Radio Selection and Data Partitioning for Energy-Efficient Wireless Data Transfer in Real-Time IoT Applications

Di Mu, Mo Sha¹, Kyoung-Don Kang, Hyungdae Yi

Department of Computer Science, State University of New York at Binghamton

Abstract

The importance of real-time wireless data transfer is rapidly increasing for Internet of Things (IoT) applications. For example, smart glasses worn by a doctor need to transmit real-time data to a hospital information system, which performs face detection and recognition, for real-time interaction with recognized patients within a certain deadline, which is ideally a few hundred milliseconds. Other emerging IoT applications, e.g., structural health monitoring, clinical monitoring, and industrial process automation, also require real-time wireless data transfer. Those applications have critical demands for real-time and energy-efficient communication through wireless medium. However, it is very challenging to support stringent timing constraints energy-efficiently through wireless medium due to its inherent unreliability and timing-unpredictability. Fortunately, heterogeneous radios are becoming increasingly available in modern embedded devices, offering new opportunities to use multiple wireless technologies to accommodate the needs of real-time applications. In this paper, we formulate the runtime radio selection and data partitioning for real-time IoT applications as an Integer Linear Programming (ILP) problem and present an optimal algorithm that makes quick and optimal decisions when selecting between two radios, a heuristic algorithm for the platforms with more radios, and a runtime algorithm that reduces deadline miss ratio when facing tight deadlines.

Keywords: Real-Time Data Transfer, Radio Selection, Data Partitioning, Energy Efficiency, IoT

¹Part of this article was published in Proceedings of the DCOSS [1]

Email addresses: dmu1@binghamton.edu (Di Mu), msha@binghamton.edu (Mo Sha), kang@binghamton.edu (Kyoung-Don Kang), hyi2@binghamton.edu (Hyungdae Yi)

¹Corresponding author

Preprint submitted to Ad Hoc Networks June 16, 2020
1. Introduction

The importance of real-time wireless data transfer is rapidly increasing for the Internet of Things (IoT) applications. For example, smart glasses worn by a doctor need to transmit real-time data to a hospital information system, which performs face detection and recognition, for real-time interaction with recognized patients within a certain deadline, which is ideally a few hundred milliseconds [2]. As another example, periodic sensor readings from unmanned aerial vehicles (UAVs) should be delivered every second to a georeferencing system that analyzes the data to determine the real-time position and altitude of UAVs [3]. Other emerging IoT applications, e.g., structural health monitoring [4], clinical monitoring [5], and industrial process automation [6, 7], also require real-time wireless data transfer. In such applications, missing data delivery deadlines may result in cognitive distraction, injury, structural damage, or safety hazard. However, it is very challenging to support stringent timing constraints through wireless medium due to its inherent unreliability and timing-unpredictability. Moreover, the energy constraints significantly amplify the challenge, since most of those IoT devices are battery-powered and achieving high energy efficiency is critical for those applications.

Fortunately, embedded system hardware and radio technologies are advancing fast in recent years. As a result, more and more embedded devices are equipped with heterogeneous radios. For example, Firestorm [8] supports ZigBee and Bluetooth Low Energy (BLE) in one device and TI CC2650 [9] integrates those two radios on a single chip. IOT-Gate-iMX7 [10] is an industrial IoT gateway, which supports 4G/LTE, WiFi, Bluetooth, and Zigbee. LX Cellular Core [11] is a small-sized IoT platform, which features 2G/3G, WiFi, BLE, ANT+, LoRa, Taggle, and SigFox. Heterogeneous radios are becoming increasingly available in modern embedded devices, offering new opportunities to use multiple wireless technologies for real-time applications. However, using multiple heterogeneous radios may enhance the timeliness at the expense of higher energy consumption or vice versa. It is even more challenging to strike a good balance between the two potentially conflicting requirements.

This paper aims to address the previously stated challenges and presents an energy-efficient radio switching and bundling solution to minimize the energy consumption of battery-powered IoT devices\(^2\) for real-time applications.

\(^2\)In this paper, we focus on minimizing the energy consumption on the sender side (IoT end devices), since the IoT gateways are usually not or much less energy-constrained.
and reduce the deadline miss ratio when facing tight deadlines, leveraging the aforementioned hardware advancements. To assure the timeliness, we target at a single-hop application scenario, since most existing solutions relying on multi-hop mesh networks suffer from long latency and high complexity. Our approach conforms to the advanced wireless network technology trend as the industry is investing heavily in network infrastructure to support IoT visions such as smart cities. As a result, more and more access points and edge servers are becoming readily available to support various IoT applications. Specifically, this paper makes the following contributions:

- We formulate the runtime radio switching and bundling as an Integer Linear Programming (ILP) problem;
- We design the Real-Time radio Selection (RT-Select) algorithm that optimally and quickly selects between two radios and partitions data between them at runtime to minimize the energy consumption;
- Based on RT-Select, we design the RT-Select-General algorithm for the platforms with more radios.
- We design the Real-Time traffic Balance (RT-Balance) algorithm that balances the traffic assigned to different radios at runtime to reduce deadline miss ratio when facing tight deadlines.
- We develop the Real-time Radio Switching and Bundling (RRaSB) system that runs on our embedded platform equipped with five heterogeneous radios, selectively makes a subset of radios available at runtime, and allows dynamic radio switching and bundling among them;
- We implement RT-Select, RT-Select-General, and RT-Balance in RRaSB and evaluate them experimentally; experimental results show that our RT-Select and RT-Select-General significantly outperform the baseline (GreenBag) and RT-Balance effectively help RT-Select and RT-Select-General reduce deadline miss ratios.

