Energy Efficient Sampling Policies for Edge Computing Feedback Systems

Vishnu Narayanan Moothedath, Jaya Prakash Champati, and James Gross

Abstract—We study the problem of finding efficient sampling policies in an edge-based feedback system, where sensor samples are offloaded to a back-end server that processes them and generates feedback to a user. Sampling the system at maximum frequency results in the detection of events of interest with minimum delay but incurs higher energy costs due to the communication and processing of redundant samples. On the other hand, lower sampling frequency results in higher delay in detecting the event, thus increasing the idle energy usage and degrading the quality of experience. We quantify this trade-off as a weighted function between the number of samples and the sampling interval. We solve the minimisation problem for exponential and Rayleigh distributions, for the random time to the event of interest. We prove the convexity of the objective functions by using novel techniques, which can be of independent interest elsewhere. We argue that adding an initial offset to the periodic sampling can further reduce the energy consumption and jointly compute the optimum offset and sampling interval. We apply our framework to two practically relevant applications and show energy savings of up to 36% when compared to an existing periodic scheme.

Index Terms—Energy minimisation, optimal sampling, edge computing, feedback systems, event detection, cyber physical systems, video analytics systems

1 Introduction

With the advent of next-generation mobile networks such as 5G Release 15 and 16, there is an increasing interest in realising various real-time services and applications [1]–[3]. Perhaps most prominently, this materialises with the Release 16 features of URLLC (ultra-reliable low latency communication) targeting sub-millisecond end-to-end delays primarily for industrial automation applications. However, in addition to these extreme use cases, a plethora of new applications are arising that all process states of reality and accurately provide feedback either to devices or humans. Examples of such feedback systems with low latency requirements are human-in-the-loop applications like augmented reality, wearable cognitive assistants (WCA), or ambient safety. Also, in the domain of cyber-physical systems (CPS), such applications are prominent, for example, in the context of automated fault detection by vibration analysis where the acoustic data is analysed to potentially trigger some emergency procedures [4]. All these applications have in common that feedback depends on state capture and timely processing, whereas essential state changes are random events and hence an efficient operation of the application becomes a central aspect of the system. This is even more emphasised by the recent trend to place most of the processing logic of such feedback systems with edge computing facilities, leveraging supposedly ubiquitous real-time compute capabilities with the additional costs of offloading compute tasks (in terms of communication delays and energy consumption).

In this paper, we study approaches that enable capturing the relevant system changes in edge-based feedback systems while striking a balance with the total energy consumption. We consider feedback systems that monitor a process via sampling, while only reacting to a sub-set of samples, referred to as *essential events*, that lead to a system change – for instance, a new augmentation towards a human user in a WCA, or an activation of emergency shutdown in a CPS. After an essential event is captured and processed (including the generation of feedback), the feedback system transits to the next state where it starts monitoring for an essential next event to happen. The trade-off that we study relates to the strategy applied to sample the process. More frequent sampling leads to a timely capture of the essential event. However, it also leads to the capture of unimportant samples of the process, wasting system resources in terms of energy, communication bandwidth, and computing cycles. We are interested in mathematically characterising this trade-off and studying the implied consequences.

Our key contributions are listed below.

- We pose the problem to find the energy-minimising periodic sampling interval of an edge-based feedback system as an optimisation problem. The convexity of this problem is proved for certain distributions by developing a novel approach that uses the Poisson sum formula and Fourier transforms – which can be of independent academic interest elsewhere. The problem is solved using a lightweight bisection algorithm that converges exponentially to the optimum.
- We prove that adding an initial offset to the periodic sampling further reduces the expected energy consumption when the TTE distribution is not exponential. Thus we pose a more generic optimisation problem in two variables and propose an algorithm to find a solution that achieves near-optimality.

- *Vishnu Narayanan Moothedath and James Gross are with the Department of Intelligent Systems, KTH Royal Institute of Technology, Malvinas Väg 10, Stockholm 11428, Sweden. E-mail: {vnmo,jamesgr}@kth.se*
- *Jaya Prakash Champati is with the Edge Networks Group, IMDEA Networks Institute, Avda. del Mar Mediterraneo 22, 28918 Leganes (Madrid), Spain. E-mail: jaya.champati@imdea.org*

- Using simulations, we study the energy reduction on systems with a wide range of parameterisations by considering a CPS and a video analytic service (VAS) use-case. Particularly for the VAS, we show the inefficiency of systems used in practice and the larger potential of energy optimisation where we obtain up to 40% increase in battery life. One reason for this is the larger communication content and power requirements for the VAS, which point to the increased relevance of the proposed solution in these future systems. We also observe that oversampling needs to be particularly avoided as it introduces a substantial increase in energy usage.

The paper is organised as follows. In Section 2, we go through the related research and explain the novelty in the contributions. In Section 3, we discuss the system model. In Sections 4 and 5, we lay down the general solution approach and solutions for exponential and Rayleigh distributed TTEs. We discuss the numerical results in Section 6 and conclude in Section 7.

