the source rate $R_h$ and the link rates $x_{hl}$ are not monotonously increased or decreased, they are allocated at each node accordingly subject to the constraints and the objective of the optimization problem (15).

We show the reconstructed picture of frame 1 in Foreman CIF sequence in Fig. 8. If all the video sensors are required with an upper bound of encoding distortion $D_h = 300.0, \forall h \in V$, the reconstructed frame 1 has a PSNR of 24.55 dB, as shown in Fig. 8(a). By sacrificing the quality of the collected videos, the WVSN can operate for a long network lifetime, $1.16 \times 10^7$ seconds. If $D_h$ is reduced to 100.0, the sink can receive frame 1 at a PSNR of 28.68 dB, as shown in Fig. 8(b). If $D_h$ is further reduced to 10.0, the PSNR of frame 1 is increased to 38.38 dB as shown in Fig. 8(c), indicating that the sink collects the videos at an excellent quality. To obtain such excellent quality, each video sensor has to consume more power on encoding and transmitting the bit streams with a high rate, thus shortening the network lifetime to $7.08 \times 10^6$ seconds.

The detailed revision is shown in the sixth paragraph of Section IV-A, Fig. 8, and the third to the last paragraph of Section IV-C in the revised manuscript.

Comment 3:
From the distributed algorithm, the iterative operations in all nodes are also involved with the energy consumption. Why did they not consider it in the problem formulation?

Response:
We agree with the reviewer that the proposed distributed algorithm itself is also involved with the energy consumption. The proposed algorithm is distributed and scalable. The computation burden is shared among all the nodes in the network. At each iteration, each node only updates
the dual variables and primal variables associated with itself. The computation is very small compared to the video encoding computation performed at each sensor node. Therefore we assume that the power consumption introduced by the distributed algorithm itself is neglectable. This assumption has been accepted in many other distributed algorithms [11][12][13] for network lifetime maximization in the conventional wireless sensor network.


Comment 4:

Besides, authors claimed (15) satisfies the Slater’s condition to maintain the strong duality. In view of the relaxed approximation in (15), could authors provide the coincidence description between (15) and (13)?

Response:

The maximum network lifetime obtained from the optimization problem (14) is equal to that obtained from the optimization problem (13). In the optimization problem (13), since $q = (1/T_{net}) > 0$, the objective that minimizes $q$ is equivalent to the one that minimizes $|N|q^2$, where $|N|$ is the number of nodes in the WVSN. By using auxiliary variable $q_i (\forall i \in N)$ to replace the common $q$, the objective function $|N|q^2$ is equal to $\sum_{i \in N} q_i^2$ under the equality constraint $q_i = q_j, \forall i \in N$. The equality constraint $q_i = q_j, \forall i \in N$ can be expressed in an equivalent way $\sum_{i \in N} a_{il}q_i = 0, \forall l \in L$. Therefore, the optimization problem (13) is converted to the equivalent formulation (14).

The maximum network lifetime obtained from the optimization problem (15) is not equal to but close to that obtained from the optimization problem (14). Let us denote by $(R^*, x^*, P^*_s, q^*)$ the optimal solution to the optimization problem (14), and $(\tilde{R}, \tilde{x}, \tilde{P}_s, \tilde{q})$ the optimal solution to
the optimization problem (15). Based on the optimization problem (14), we have \(\sum_{i\in\mathbb{N}}(q_i^*)^2 \leq \sum_{i\in\mathbb{N}}\tilde{q}_i^2\) and \(q_i^* = q_j^* = q^*, \forall i \in \mathbb{N}\). Based on the optimization problem (15), we have \(\sum_{i\in\mathbb{N}}\tilde{q}_i^2 + \sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} \delta\tilde{x}_{hi}^2 + \sum_{h \in \mathcal{V}} \delta\tilde{R}_h^2 \leq \sum_{i\in\mathbb{N}}(q_i^*)^2 + \sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} \delta(x_{hi}^*)^2 + \sum_{h \in \mathcal{V}} \delta(R_h^*)^2\)

and \(\tilde{q}_i = \tilde{q}_j = \tilde{q}, \forall i \in \mathbb{N}\). From the above relationships, we then have the following inequalities: \((q^*)^2 \leq \tilde{q}^2, \sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} \delta(x_{hi}^*)^2 + \sum_{h \in \mathcal{V}} \delta(R_h^*)^2 \geq \sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} \delta\tilde{x}_{hi}^2 + \sum_{h \in \mathcal{V}} \delta\tilde{R}_h^2\), and \(\sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} \delta\tilde{x}_{hi}^2 + \sum_{h \in \mathcal{V}} \delta\tilde{R}_h^2 \leq \sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} \delta(x_{hi}^*)^2 + \sum_{h \in \mathcal{V}} \delta(R_h^*)^2\).