The remainder of the paper is organized as follows. Section 2 introduces our problem formulation. Section 3 presents the design of RT-Select, RT-Select-General, and RT-Balance. Section 4 describes RRaSB. Section 5 presents our experimental evaluation. Section 6 reviews related work and Section 7 concludes the paper.
2. Problem Formulation

In this section, we formulate the runtime radio selection and data partitioning for real-time applications as an ILP problem. We first introduce some related radio characteristics and then define the objective function and constraints of the ILP problem.

We assume that \( m \) radios, \( R_1, ..., R_m \), are available on an IoT end device. The characteristics of each radio \( R_i \) (\( 1 \leq i \leq m \)) are separated into two categories:

1. variable characteristics related to the bandwidth and reliability of the wireless link between \( R_i \) and the IoT gateway:
   - throughput, \( TH_i \), is the maximum number of data packets which \( R_i \) is able to successfully deliver to the IoT gateway per second;
   - expected transmission count, \( ETX_i \), is the average number of transmission(s) which \( R_i \) needs to attempt to successfully deliver a packet to the IoT gateway.

2. constant characteristics related to energy and time consumption of \( R_i \):
   - switching energy, \( E_{sw,i} \), is the total energy consumed to switch \( R_i \) on and off\(^3\);
   - switching time, \( T_{sw,i} \), is the time taken to switch \( R_i \) on\(^4\);
   - radio base power, \( P_{rb,i} \), is the base power consumed by \( R_i \) when the radio is on and idle;
   - per-transmission energy, \( E_{ta,i} \), stands for the additional energy consumed by \( R_i \) for each packet transmission attempt.

We define the deadline miss ratio as the number of data transfers which are not completed before their deadlines divided by the total number of data transfers. Since the deadline miss ratio directly reflects the performance of real-time applications, we minimize the deadline miss ratio instead of the absolute latency. Thus, our optimization goal is to minimize the radio energy consumption, while meeting the data rate and deadline requirements. To achieve the objective, we select the radio(s) and assign data packets to them.

We assume that there are \( N \) packets required to be delivered by deadline

\(^3\)\( R_i \) is turned off by default after it transmits all assigned packets if the future traffic demand is unknown.

\(^4\)The time taken to switch \( R_i \) off is not included since the radio can be turned off after the deadline if it is not selected for use in the next period.
Let us also assume that $X_i$ packets are assigned to radio $R_i$, where $0 < X_i \leq N$ if $R_i$ is selected or $X_i = 0$ if $R_i$ is not selected. The objective function to minimize is the sender’s energy consumption $E$, which is the sum of the radio switching energy, radio base energy, and radio transmission energy consumed by the selected radios as shown in Eq. 1, where the radio base energy is $P_{rb,i}$ multiplied by the transmission time ($X_i/TH_i$), the radio transmission energy is $E_{ta,i}$ multiplied by $ETX_i$ and $X_i$, and the set $S$ is composed of the indices of all selected radios:

$$\min \left\{ \sum_{i \in S} (E_{sw,i} + P_{rb,i} \times \frac{X_i}{TH_i} + E_{ta,i} \times ETX_i \times X_i) \right\}$$  \hspace{1cm} (1)$$

There are three constraints on variable $X_i$ (the number of packets assigned to $R_i$): (i) $X_i$ is a non-negative integer not greater than $N$ as specified in Eq. 2 (ii) $X_i$ should not exceed the maximum packet delivery capacity of the radio link ($X_{max,i}$) for the deadline $D$ as stated in Eq. 3 and (iii) the total number of packets assigned to all radios should be equal to $N$ as specified in Eq. 4. Therefore, the following constraints should be met to satisfy the traffic demand and deadline requirements:

$$0 \leq X_i \leq N \ (X_i \in \mathbb{N})$$ \hspace{1cm} (2)$$

$$X_i \leq X_{max,i} \equiv (D - T_{sw,i}) \times TH_i$$ \hspace{1cm} (3)$$

$$\sum_{i=1}^{m} X_i = N$$ \hspace{1cm} (4)$$

In addition, let us introduce a Boolean variable, $Y_i$, to indicate whether or not the radio $R_i$ is selected. $Y_i = 1$ if $R_i$ is selected ($X_i > 0$) and $Y_i = 0$ if $R_i$ is not selected ($X_i = 0$).

Given Eq. 2–4, we simplify the objective function $E$ in terms of variables $X_i$ and $Y_i$ as well as coefficients $A_i$ and $B_i$ as follows:

$$\min \left( \sum_{i=1}^{m} [A_i Y_i + B_i X_i] \right)$$ \hspace{1cm} (5)$$

where

$$A_i = E_{sw,i}$$

$$B_i = \frac{P_{rb,i}}{TH_i} + E_{ta,i} \times ETX_i$$ \hspace{1cm} (6)$$

Eq. 2–6 form an ILP problem, which is NP-hard.
Many resource-constrained IoT devices cannot afford to execute an ILP solver to solve the problem at runtime for real-time applications. This motivates us to develop lightweight algorithms tailored for the runtime radio selection and data partitioning problem.

3. Algorithm Design

One of the primary design goals of our algorithms is to be time-efficient. With the consideration of the demand of fast responses, our decision-making strategies can be processed fast by the IoT devices to guide the runtime radio selection and data partitioning in response to the current wireless link state and application timing requirement. Specifically, we first design the RT-Select algorithm that optimally solves the two-radio case of the problem and prove its optimality. Then, based on the insights from the design of RT-Select, we design the RT-Select-General algorithm to solve the general form of the problem involving \( m \) radios. Finally, we design the RT-Balance algorithm that balances the traffic assigned to different radios at runtime to reduce deadline miss ratio when facing tight deadlines. All of our algorithms take the inputs of the traffic demand (i.e., \( N \) packets) and the delivery deadline \( D \) specified by the application and the pre-measured radio characteristics. While RT-Select and RT-Select-General output the radio selection decision, RT-Balance adjusts the traffic assignments at runtime and outputs the result whether the deadline is met successfully.