2 Related Works

Some of the earliest ideas on detecting the relevant system changes (or essential events) come from control systems with applications in manufacturing and observable models [5]. The survey [6] looks at some important works that focus on failure events in complex systems. This is viewed from a statistical perspective in [7] and is later revisited in [8] where the authors look for stochastic changes in the system with an aim of quickest detection. Moving on to the recent research on edge computing and CPS, some important works use data mining and post-processing to detect events from the collected data [9]. We, on the other hand, are interested in detecting live events and generating corresponding feedback to dictate the process. Examples of such an event detection occur in real-time video analytics systems which are explored extensively in recent years from multiple perspectives [10], [11]. However, the aspect of energy consumption is not considered in the above works. Instead of using minimal energy, they look only at detecting the relevant system changes as fast as possible.

Wireless video surveillance, where the video frames are captured by sensors and are sent to the processing node over wireless sensor networks (WSN) is studied in [12]. The authors discuss the challenges faced by these systems including the energy consumption – that we are particularly interested in from a general event detection perspective. The authors discuss some of the adopted methods to reduce energy consumption. These methods include optimising sensor topologies [13], optimising video coding and transmission techniques [14], and forcing node cooperation between multiple sensors [15]. In a similar context of object detection and tracking, [16] discusses the energy saving by sending the camera to an idle state where the frames are dropped for a duration determined adaptively based on the speed of the object. Energy-efficient surveillance and tracking using a network of sensors are also discussed in [17], where only a subset of the sensors are activated at any given time. A different but widely studied method to save energy is offloading the sensor data [18], [19]. The drawback of an increased latency compounded on a large number of samples during an offloading is addressed (to a certain extent) by making offloading decisions [20] for the samples. This includes binary decisions [21]–[26], partial offloading decisions [27]–[29], and stochastic decisions [30].

In this paper, we focus on detecting the essential events of an edge-based feedback system in real-time in an energy-efficient manner. In contrast to all the above-mentioned works, our approach minimises energy consumption by reducing the data generated by the sensors, thereby reducing the total amount of data in the communication and processing pipeline. While some of the existing works reduce energy by lightly processing and filtering the samples before transmitting, or taking a cluster wise sensing decision, we look at statistically determining the optimum sensing time instances. In the control theory literature, an event is generally termed to have occurred when the sensor reading crosses a threshold. This is an example of an event-triggered control and the event is detected at the sensor [31]. Contrary to our model, these are not based on any further assumptions on remote processing and detection of the events, the size of the data being communicated to the remote node, or the necessity of feedback for the control.

While the works that focus on offloading tends to concentrate on the sensor side and take a hit at total energy consumption by shifting the energy usage to the edge device, our work provides a framework for reducing the total energy consumption in the system. To the best of our knowledge, our work is the first attempt to find the energy-optimal sampling points of an edge-based feedback system for capturing the relevant system changes.

For illustrating the performance of the proposed design, we focus on two relevant and practical systems. One commonality between these systems that make it particularly suited for our model is the necessity (or preference) of processing the samples at the edge, thereby taking out the possibility of approaches such as the state-of-the-art event-triggered control. The first system is a CPS that aims at fault detection, for instance, where the acoustic or vibration data from various metal parts (like bearings) are sensed and passed on to an edge device to jointly process and detect mechanical failures. Such systems are typically characterised by exponentially distributed failures, a relatively small amount of data transfer, and a low-power communication technology. A basic direction on how to approach the above-mentioned trade-off for a CPS is given by the authors in [32]. The second system is a video analytics system (VAS) motivated by a WCA system, where a human task progress is monitored continuously for the detection of task completion. The task completion time comprises of multiple system delays (communication, processing etc.) and a delay that is tied to the response time and skill of the human user. Previous works on WCA [33] and general distribution fitting indicates that this task completion time can be modelled as a Rayleigh distributed random variable. Other characteristics of a VAS are large data transfer and the requirement of a high throughput communication technology, owing to the continuous sequence of video frames that needs to be send to the processing node.

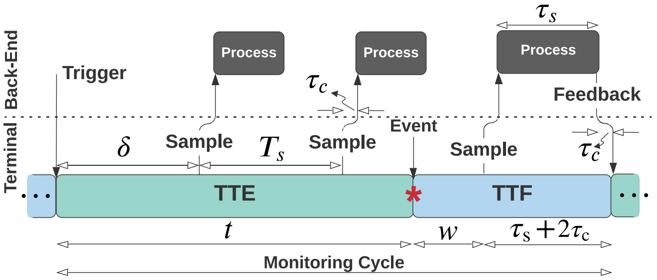


Fig. 1: Timing diagram of an arbitrary monitoring cycle.

$\mathcal{S}$	number of samples	$\mathcal{W}$	wait time
$\mathcal{T}$	time to event (TTE)	$\mathcal{T}_0$	idle time
$F_{\mathcal{X}}(\cdot)$	CDF X	$\bar{F}_{\mathcal{X}}(\cdot)$	CCDF of $\mathcal{X}$
T_s	sampling interval	δ	offset
τ_c	communication delay	τ_s	processing delay
P_c	communication power	P_0	idle power
E	energy	$\mathcal{E}$	energy penalty
$\mathbb{N}$	{0,1,2,. . . }	$\mathbb{N}^+$	{1,2,3,. . . }

TABLE 1: Table of notations.