Then we get the range of \(\tilde{q}^2\): \((q^*)^2 \leq \tilde{q}^2 \leq (q^*)^2 + \frac{\delta}{|\mathcal{N}|}(\sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} (x_{hi}^*)^2 + \sum_{h \in \mathcal{V}} (R_h^*)^2 - \sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} \tilde{x}_{hi}^2 - \sum_{h \in \mathcal{V}} \tilde{R}_h^2).\) Since \(q^* = \frac{1}{T_{net}}\) where \(T_{net}\) is the maximum network lifetime obtained from the optimization problem (14), and \(\tilde{q} = \frac{1}{T_{net}}\) where \(T_{net}\) is the maximum network lifetime obtained from the optimization problem (15), we get the range of \(\tilde{T}_{net}^2\): \(\frac{1}{(T_{net})^2} \leq \frac{1}{(T_{net})^2} - \frac{1}{(T_{net})^2} + \frac{\delta}{|\mathcal{N}|}(\sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} (x_{hi}^*)^2 + \sum_{h \in \mathcal{V}} (R_h^*)^2 - \sum_{h \in \mathcal{V}} \sum_{i \in \mathcal{L}} \tilde{x}_{hi}^2 - \sum_{h \in \mathcal{V}} \tilde{R}_h^2).\) In summary, the maximum network lifetime obtained from the optimization problem (15) is smaller than that obtained from the optimization problem (14). However, the loss of the network lifetime is small when the regularization factor \(\delta\) is a small number.

The revision is shown in the last three paragraphs of Section IV-A in the revised manuscript.

**Comment 5:**

Please clarify Figure 4 to help one understand the performance difference between two schemes (the proposed scheme and ROS scheme). From Figure 4, we could observe that ROS maintains constant encoding power for each node, and thus provides constant video quality. For the proposed scheme, however, the total power keeps constant by dynamically allocating power consumption between encoding and communication operation. Thus, the nodes close to the sink have to spend more power to relay the information from nodes far from the sink, e.g. the node 5, 7, 8 have lower encoding power consumption to generate poor video quality. How do we justify the phenomenon?

**Response:**

According to Power-Rate-Distortion model \(d_{sh} = \sigma^2 e^{-\gamma R_h} P_{sh}^{2/3}\), the encoding distortion depends on the source rate \(R_h\) and the encoding power \(P_{sh}\). For a distortion requirement, the video sensor can either encode the video at a lower source rate with a higher encoding power, or encode the video at a higher source rate with a lower encoding power.

For the simulation results in Fig. 4, the upper bound of the encoding distortion \(D_h(\forall h \in \mathcal{V})\)
for session \( h \) is set to the same value 100, and each sensor node has the same initial energy. Subject to the same encoding distortion upper bound, the proposed algorithm optimally allocate the source rate and the encoding power at each video sensor in order to maximize the network lifetime. For the video sensor far away from the sink, the video is encoded at a lower source rate with a higher encoding power. A lower source rate at the far node helps to reduce the transmission and reception power consumption at the relay nodes. For example, node 1 is far away from the sink, it encodes the video at 0.105 Mbps with the encoding power of 0.48 W, leading to an encoding distortion of 99.96 in MSE. We can observe from Fig. 4 (bar C) that the encoding power at node 1 takes a major part of the total power consumption. On the other hand, the video sensor close to the sink encodes the video at a higher source rate with a lower encoding power. The reasons are: 1) the video sensor close to the sink needs to save more power on relaying the streams from the other video sensors; 2) the encoded stream from the video sensor close to the sink requires less relays to reach the sink. For example, node 8 is close to the sink, it encodes the video at 0.140 Mbps with an encoding power of 0.31 W, as shown in Fig. 4 (bar C). The encoding distortion at node 8 is 99.58 in MSE. Each video sensor generates the same video quality even though it encodes the video with different encoding power.

In Routing-Optimized Scheme (ROS), the routing scheme is optimized while the source is not optimized. In ROS, each sensor encodes the video at the same source rate 0.1197 Mbps with the same encoding power 0.391 W. Each video sensor in ROS generates the same video quality as to that in the proposed scheme. However, the total power dissipation at each sensor node in ROS in different. Some sensor consume more power, and will die before the other sensors. The network lifetime is defined as the minimum node lifetime. Therefore, the maximum network lifetime in ROS is lower than that in the proposed scheme.

The detailed revision is shown in Fig. 4 and Paragraph 4 to 6 of Section IV-C in the revised manuscript.

REFERENCES