---

**Algorithm 1: RT-Select**

**Input:** \( N, D, RC_1, RC_2 \)

**Output:** \( X_1, X_2 \)

1. Compute \( A_i, B_i, X_{\max,i} | i = 1, 2 \);
2. \( (idx_1, idx_2) = \text{sort}\{A_i + B_i \times N \mid i = 1, 2\}; \)
3. \( (idx_1', idx_2') = \text{sort}\{B_1, B_2\}; \)
4. if \( X_{\max}(idx_1) \geq N \) then
   5. \( X_{idx_1} \leftarrow N; \)
   6. else if \( X_{\max}(idx_1) < N \) and \( X_{\max}(idx_2) < N \) then
      7. \( X_{idx_1'} \leftarrow X_{\max}(idx_1'); \)
      8. if !Conflict() then
         9. \( X_{idx_2'} \leftarrow N - X_{idx_1'}; \)
      10. end
   11. else if \( B_{idx_2} < B_{idx_1} \) or
        \( A_{idx_1'}/(B_{idx_2} - B_{idx_1'}) > X_{\max}(idx_1') \) or Conflict()
        then
           13. \( X_{idx_2} \leftarrow N; \)
        14. else
           15. \( X_{idx_1'} \leftarrow X_{\max}(idx_1'); \)
           16. \( X_{idx_2} \leftarrow N - X_{idx_1'}; \)
        17. end
   18. end
For simplicity, we use $RC_i$ to represent the characteristics of each radio $R_i$ including $TH_i$, $ETX_i$, $E_{sw,j}$, $T_{sw,j}$, $P_{rb,i}$ and $E_{ta,i}$ (see Section 2).

Please note that an embedded device may not allow to use some of its radios simultaneously due to hardware conflicts. For example, the ZigBee and BLE radios on the TI CC2650 [9] cannot operate simultaneously, since they share a single DSP modem and a digital PLL. Our algorithms always consider such hardware conflicts when selecting radios.

3.1. RT-Select Algorithm for Selection between Two Radios

Algorithm 1 shows RT-Select algorithm that selects between two radios to minimize the energy consumption, while meeting the application specified traffic demand and deadline requirements. We have proven the optimality of Algorithm 1 [1]. RT-Select first computes the $A_i$, $B_i$, and $X_{max,i}$ values for both radios based on Eq. 6 and Eq. 3 (Line 1). It then sorts the two radios based on the energy consumption for each radio to transmit $N$ packets by itself ($A_i + B_i \times N$) and stores the radio indices to $(idx_1, idx_2)$ in ascending order (Line 2). Therefore, the radio $R_{idx,1}$ is more energy-efficient than $R_{idx,2}$. Similarly, RT-Select sorts the two radios based on the average energy consumption per packet $B_i$ without considering radio switching energy consumption $A_i$ and stores the radio indices to $(idx_1', idx_2')$ in ascending order (Line 3). Therefore, the radio $R_{idx,1'}$ is more energy-efficient than $R_{idx,2'}$ without considering radio switching energy consumption $A_i$. The radio hardware conflict checker “Conflict()” gets the boolean information on whether there is a hardware conflict between the two radios which prevents them from being used simultaneously. Finally, RT-Select makes radio selection decisions based on three different cases:

1. if the more energy-efficient radio $R_{idx,1}$ can deliver all packets before the deadline by itself, RT-Select uses $R_{idx,1}$ alone and assigns all $N$ packets to it. (Line 4-5)

2. if none of the radios can deliver all packets before the deadline by itself, RT-Select attempts to use both radios. First, RT-Select assigns $X_{max, (idx_1, idx_2)}$ packets to $R_{idx,1'}$. Then, the remaining packets are assigned to the other radio if there is no hardware conflict between the two radios. (Line 6-10)

3. if only the less energy-efficient radio $R_{idx,2}$ can deliver all packets before the deadline, RT-Select needs to decide whether to use it alone or use both radios. In case $R_{idx,2}$ has the smaller $B_i$ of the two radios or
$X_{\text{max},(idx,1')}$ is smaller than $A_{idx,1'}/(B_{idx,2} - B_{idx,1'})$ \(^5\), RT-Select uses the less energy-efficient radio $R_{idx,2}$ alone and assigns all $N$ packets to it. If there exists a hardware conflict between the two radios, $R_{idx,2}$ is also used alone to avoid the conflict. Otherwise, RT-Select selects both radios and assigns $X_{\text{max},(idx,1')}$ packets to $R_{idx,1'}$ and the remaining packets to the other radio. (Line 12-17)

3.2. RT-Select-General Algorithm for Selection among Multiple Radios

Based on the insights collected during our algorithm design for the two-radio special case, we design RT-Select-General that solves the general form of the problem involving $m$ radios. As shown in Algorithm 2, RT-Select-General first computes the $A_i$, $B_i$, and $X_{\text{max},i}$ values for all $m$ radios (Line 1). Similar to RT-Select, RT-Select-General sorts all $m$ radios based on the energy consumption to transmit $N$ packets for each single radio $(A_i + B_i \times N)$ and stores the sorted radio indices to $(idx,1,...,idx,m)$ in ascending order (Line 2). RT-Select-General sorts all radios again based on the average energy consumption per packet $B_i$ without considering radio switching energy consumption $A_i$ and stores the radio indices to $(idx,1',...,idx,m')$ in ascending order (Line 3). The radio hardware conflict checker “Conflict($R_x$, $R_y$)” gets the boolean information on whether there is a hardware conflict between the radio $R_x$ and any radio in $R_y$, where $R_y$ is a set that consists of one or more radios.