3 System Model and Problem Statement

Consider a feedback system consisting of a mobile terminal (simply terminal) and a back-end server (simply back-end), that is designed to monitor a process through sampling. The terminal capture and send the samples to the back-end for processing. Immediately after the occurrence of an essential event (simply event), the process moves to an intermediate state where no more events are expected. The next sample drawn at or after this transition point – referred to as a *successful sample* – indicates the event detection at the back-end's processor, and causes feedback to the terminal. The reception of this feedback triggers the start of a fresh *monitoring cycle* to detect the next event. Within a monitoring cycle, the feedback is generated only to the successful sample, and all other samples are discarded by the back-end. The time taken from the start of a monitoring cycle to the event is termed as *Time to event (TTE)*, and the time between this event and the corresponding feedback is termed as *Time to feedback (TTF)*. We denote the TTE using the random variable $\mathcal{T}$, and a value for $\mathcal{T}$ is denoted by t. The timing diagram of the system is given in Fig. 1. Here we focus on modelling and designing the monitoring cycle corresponding to a single event, and the proposed design can be independently extended to all the (potentially predetermined set of) events that the system is supposed to detect.

For example, in a CPS, the event might correspond to the detection of a system failure. The feedback potentially triggers a reset and the terminal starts monitoring for the next failure. In the VAS (or WCA) system studied in [34], [35], the user is assigned to complete a set of tasks and the events correspond to the completion of each task. The terminal takes snapshots (video frames) of the user activity using a camera and sends them to the back-end for image processing.

The terminal can sample the process, transmit it to the back-end and can receive feedback whenever available. The back-end on the other hand is capable of receiving the sample, processing it and transmitting the feedback after an essential event detection. At all other times, the terminal and back-end go to their respective idle mode to save energy, and the power consumed while in this mode is denoted by P_0. Though the back-end can possibly serve and manage multiple terminals, we restrict our study to the back-end's interaction with a single terminal.

Sampling the system for detecting the event is governed by a set of sampling policies Π. We consider sampling policies $\pi \in \Pi$ that sample the system periodically with a sampling interval T_s, except for the first sample, which is sampled after waiting for a duration $\delta \geq T_\mathrm{s}$. We refer to δ as *offset*. This offset takes care of a minimum time threshold, sampling before which only adds to the energy wastage. We elaborate this in detail in Section 5. Apart from the set of sampling policies Π, we also consider a special case with $\delta = T_\mathrm{s}$, referred to as $\hat{\Pi}$. In other words, under Π the samples are taken at $t = kT_\mathrm{s} + \delta,\ k \in \mathbb{N}$ whereas under $\hat{\Pi}$ the samples are taken at $t = kT_\mathrm{s},\ k \in \mathbb{N}^+$. While the optimum policy under Π comprises of an optimum sampling interval T_s^* and an optimum phase δ^*, the optimum policy under the analytically simpler subset $\hat{\Pi}$ corresponds to (only) an optimum sampling interval $T_\mathrm{s}^\#$. We use the notation $(\cdot)^\#$ instead of $(\cdot)^*$ to have a clear distinction between the underlying set of policies. Following the same conventions, we refer to these optimum policies as π^* and $\pi^\#$, respectively.

The number of samples taken is denoted by a random variable $\mathcal{S}$, and a value for $\mathcal{S}$ is denoted by s. This includes both the discarded samples taken during TTE as well as the final sample that leads to detecting the event. This TTF consists of a (potential) random lag $\mathcal{W}$ until the next sample is drawn as well as a deterministic processing and two-way communication delays. The random lag $\mathcal{W}$ is referred to as *wait* and w denotes a value of $\mathcal{W}$. Note that only the processing and communication delay corresponding to the final sample contributes to the delay after the wait. In this work, we assume that the total power consumption during transmission and reception is the same at both the terminal and the back-end and we refer to it simply as communication power denoted by P_c, which is typically much larger than the idle power P_0. We also assume that the communication delay in either direction is τ_c and the processing time of the successful sample be τ_s.

When not performing a transmission, reception, or processing, both the terminal and the back-end enter idle mode, which incurs an idle power consumption. The total time within a state, during which the terminal or back-end is in an idle mode is referred to as the *idle time* and is denoted by the random variable $\mathcal{T}_0$. The notations used and their meanings are reiterated in TABLE 1 for readability.

3.1 Problem Statement

An ideal sampling policy should sample the system immediately after the event so that the wait is zero and the number of samples required per event is exactly 1. However, such a sampling policy is unattainable given the fact the TTE is randomly distributed. As a result, we have to settle with a sampling policy that simultaneously reduces the expected wait and expected number of discarded samples to yield the maximum attainable benefit. Note that this is not straightforward as the wait time and the number of samples shows opposite behaviour with a change in sampling frequency. For instance, an aggressive sampling reduces the wait time, but it also increases the number of samples. In this work, we use energy as a metric to quantify this opposing behaviour. Each sample warrants energy in terms of communication and processing; and during the wait w, energy is expended as governed by the idle power. It is this optimal trade-off of energy usage between the sampling frequency and the number of samples that we seek to quantify.

Note that the fundamental random variable here is the TTE $\mathcal{T}$ and the other random variables – $\mathcal{S}, \mathcal{W}$ and $\mathcal{T}_0$ – are derived from $\mathcal{T}$ through the selection of the optimization variables T_s and δ. We can compute the idle time denoted by $\mathcal{T}_0$ in terms of other parameters as follows:

$$\mathcal{T}_0 = \mathcal{T} + \mathcal{W} + \tau_\mathrm{s} - (\mathcal{S} - 1)\tau_\mathrm{c} \ . \tag{1}$$

Let E be the energy required for detecting one event.

$$\begin{aligned} \mathrm{E} &= (\mathcal{S}+1)\tau_\mathrm{c} P_\mathrm{c} + \mathcal{T}_0 P_0 \\ &= \mathcal{S}\tau_\mathrm{c}(P_\mathrm{c} - P_0) + \mathcal{W}P_0 + (\mathcal{T} + \tau_\mathrm{c} + \tau_\mathrm{s})P_0 + \tau_\mathrm{c} P_\mathrm{c} \ . \end{aligned} \tag{2}$$

We aim to minimise the expected energy for a given TTE statistics. In (2), the terms except those containing the number of samples $\mathcal{S}$ or the wait $\mathcal{W}$ are either constants or have constant expectations for a fixed distribution of $\mathcal{T}$. Hence these terms are irrelevant in the optimisation where we minimise the expected energy. Let $\mathcal{E}(T_\mathrm{s}, \delta)$ be the component of the total energy which is relevant for the optimisation and let us call it *energy penalty*. We obtain

$$\mathcal{E}(T_\mathrm{s}, \delta) = \alpha \mathbb{E}[\mathcal{S}] + \beta \mathbb{E}[\mathcal{W}] \ , \tag{3}$$

where we use the constants α and β for mathematical tractability in the upcoming sections. Here,

$$\alpha = \tau_\mathrm{c}(P_\mathrm{c} - P_0) \ \text{ and } \ \beta = P_0 \ . \tag{4}$$

In (3), $\beta\,\mathbb{E}[\mathcal{W}]$ corresponds to the additional energy expended for waiting, and α represents the energy wasted per discarded sample due to the additional communication and processing. The minimum values of $\mathcal{E}$ under the policies Π and $\hat{\Pi}$ are referred to as $\mathcal{E}^*$ and $\mathcal{E}^\#$, respectively. That is

$$\mathcal{E}^* = \mathcal{E}(T_\mathrm{s}^*, \delta^*) \text{ and } \mathcal{E}^\# = \mathcal{E}(T_\mathrm{s}^\#, T_\mathrm{s}^\#) \ .$$

We study the general optimisation problems $\mathcal{P}$ and $\hat{\mathcal{P}}$ under the set of sampling policies Π and $\hat{\Pi}$, respectively. These optimisation problems are defined as follows:

$$\begin{aligned} \mathcal{P} &: \ \pi^* = \{T_\mathrm{s}^*, \delta^*\} = \underset{\{T_\mathrm{s}, \delta\}}{\arg\min}\, \mathcal{E}(T_\mathrm{s}, \delta) \\ \hat{\mathcal{P}} &: \ \pi^\# = \{T_\mathrm{s}^\#\} = \underset{T_\mathrm{s}}{\arg\min}\, \mathcal{E}(T_\mathrm{s}, T_\mathrm{s}) \ , \end{aligned} \tag{5}$$

where π^* and $\pi^\#$ are the optimum policies for $\mathcal{P}$ and $\hat{\mathcal{P}}$, respectively. In practice, given the distribution of the TTE, solutions to these optimisation problems are computed one time prior to starting the sampling. The computations can be done at the sensor side - if it is capable of it - or at the back-end as part of the admission control (initial setup). As a result of this one time apriori computation, there is no signalling overhead demanded by the proposed solution. In what follows, we will solve these problems for the CPS and the VAS where the TTEs follow exponential and Rayleigh distributions, respectively.

4 Solution to the Problem $\hat{\mathcal{P}}$

Being the simpler one, we start with the optimisation problem $\hat{\mathcal{P}}$ in this section and find the optimum periodic sampling interval by minimising $\mathcal{E}(T_\mathrm{s}, \delta)$ given in (3). Throughout this section, we will drop the redundant second argument δ ($=T_\mathrm{s}$ under $\hat{\mathcal{P}}$) from the energy penalty for simplicity. As the solution is specific to the distribution of $\mathcal{T}$, we will first lay down the solution approach to finding the energy penalty for any general distribution and later apply this to particular distributions.

4.1 General TTE Distribution

Here, we derive the expressions for the expected number of samples and the expected wait time under a general TTE distribution, which together constitutes $\mathcal{E}(T_\mathrm{s})$.