RT-Select-General makes radio selection decisions based on three cases similar to RT-Select:

1. if the most energy-efficient radio $R_{idx,1}$ can deliver all packets before the deadline by itself, RT-Select-General uses it alone and assigns all $N$ packets to it. (Line 4-5)

2. if none of the radios can deliver all packets before the deadline by itself, RT-Select-General has to use multiple radios. Similar to RT-Select, RT-Select-General prefers to use the radios with small $B_i$s, thus it selects the radios one by one based on the sorted indices $(idx,1',...,idx,m')$ and lets them transmit with their maximum capacity until the selected radios can deliver all $N$ packets before the deadline. If there exists a radio hardware conflict between $R_{idx,i'}$ and any radio $R_k$ which has already been selected ($X_k > 0$), the radio $R_{idx,i'}$ is skipped to avoid the conflict. (Line 6-17)

---

\(^5\)This comparison decides whether it consumes less energy to use the less energy-efficient radio alone. The equation comes from the optimality proof in [1].
Algorithm 2: RT-Select-General

Input: $N, D, RC_1, RC_2, ..., RC_m$

Output: $X_1, X_2, ..., X_m$

1. Compute $\{A_i, B_i, X_{\max,i} \mid i = 1, ..., m\}$;
2. $(idx_1, ..., idx_m) = \text{sort}\{A_i + B_i \times N \mid i = 1, ..., m\}$;
3. $(idx_1', ..., idx_m') = \text{sort}\{B_i \mid i = 1, ..., m\}$;
4. if $X_{\max,idx_1} \geq N$ then
   5. $X_{idx,1} \gets N$
   6. else if $\max\{X_{\max,idx,i} \mid i = 1, ..., m\} < N$ then
      7. for $i = 1$ to $m$ do
         8. if $\text{Conflict}(R_{idx,i'}, \{R_k \mid X_k > 0\})$ then
            9. continue;
         10. end
         11. if $X_{\max,idx,i'} < N - \sum\{X_{idx,k} \mid k < i\}$ then
            12. $X_{idx,i'} \gets X_{\max,idx,i'}$
         13. else
            14. $X_{idx,i'} \gets N - \sum\{X_{idx,k} \mid k < i\}$
            15. break;
         16. end
      7. end
   6. else
      7. for $i = 2$ to $m$ do
         8. if $X_{\max,idx,i} < N$ then
            9. continue;
         10. end
         11. if $B_{idx,i} = B_{idx,i'}$ or $A_{idx,1'}/(B_{idx,i} - B_{idx,i'}) > X_{\max,idx,1'}$
            or $\text{Conflict}(R_{idx,i}, R_{idx,i'})$ then
            12. $X_{idx,i} \gets N$
         13. else
            14. $X_{idx,i'} \gets X_{\max,idx,i'}$
            15. $X_{idx,i} \gets N - X_{idx,i'}$
         16. end
         17. break;
      7. end
   6. end
5. end
3. if there exists a radio $R_{idx,i}$ which can deliver all packets before the deadline by itself but is not the most energy-efficient one ($i > 1$), then RT-Select-General needs to decide whether to use it alone or combine it with another radio\(^6\). Inspired by Algorithm 1, we consider the radio $R_{idx,i'}$ (the one with the smallest $B_i$ of all radios) for the possible combination with $R_{idx,i}$. If $R_{idx,i}$ has the smallest $B_i$ or $X_{max,i}(idx,i')$ is smaller than $A_{idx,i'}/(B_{idx,i} - B_{idx,i'})$, RT-Select-General selects $R_{idx,i}$ only and assigns all packets to it. If there exists a hardware conflict between $R_{idx,i}$ and $R_{idx,i'}$, $R_{idx,i}$ is also selected to be used alone. Otherwise, RT-Select-General combines $R_{idx,i}$ with $R_{idx,i'}$ and let $R_{idx,i'}$ transmit with its maximum capacity and assigns the remaining packets to $R_{idx,i}$. (Line 19-30)

The constraints reflecting the hardware conflicts can be added into case 2) and case 3) of Algorithm 1 and Algorithm 2. RT-Select-General behaves identically to RT-Select when $m = 2$, making the latter a special case providing optimal selections. The time complexity of RT-Select-General is $O(m \log m)$ (dominated by the complexity of sorting), which is acceptable to support real-time decision-making since $m$ is not expected to be very large in practice ($m \leq 16$ today to our knowledge).

3.3. RT-Balance Algorithm for Runtime Traffic Balancing

As discussed in Section 3.1 and Section 3.2, RT-Select and RT-Select-General are designed to ensure that all packets can be delivered to their destination by the deadline if they can find feasible radio selection and data partitioning solutions with the assumption that the actual runtime throughput follows the predicted value $TH_i$. In reality, there does not exist any throughput predictor which achieves 100% prediction accuracy. To study the impact of inaccurate throughput prediction, we perform an empirical study. We use Holt-Winter predictor [12], one of the most effective time series forecasting algorithms, to predict throughput based on historical measurements, run RT-Select to select radios and partition the traffic, and record the deadline misses. We observe that a deadline miss occurs when the traffic assigned to the radio $R_i$ is close to its maximum packet delivery capacity $X_{max,i}$ and the actual throughput of the radio $R_i$ is smaller than the predicted value in that period. Figure 1 plots the throughput prediction errors when both the WiFi and ZigBee radios are selected by RT-Select to

\(^6\)We select at most two radios in this case in consideration of designing a light-weight algorithm for runtime use.
transmit 500 packets (64KB data) with a deadline of 0.8s. Based on line 7-8 in Algorithm 1, the traffic assigned to the WiFi radio has about 478 packets, which is very close to the WiFi radio’s capacity, while only about 22 packets are assigned to the ZigBee radio. As Figure 1(a) shows, the packet deliveries through the WiFi link miss the deadline in three periods (45s, 95s, and 100s), when the actual throughput measurements are smaller than the predictions by at least 30 packets/s. Figure 1(b) shows that the packet deliveries through the ZigBee link always meet the deadline because the traffic assigned to the ZigBee radio is far below its capacity. From the results, we can see that the deadline misses occur when the traffic assigned to a radio is very close to its capacity.

To address this issue, we reserve a small portion of the predicted throughput (e.g., 5%) as a guard space, compute $X_{max,i}$ based on the rest (e.g., 95%), and design a runtime algorithm, namely RT-Balance, which balances the traffic assigned to different radios. Algorithm 3 shows the RT-Balance algorithm. When facing tight deadlines, RT-Balance creates a process for

Algorithm 3: RT-Balance

<table>
<thead>
<tr>
<th>Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Compute ${X_{max,i}</td>
</tr>
<tr>
<td>2</td>
<td>if $\sum_{i=1}^{m} X_{max,i} &gt; N$ then</td>
</tr>
<tr>
<td>3</td>
<td>goto RT-Select(-General) ;</td>
</tr>
<tr>
<td>4</td>
<td>end</td>
</tr>
<tr>
<td>5</td>
<td>for $i = 1$ to $m$ do</td>
</tr>
<tr>
<td>6</td>
<td>if $fork() &gt; 0$ then</td>
</tr>
<tr>
<td>7</td>
<td>continue ;</td>
</tr>
<tr>
<td>8</td>
<td>end</td>
</tr>
<tr>
<td>9</td>
<td>while seq &lt; N do</td>
</tr>
<tr>
<td>10</td>
<td>if isReady $(R_i)$ then</td>
</tr>
<tr>
<td>11</td>
<td>$Tx (R_i, ++seq)$ ;</td>
</tr>
<tr>
<td>12</td>
<td>end</td>
</tr>
<tr>
<td>13</td>
<td>if $time() &gt; D$ then</td>
</tr>
<tr>
<td>14</td>
<td>return FAIL ;</td>
</tr>
<tr>
<td>15</td>
<td>end</td>
</tr>
<tr>
<td>16</td>
<td>end</td>
</tr>
<tr>
<td>17</td>
<td>return OK ;</td>
</tr>
<tr>
<td>18</td>
<td>end</td>
</tr>
</tbody>
</table>
each radio that repeatedly transmits a packet when it is ready (Line 5-18). In this way, RT-Balance minimizes the latency to meet the deadline and achieves natural load balance among the radios. Specifically, a global variable “seq”, storing the sequence number of the current packet assigned for transmission, is shared by all processes and initialized as 0. Algorithm 3 first computes the packet delivery capacity \( X_{max,i} \) of each radio \( R_i \) (Line 1), where only the radios without hardware conflict are considered. Then, if the sum of all radios’ packet delivery capacities is larger than the traffic demand, RT-Select or RT-Select-General is used to select radios and partition data (Line 2-3). Otherwise, the load balancing is invoked and \( m \) child processes are created for the \( m \) radios using “fork()” (Line 5-7). Each child process uses a loop to request packets for transmission until all packets have been assigned. If there is any unassigned packet and the radio \( R_i \) is ready to transmit, seq is incremented to be the sequence number of a new packet, which is assigned to the radio \( R_i \) for transmission (Line 9-11). The time that has passed since the program starts is checked in each loop. If the deadline has passed before all packets have been transmitted, the child process terminates and indicates that the deadline has been missed (Line 13-14). Otherwise, the child process finishes after all transmission is complete (Line 17).

4. System Design and Implementation

To realize our designs, we develop the RRaSB system that makes multiple radios available at runtime and allows dynamic radio switching and bundling among them. Figure 2 shows the system architecture. The radio characteristics including energy consumption of radio switching \( (E_{sw}) \), radio switching time \( (T_{sw}) \), power consumption when the radio is idle \( (P_{rb}) \), and average energy consumption per transmission attempt \( (E_{ta}) \) are measured offline and stored in the Radio Characteristics component, serving as inputs to the radio selection algorithm. The Throughput Predictor predicts the throughput in the next period based on the historical data and the Link Quality Predictor estimates the expected transmission counts (ETX) in the next period based on previous ETX measurements using the Holt-Winters method. If a radio has not been used for a long time, Link Quality Predictor transmits some probing packets through it to keep its link quality measurements updated. The Radio Selection Engine takes radio characteristics, estimated throughput and ETX, and traffic demand and deadline specified by the application as inputs and runs the radio selection algorithm to select the radio(s) that is/are best suited for the current
network traffic and operating conditions and then assigns packets accordingly. Multiple **Radio Controller** modules exist in RRaSB. Each Radio Controller controls the on/off state of a radio based on the decision made by the Radio Selection Engine and measures the actual throughput and ETX fed into the Throughput Predictor and Link Quality Predictor, respectively. RRaSB is configured to perform the radio selection in each period based on the measured throughput and ETX of the radio links as well as the traffic demand and deadline specified by the benchmark application. If the current radio selection is found to be the best-suited, it is retained; otherwise, our system switches to a new best-suited setting. Radios are turned off after the last transmission in each period if they are not selected for use in the next period and the unselected ones are kept off to reduce energy consumption. If multiple transmitters exist, they access the channel in a TDMA fashion. We have implemented RRaSB in Raspbian Linux [13] and Contiki [14] and two prototypes: one with two radios and the other with five radios. A power monitor from Monsoon Solutions [15] is connected to the sender to measure the energy consumption. More implementation details can be found in [1].