4.1.1 Expected number of samples

Recall that $\mathcal{S}$ is the random number of samples taken for detecting an event. We have

$$\begin{aligned} \mathbb{P}(\mathcal{S} = k) &= \mathbb{P}\big(\lceil t/T_\mathrm{s} \rceil = k\big), \ \forall k \geq 1 \\ &= \mathbb{P}(k - 1 < t/T_\mathrm{s} \leq k) \\ &= F_\mathcal{T}(kT_\mathrm{s}) - F_\mathcal{T}\big((k-1)T_\mathrm{s}\big) \\ \Rightarrow \mathbb{E}[\mathcal{S}] &= \sum_{k=1}^{\infty} k\Big(F_\mathcal{T}(kT_\mathrm{s}) - F_\mathcal{T}\big((k-1)T_\mathrm{s}\big)\Big) \ . \end{aligned} \tag{6}$$

4.1.2 Expected wait time

Recall that w denotes a value of the random wait time $\mathcal{W}$. We have a fixed set of sampling instants governed by T_s. The CDF of the wait $F_\mathcal{W}(w)$ can be obtained by taking the probability of the TTE to fall at most w short of any sampling instant. Even though the TTEs are finite in practice, we consider they can be arbitrarily large for the sake of generalised analysis. As a result, a successful sample can be located anywhere from the first sampling instant to possibly infinity. For real systems, however, there is an upper bound for the TTE (preemption or otherwise) and it implies that the corresponding probability is zero beyond this point. We can compute the CDF of $\mathcal{W}$ as follows.

$$\begin{aligned} F_\mathcal{W}(w) &= \sum_{k=1}^{\infty} \mathbb{P}\big(kT_\mathrm{s} - w < t \leq kT_\mathrm{s}\big) \\ &= \sum_{k=1}^{\infty} \big(F_\mathcal{T}(kT_\mathrm{s}) - F_\mathcal{T}(kT_\mathrm{s} - w)\big) \ . \end{aligned} \tag{7}$$

Since $\mathcal{W}$ is a non-negative random variable, we have

$$\begin{aligned} \mathbb{E}[\mathcal{W}] &= \int_0^\infty \big(1 - F_\mathcal{W}(w)\big)\,\mathrm{d}w \\ &= \int_0^{T_\mathrm{s}} \Big(1 - \sum_{k=1}^{\infty} \big(F_\mathcal{T}(kT_\mathrm{s}) - F_\mathcal{T}(kT_\mathrm{s} - w)\big)\Big)\,\mathrm{d}w \ . \end{aligned}$$

Here, integral limits and infinite sum are finite, and the summand is non-negative. Thus

$$\mathbb{E}[\mathcal{W}] = T_\mathrm{s} - \sum_{k=1}^{\infty} \int_0^{T_\mathrm{s}} \big(F_\mathcal{T}(kT_\mathrm{s}) - F_\mathcal{T}(kT_\mathrm{s} - w)\big)\,\mathrm{d}w \ . \tag{8}$$

Using (6) and (8), we can compute the energy penalty given in (3). Clearly, $\mathbb{E}[\mathcal{S}]$ decrease with an increase in T_s.

We can see that the summation containing the integral in the expected wait time expression also decrease with an increase in T_{s}, thus increasing $\mathbb{E}[\mathcal{W}]$. This opposing behaviour of the two penalties result in a minima in their weighted sum $\mathcal{E}(T_{\mathrm{s}})$ (with weights α and β) at the optimum sampling interval $T_{\mathrm{s}}^{\#}$. Depending on the distribution of $\mathcal{T}$ and the resultant energy penalty, $T_{\mathrm{s}}^{\#}$ can be computed using known optimisation techniques or numerical solvers [36].

4.2 Exponentially Distributed TTE

In this subsection, we look into a feedback system where the TTEs are exponentially distributed. This part of the work describing about the solution of $\hat{\mathcal{P}}$ under an exponentially distributed TTE was published earlier by the authors in [32].

Lemma 1. *For exponentially distributed TTE with mean* $1/\lambda$, *the energy penalty* $\mathcal{E}(T_s)$ *imparted by a periodic sampling policy with a period* T_s *is given by*

$$\mathcal{E}(T_s) = \frac{\alpha\lambda + \beta\big(e^{-\lambda T_s} + \lambda T_s - 1\big)}{\lambda(1 - e^{-\lambda T_s})} . \tag{9}$$

Proof. Given that the TTEs are exponentially distributed with rate parameter λ, we have $F_{\mathcal{T}}(t) = 1 - e^{-\lambda t}$. From (6), the expected number of samples can be computed as

$$\mathbb{E}[\mathcal{S}] = \frac{1}{1 - e^{-\lambda T_{\mathrm{s}}}} . \tag{10}$$

Substituting $F_{\mathcal{T}}(t)$ in (7)

$$\begin{aligned} F_{\mathcal{W}}(w) &= \sum_{k=1}^{\infty} \big((1 - e^{-\lambda k T_{\mathrm{s}}}) - (1 - e^{-\lambda(kT_{\mathrm{s}} - w)})\big) \\ &= \sum_{k=1}^{\infty} e^{-\lambda k T_{\mathrm{s}}}(e^{\lambda w} - 1) = \frac{e^{\lambda w} - 1}{e^{\lambda T_{\mathrm{s}}} - 1} \\ \Rightarrow f_{\mathcal{W}}(w) &= \frac{\lambda e^{\lambda w}}{e^{\lambda T_{\mathrm{s}}} - 1} \\ \Rightarrow \mathbb{E}[\mathcal{W}] &= \frac{e^{-\lambda T_{\mathrm{s}}} + \lambda T_{\mathrm{s}} - 1}{\lambda(1 - e^{-\lambda T_{\mathrm{s}}})} . \end{aligned} \tag{11}$$

Substituting (10) and (11) in (5), we get the energy penalty $\mathcal{E}(T_{\mathrm{s}})$ for using a particular sampling interval T_{s} as

$$\mathcal{E}(T_{\mathrm{s}}) = \frac{\alpha\lambda + \beta\big(e^{-\lambda T_{\mathrm{s}}} + \lambda T_{\mathrm{s}} - 1\big)}{\lambda(1 - e^{-\lambda T_{\mathrm{s}}})} .$$

□

Lemma 2. *The energy penalty* $\mathcal{E}(T_s)$ *is convex in* T_s .

Proof. In the following, we drop the argument T_{s} from $\mathcal{E}(T_{\mathrm{s}})$ for simplicity in presentation. Note from (4) that $\beta > 0$.

$$\begin{aligned} \mathcal{E}' &:= \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}T_{\mathrm{s}}} = \frac{\beta\big(1 - e^{-\lambda T_{\mathrm{s}}}(\lambda T_{\mathrm{s}} + \frac{\alpha}{\beta}\lambda + 1)\big)}{(1 - e^{-\lambda T_{\mathrm{s}}})^2} , \\ \mathcal{E}'' &:= \frac{\mathrm{d}^2\mathcal{E}}{\mathrm{d}T_{\mathrm{s}}^2} = \frac{\beta\lambda e^{-\lambda T_{\mathrm{s}}}\big((1 + e^{-\lambda T_{\mathrm{s}}})(\lambda T_{\mathrm{s}} + \frac{\alpha}{\beta}\lambda) + 2e^{-\lambda T_{\mathrm{s}}} - 2\big)}{(1 - e^{-\lambda T_{\mathrm{s}}})^3} . \end{aligned}$$

$$\begin{aligned} \mathcal{E}'' \geq 0 \Rightarrow \tilde{\mathcal{E}} &:= (1 + e^{-\lambda T_{\mathrm{s}}})(\lambda T_{\mathrm{s}} + \tfrac{\alpha}{\beta}\lambda) + 2e^{-\lambda T_{\mathrm{s}}} - 2 \geq 0 , \\ \tilde{\mathcal{E}}' &:= \frac{\mathrm{d}\tilde{\mathcal{E}}}{\mathrm{d}T_{\mathrm{s}}} = \lambda(1 - e^{-\lambda T_{\mathrm{s}}}(\lambda T_{\mathrm{s}} + \tfrac{\alpha}{\beta}\lambda + 1)) , \\ \tilde{\mathcal{E}}'' &:= \frac{\mathrm{d}^2\tilde{\mathcal{E}}}{\mathrm{d}T_{\mathrm{s}}^2} = \lambda^2 e^{-\lambda T_{\mathrm{s}}}(\lambda T_{\mathrm{s}} + \tfrac{\alpha}{\beta}\lambda) . \end{aligned}$$

From the above expressions, since $\tilde{\mathcal{E}}'' \geq 0 \ \forall T_{\mathrm{s}}$, we can conclude that $\tilde{\mathcal{E}}$ is globally convex and any infimum point is its minimum. To find this infimum:

$$\tilde{\mathcal{E}}' = 0 \Rightarrow \lambda T_{\mathrm{s}} + \tfrac{\alpha}{\beta}\lambda = e^{\lambda T_{\mathrm{s}}} - 1 .$$

Substituting in the above expression for $\tilde{\mathcal{E}}$, we obtain

$$\begin{aligned} \min_{T_{\mathrm{s}} > 0}\{\tilde{\mathcal{E}}\} =&(1 + e^{-\lambda T_{\mathrm{s}}})(e^{\lambda T_{\mathrm{s}}} - 1) + 2e^{-\lambda T_{\mathrm{s}}} - 2 \\ =&e^{\lambda T_{\mathrm{s}}} + e^{-\lambda T_{\mathrm{s}}} - 2 \\ \geq&0 \ \forall T_{\mathrm{s}} > 0 \\ \Rightarrow \tilde{\mathcal{E}} \geq&0 \ \forall T_{\mathrm{s}} > 0 \\ \Rightarrow \mathcal{E}'' \geq&0 \ \forall T_{\mathrm{s}} > 0 . \end{aligned}$$