5. Evaluation

To examine the efficacy of our radio selection and traffic partitioning solution, we perform a series of experiments on our embedded platform presented in Section 4. We start by demonstrating the time efficiency of RT-Select-General and the effectiveness of the throughput and link quality predictors. We then run experiments to measure the radio energy consumption.
and deadline miss ratio with our prototype hosting two radios and repeat the experiments with five radios. We compare our approaches against two baselines: GreenBag using GB-E configuration [16] and GLPK (GNU Linear Programming Kit) [17]. GreenBag is a practical state-of-the-art radio selection approach designed for real-time applications. GreenBag supports multi-radio mode and single-radio mode under GB-E and GB-P configurations. In multi-radio mode, GreenBag seeks to minimize the transmission time by balancing the load on multiple radios based on link throughput prediction, while the most energy-efficient radio is selected in single-radio mode. GB-E chooses single-radio mode to reduce the energy consumption and switches to multi-radio mode when the bandwidth is insufficient, while GB-P uses multi-radio mode only. GLPK provides the optimal results to the ILP problems. Please note that GLPK cannot be used for real-time applications with short deadlines because of its heavy computation overhead as presented in Section 5.1. We run GLPK offline and exclude its energy consumption in the results of optimal solutions (Figure 8(a) and 9(a)).

In all experiments, we deploy two real-time benchmark applications on top of our system which generate data packets periodically. The first benchmark application (benchmark application A) emulates a health care scenario where doctors use smart glasses to take ambient pictures or videos of patients and send them to the hospital information system for real-time face detection and recognition [2]. In this application, a fixed traffic demand is employed by the smart glasses but the application may specify different deadlines based on its quality of service (QoS) needs. The second benchmark application (benchmark application B) emulates a real-time georeferencing scenario where UAVs capture images of the land from the air and transmit them together with GPS locations to a ground station [3]. In this application, a fixed deadline (e.g., 1 second) of image delivery is adopted by the UAVs to ensure the accuracy of the real-time location but the traffic demand (image size) may vary to meet different needs. Both benchmark applications generate periodic traffic whose deadline is equal to its period. The two benchmark applications allow us to examine the performance of our system (i) at a fixed data rate with different data delivery deadlines and (ii) at various data rates with a fixed deadline.

5.1. Time Efficiency of RT-Select-General

We first measure the execution time of RT-Select-General and two baseline approaches (GreenBag and GLPK) on the Raspberry Pi 3 with a 1.2 GHz 64-bit quad-core ARMv8 CPU. We measure the time duration between feeding the input into the Radio Selection Engine and receiving the output
from it. We repeat the experiments 20 times using random inputs for each $m$ (the number of radios). Figure 3 shows the average execution time of GreenBag, GLPK and RT-Select-General for different number of radios ($m$ ranging from 2 to 16) in the logarithmic scale. As Figure 3 shows, the average execution time of RT-Select-General increases from 4 $\mu$s to 26 $\mu$s when $m$ increases from 2 to 16, which is slightly ($2\sim17\mu$s) longer than what GreenBag uses. The average execution time of GLPK ranges from 6267 $\mu$s to 8670 $\mu$s, which is $336\sim1412$ times longer than what RT-Select-General consumes. Therefore, it is not feasible to use the time-consuming GLPK to support the real-time applications with short deadlines, especially when running on the platforms with limited hardware resources. As a comparison, our RT-Select-General can time-efficiently make decisions achieving performance close to what GLPK offers (see Section 5.4).

5.2. Effectiveness of Link Condition Predictors

We then perform a set of controlled experiment to evaluate the effectiveness of our Throughput Predictor and Link Quality Predictor employing the Holt-Winters method. In this set of experiments, we measure the throughput and ETX of radio links under controlled interference and compare them against the predicted values. Figure 4 plots the example traces showing the throughput and ETX changes of a WiFi link when encountering the controlled interference. An interferer begins the transmission in the same channel from the 31st second to the 100th second. As Figure 4 shows, the predictions are very close to the measurements during the process. The standard deviation on the throughput difference is 152 packets/s and 80% of the prediction errors are less than 125 packets/s. The standard deviation on the ETX difference is 0.25 and 80% of the prediction errors are less than 0.2.
5.3. Experiments with Two Radios

We run experiments on our prototype hosting two radios [1] (i.e., the CC2650 ZigBee radio and the RT5370 WiFi radio) to evaluate the effectiveness of RT-Select and its impact on radio energy consumption and real-time performance. Since the output of RT-Select is proved to be optimal, we only compare RT-Select against GreenBag in this set of experiments.