As T_{s} is non-negative, the energy penalty $\mathcal{E}$ is convex. □

Proposition 1. *The optimum sampling interval* $T_s^{\#}$ *under an exponentially distributed TTE with mean* $1/\lambda$ *is the solution to the expression*

$$e^{\lambda T_s^{\#}} - \lambda T_s^{\#} = \frac{\alpha}{\beta}\lambda + 1 . \tag{12}$$

Proof. The proof follows from Lemma 1 and Lemma 2. We can find the optimum by equating the first derivative of total penalty to zero.

$$\begin{aligned} &\frac{\mathrm{d}\mathcal{E}(T_{\mathrm{s}})}{\mathrm{d}T_{\mathrm{s}}} = 0 \Rightarrow 1 - e^{-\lambda T_{\mathrm{s}}}(\lambda T_{\mathrm{s}} + \tfrac{\alpha}{\beta}\lambda + 1) = 0 \\ &\Rightarrow e^{\lambda T_{\mathrm{s}}} - \lambda T_{\mathrm{s}} = \tfrac{\alpha}{\beta}\lambda + 1 . \end{aligned}$$

□

We know that $e^x - x$ is a monotonically increasing convex function in x with $e^x - x \geq 1, \ \forall x \geq 0$. Hence, this single variable expression can be solved using well-known numerical solvers.

Though the value of energy penalty depends on the values of β and α, the optimum sampling interval $T_{\mathrm{s}}^{\#}$ only depends on their ratio. Furthermore, from (4), it can be seen that this ratio β/α does not depend on the individual power figures but only on the percentage additional power necessary for communication or processing when compared to their respective idle power requirement. In other words, for fixed λ, β/α and thus the optimum sampling interval is a function of only τ_{c} and P_{c}/P_0.

4.3 Rayleigh Distributed TTE

We will now consider a Rayleigh distributed TTE with mean $\mu = \sigma\sqrt{\frac{\pi}{2}}$ and CDF $F_{\mathcal{T}}(t) = 1 - e^{-t^2/2\sigma^2}$; where σ is the scale parameter.

Lemma 3. *The expected number of samples is given by*

$$\mathbb{E}[\mathcal{S}] = \sum_{k=0}^{\infty} e^{-k^2 T_s^2/2\sigma^2} . \tag{13}$$

Proof. Substituting the Rayleigh CDF in (6) gives

$$\mathbb{E}[\mathcal{S}] = \sum_{k=1}^{\infty} k\left(e^{-(k-1)^2 T_{\mathrm{s}}^2/2\sigma^2} - e^{-k^2 T_{\mathrm{s}}^2/2\sigma^2}\right) .$$

Since the positive and negative terms converge individually to a finite value, we can rearrange the terms to complete the proof. □

Lemma 4. *The expected wait time is given by*

$$\mathbb{E}[\mathcal{W}] = T_{\rm s} \sum_{k=0}^{\infty} e^{-(kT_{\rm s})^2/2\sigma^2} - \sigma\sqrt{\tfrac{\pi}{2}} \ . \tag{14}$$

Proof. Substituting the Rayleigh CDF in (8) gives

$$\begin{aligned}\mathbb{E}[\mathcal{W}] &= T_{\rm s} - \sum_{k=1}^{\infty} \int_0^{T_{\rm s}} \left(e^{-(kT_{\rm s}-w)^2/2\sigma^2} - e^{-(kT_{\rm s})^2/2\sigma^2}\right) {\rm d}w \\ &= T_{\rm s} + T_{\rm s} \sum_{k=1}^{\infty} e^{-(kT_{\rm s})^2/2\sigma^2} - \sum_{k=1}^{\infty} \int_0^{T_{\rm s}} e^{-(kT_{\rm s}-w)^2/2\sigma^2} {\rm d}w \\ &= T_{\rm s} \sum_{k=0}^{\infty} e^{-(kT_{\rm s})^2/2\sigma^2} - \sum_{k=1}^{\infty} \int_0^{T_{\rm s}} e^{-(kT_{\rm s}-w)^2/2\sigma^2} {\rm d}w \ .\end{aligned}$$

Let $erf(x)$ be the error function defined as

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \ {\rm d}t \ .$$

Also, let $erfc(x) = 1 - erf(x)$ be the complimentary error function. Substituting $\dfrac{(kT_{\rm s} - w)}{\sigma\sqrt{2}} = t$, we get