We configure the benchmark application A to transmit a 23KB image (480×480 JPEG) in every period and repeat the experiments with 12 different deadlines ranging from 0.60s to 1.04s according to the response time of Amazon face recognition applications [18]. Figure 5(a) shows the energy saving of RT-Select over GreenBag per period and Figure 5(b) plots the deadline miss ratio. RT-Select shows significant energy saving (ranging from 8mJ to 37mJ\(^7\)) when the deadline is greater than 0.64s with the deadline miss ratios no higher than 1%. The energy savings benefit from RT-Select’s decision on keeping only the WiFi radio active rather than using both radios suggested by GreenBag. High deadline miss ratios are observed under both RT-Select and GreenBag when the deadline is shorter than 0.68s, not enough to turn on the WiFi radio or send all packets using the ZigBee radio. The results show that RT-Select consistently outperforms GreenBag under various deadlines.

Similarly, we configure the benchmark application B to transmit a JPEG image with the fixed deadline (0.80s) in every period, and repeat the experiments with 12 image sizes ranging from 31KB (640×480 JPEG) to 108KB (1280×720 JPEG). As Figure 6(a) and Figure 6(b) show, RT-Select consumes 27~54mJ less energy compared to GreenBag without missing any

\(^7\)As a comparison for energy saving values, the CC2650 radio consumes 30mW power when transmitting at 5dBm [9].
(a) Energy saving over GreenBag.

(b) Comparison on deadline miss ratio.

Figure 6: Performance under RT-Select and GreenBag with two radios when the application transmits at different data rates with the same deadline.

deadline when the image size is between 31KB and 66KB. The energy savings benefit from RT-Select’s decision on keeping only the WiFi radio active rather than using both radios suggested by GreenBag. The energy saving is marginal when the image size is 73KB or 80KB. This is because both RT-Select and GreenBag decide to use only the WiFi radio when it becomes the more energy-efficient radio under high traffic demand and can deliver all data packets by the deadline. When the image size is 87KB, both RT-Select and GreenBag suggest using both radios. However, RT-Select assigns 94.6% of packets to the WiFi radio and 5.4% to the ZigBee radio and lets WiFi transmit for the entire period and ZigBee finish early, while GreenBag assigns 85.9% of packets to the WiFi radio and 14.1% to the ZigBee radio and lets both radios finish their transmissions at the same time, resulting RT-Select consumes 37mJ less energy than GreenBag. High deadline miss ratios are observed under both RT-Select and GreenBag when the image size is larger than 87KB, beyond the capacity of two radios with the consideration of radio switching overhead. The results show that RT-Select always provides the better radio selections on various data rates.

To evaluate the performance of RT-Balance, we configure the benchmark application A to transmit a fix sized image of 64KB with some tight deadlines ranging from 0.35s to 0.50s. Since the deadlines are very tight, both radios have to keep active for the entire period. As Figure 7(a) and 7(b) show, RT-Balance significantly reduces the deadline miss ratio by 34.5%, 48.9% and 21.7% compared to RT-Select when the deadlines are 0.40s, 0.45s and 0.50s, respectively. At these deadlines, RT-Balance only increases the energy consumption by 11mJ, 12mJ and 8mJ per period. The slight increase in energy consumption is in exchange for a proportionally much-larger reduction in deadline miss ratio. The reduction on the deadline miss ratio benefits from RT-Balance’s runtime traffic balancing between the two radios, in contrast to RT-Select and GreenBag which assign packets to each
5.4. Experiments with Five Radios

In this set of experiments, we examine the effectiveness of RT-Select-General with our prototype device hosting five radios [1]. We compare RT-Select-General against GreenBag and Optimal.

We first explore RT-Select-General’s performance under a fixed traffic demand with different deadline requirements. We configure the benchmark application A to transmit a 109KB image (1280×720 JPEG) in each period and repeat the experiments with 12 different deadlines ranging from 0.80s to 1.24s. Figure 8 shows the comparisons on radio energy consumption and deadline miss ratio under GreenBag, Optimal, and RT-Select-General, respectively. As Figure 8(a) and Figure 8(b) show, all three methods suggest using all radios to accommodate the tight deadlines (i.e., 0.80s and 0.84s). High deadline miss ratios are observed when the deadline is 0.80s, beyond the capacity of all five radios together when considering radio switching overhead. When the deadline is larger than 0.84s, RT-Select-General
achieves significant energy savings ranging from 308\,mJ to 436\,mJ compared to GreenBag with the deadline miss ratios no higher than 1%. RT-Select-General makes the optimal selections for all deadlines except 0.88\,s and 0.92\,s. In those two cases, RT-Select-General selects to use the BCM43438 radio as the secondary radio based on the sorting of $B_i$ (see Section 3.2), while Optimal decides to use the CC2420 radio instead.

We also evaluate RT-Select-General’s performance under various traffic demands with a fixed deadline. We configure the benchmark application B to transmit a JPEG image with a fixed deadline (1.44\,s) in each period and repeat the experiments with 12 different image sizes ranging from 109KB (1280×720 JPEG) to 433KB (1920×1080 JPEG). As Figure 9(a) shows, RT-Select-General consistently consumes less energy (298\,mJ on average) compared to GreenBag and performs close to what Optimal offers (30\,mJ difference on average). RT-Select-General provides optimal selections to nine cases among the 12 cases. Please note that high deadline miss ratios are observed under all three methods when the image size is 433KB, beyond the capacities of all radios operating simultaneously when considering radio switching overhead. We also perform trace-driven simulations and observe similar improvements at various combinations of traffic demand and deadline [1]. The results demonstrate the effectiveness of RT-Select-General in reducing the energy consumption, while meeting satisfactory real-time requirements.

To evaluate the performance of RT-Balance, we configure the benchmark application A to transmit a fix sized image of 128KB with tight deadlines ranging from 0.46\,s to 0.52\,s. Since the deadlines are very tight, all five radios have to keep active for the entire period. As Figure 10(a) and Figure 10(b) shows, RT-Balance significantly reduces the deadline miss ratio by 28.6% and 51.4% when the deadlines are 0.48\,s and 0.50\,s, respectively, while only increases the energy consumption by 19\,mJ and 22\,mJ per period com-
5.5. Large-scale Simulation Study

Relying on the radio characteristics measured on our platform with five radios, we also perform a large-scale simulation study to measure radio energy consumption and deadline miss ratio at various combinations of traffic demands and deadlines. The grey shaded areas denote the invalid combinations that the optimal deadline miss ratio is higher than 5%. The colors in each subfigure denote the percentages of more energy consumed than Optimal, i.e., $(E(\text{RT-Select-General}) - E(\text{Optimal}))/E(\text{Optimal})$ and $(E(\text{GreenBag}) - E(\text{Optimal}))/E(\text{Optimal})$, respectively.

pared to RT-Select-General. The slight increase in energy consumption is in exchange for a proportionally much-larger reduction in deadline miss ratio. The reduction on the deadline miss ratio benefits from RT-Balance’s runtime traffic balancing between the five radios, in contrast to RT-Select-General and GreenBag which assign packets to each radio before transmission based on throughput prediction. The deadline miss ratios are nearly 100% for all approaches when the deadline is 0.46s, which is too short for the five radios.
demands and deadlines. In this set of experiments, we uniformly select 200 image sizes ranging from 94KB (1280×720 JPEG) to 847KB (3840×2160 JPEG) and 200 deadline samples ranging from 0.8s to 2.6s and then simulate radio energy consumption of running Optimal, GreenBag, and RT-Select-General, respectively, under all valid combinations of traffic demands and deadlines (optimal deadline miss ratio no higher than 5%).

Figure 11(a) is a heat map plotting the energy consumption difference between RT-Select-General and Optimal and Figure 11(b) shows the difference between GreenBag and Optimal. The white areas of Figure 11(a) shows the cases (94.4% of deadline and image size combinations) where RT-Select-General makes the optimal radio selections and traffic partitions. GreenBag only makes the optimal decisions in 5.4% of combinations, as shown in Figure 11(b). The mean energy consumption difference between RT-Select-General and Optimal is 7.1%, while the difference between GreenBag and Optimal is 60.8%. The simulation results confirm that RT-Select-General can provide optimal selections to most cases and significantly outperforms GreenBag under various combinations of data rates and deadlines.

6. Related Work

Bandwidth aggregation for a device with multiple network interfaces has also been studied for years in the literature and many techniques are readily available [19]. For instance, multipath TCP (MPTCP) [20] is one of the most widely used techniques and now a new standardized transport protocol that allows a device to take advantage of data transfer through multiple network interfaces simultaneously. Those early efforts are not directly applicable to embedded wireless devices with power constraints, since they were not designed to provide energy-efficient wireless data transfers [21, 22].

There has been increasing interest in studying the energy-aware bundling or switching between WiFi and 3G/4G radios on smartphones. For instance, Bui et al. used WiFi and/or LTE to minimize playback halts due to the buffer underflow when a stored video is streamed to a smartphone [16]. There exists commercial software, e.g., VideoBee, Super Download Lite-Booster, MPTCP in iOS, and KT’s GiGA LTE, that supports concurrent use of WiFi and cellular radios. More recently, research efforts have begun to pay more attention to energy efficiency in the context of smartphones and IoT applications. For instance, Lim et al. [23] extended MPTCP to support energy-aware data transfers over WiFi and LTE radios. Nikraves et al. conducted a real-world study of multipath for mobile settings and developed a flexible software architecture to enhance the performance of
MPTCP on smartphones [21]. Nika et al. developed an energy model for smartphones to support energy-aware WiFi and LTE radio bundling [24]. Mu et al. developed a radio and transmission power selection system for IoT applications to meet their QoS requirements [25]. Wu et al. designed an energy-efficient WiFi and LTE bandwidth aggregation method for video services on mobile devices [26]. Gu et al. developed a low-power LoRa-based control plane bundled with a ZigBee-based data-plane network [27]. These existing approaches are either unaware of timing constraints or limited to mainly WiFi and 3G/4G on smartphone platforms, thus they are not directly applicable to support timely, energy-efficient data transfer using heterogeneous radios in various IoT embedded platforms.

For real-time wireless data deliveries, novel methods (e.g., [28, 29, 30]) have recently been explored to meet timing constraints via real-time MAC protocols, packet scheduling, and routing based on the centralized Time Division Multiple Access (TDMA) scheme. However, most of them consider neither energy efficiency nor heterogeneous radios. In contrast to these real-time approaches, our work aims to support stringent timing constraints with minimal energy consumption by effectively leveraging heterogeneous radios. Our work is therefore orthogonal and complementary.

7. Conclusion and Future Work

Heterogeneous radios are becoming increasingly available in modern embedded devices, offering new opportunities to use multiple wireless technologies energy-efficiently to accommodate the needs of real-time applications. This paper formulates the runtime radio switching and bundling for real-time IoT applications as an optimization problem and presents three algorithms which select radios and partition data at runtime to minimize the energy consumption for real-time data transfer. Experimental results show that the proposed solution can significantly reduce the radio energy consumption over the state of the art, while meeting the application specified traffic demand and deadline requirement.

Acknowledgment

This work was supported by the NSF through grant CRII-1657275 (NeTS) and CNS-1526932.
References


[10] IOT-GATE-iMX7 - Industrial Internet of Things Gateway [online].


[17] GNU Linear Programming Kit (GLPK) [online].

[18] Analyzing Performance for Amazon Rekognition Apps [online].