$$\begin{aligned}\mathbb{E}[\mathcal{W}] &= T_{\rm s} \sum_{k=0}^{\infty} e^{-(kT_{\rm s})^2/2\sigma^2} - \sigma\sqrt{2} \sum_{k=1}^{\infty} \int_{(k-1)T_{\rm s}/\sigma\sqrt{2}}^{kT_{\rm s}/\sigma\sqrt{2}} e^{-t^2} \ {\rm d}t \\ &= T_{\rm s} \sum_{k=0}^{\infty} e^{-(kT_{\rm s})^2/2\sigma^2} - \sigma\sqrt{\tfrac{\pi}{2}} \sum_{k=1}^{\infty} \left(erf\left(\tfrac{kT_{\rm s}}{\sigma\sqrt{2}}\right) - erf\left(\tfrac{(k-1)T_{\rm s}}{\sigma\sqrt{2}}\right)\right) \\ &= T_{\rm s} \sum_{k=0}^{\infty} e^{-(kT_{\rm s})^2/2\sigma^2} - \sigma\sqrt{\tfrac{\pi}{2}} \sum_{k=1}^{\infty} \left(erfc\left(\tfrac{(k-1)T_{\rm s}}{\sigma\sqrt{2}}\right) - erfc\left(\tfrac{kT_{\rm s}}{\sigma\sqrt{2}}\right)\right) \\ &= T_{\rm s} \sum_{k=0}^{\infty} e^{-(kT_{\rm s})^2/2\sigma^2} - \sigma\sqrt{\tfrac{\pi}{2}} \Big(erfc(0) - erfc(\infty)\Big) \\ &= T_{\rm s} \sum_{k=0}^{\infty} e^{-(kT_{\rm s})^2/2\sigma^2} - \sigma\sqrt{\tfrac{\pi}{2}} \ .\end{aligned}$$

□

Substituting (14) and (13) in (5), we obtain the total energy penalty, given by

$$\begin{aligned}\mathcal{E}(T_{\rm s}) &= \alpha \sum_{k=0}^{\infty} e^{-k^2T_{\rm s}^2/2\sigma^2} + \beta\Big(T_{\rm s} \sum_{k=0}^{\infty} e^{-(kT_{\rm s})^2/2\sigma^2} - \sigma\sqrt{\tfrac{\pi}{2}}\Big) \\ &= (\alpha + \beta T_{\rm s}) \sum_{k=0}^{\infty} e^{-k^2T_{\rm s}^2/2\sigma^2} - \beta\sigma\sqrt{\tfrac{\pi}{2}} \ . \end{aligned}\tag{15}$$

Note that this energy penalty can be expressed using the Jacobi third theta function $\theta_3(z, q)$ given by [37]

$$\theta_3(z, q) = \sum_{n=-\infty}^{\infty} q^{n^2} e^{2niz} \ . \tag{16}$$

Although the Jacobi theta functions do not have a closed-form solution, it is finite and can be computed numerically or from tables. If required by the numerical solvers, this infinite sum can be approximated to a finite sum with a sufficiently large number of terms.

In what follows, we prove the convexity of $\mathcal{E}(T_{\rm s})$. Note that it involves a sum of infinite terms, and one can show that some of those terms are non-convex. Therefore, using typical methods to prove convexity is not applicable. Instead, we use a novel approach that involves a transformation of the infinite sum using the Poisson sum formula, which can be of independent interest in proving convexity for a sum of infinite terms, in general.

Lemma 5. *Let $f(k,t) = e^{-k^2t^2/2\sigma^2}$. Then for $t \in (0,\infty)$,*

$$\sum_{k=0}^{\infty} f(k,t) = \sqrt{2\pi}\ \frac{\sigma}{t} \sum_{k=0}^{\infty} f(k, \tfrac{2\pi\sigma^2}{t}) \ .$$

Proof. According to the Poisson sum formula, periodic summation of an aperiodic function is equal to the periodic summation of its Fourier transform. That is, for the given function $f(k,t)$

$$\begin{aligned}\sum_{k=-\infty}^{\infty} f(k,t) &= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,t) e^{-2\pi i n x} \ {\rm d}x \\ &= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2t^2/2\sigma^2} e^{-2\pi i n x} \ {\rm d}x \\ &= \sqrt{2\pi}\ \frac{\sigma}{t} \sum_{n=-\infty}^{\infty} e^{-2\pi^2n^2\sigma^2/t^2} \\ \Rightarrow \sum_{k=-\infty}^{\infty} f(k,t) &= \sqrt{2\pi}\ \frac{\sigma}{t} \sum_{k=-\infty}^{\infty} f(k, \tfrac{2\pi\sigma^2}{t}) \ .\end{aligned}$$

Since $f(k,t)$ is an even function of k, we get

$$\sum_{k=0}^{\infty} f(k,t) = \sqrt{2\pi}\ \frac{\sigma}{t} \sum_{k=0}^{\infty} f(k, \tfrac{2\pi\sigma^2}{t}) \ .$$

□

Lemma 6. *The function $f(t) = \sum_{k=0}^{\infty} f(k,t)$ is convex in t .*

Proof. We prove the convexity using the second derivative test. Since the function sequence converges, it is sufficient to prove that the terms inside the summation are independently convex. For this, we make use of the function and its alternate expression given by Lemma 5.

$$\begin{aligned}f(t) &= 1 + \sum_{k=1}^{\infty} e^{-k^2t^2/2\sigma^2} \\ \Rightarrow f''(t) &= \sum_{k=1}^{\infty} \frac{k^2}{\sigma^4}(k^2t^2 - \sigma^2) e^{-k^2t^2/2\sigma^2} \\ &\geq 0 \Leftarrow t \geq \frac{\sigma}{k} \quad ,\forall k \geq 1 \ . \\ &\text{Therefore, } f(t) \text{ is convex if } t \geq \sigma \ . \end{aligned}\tag{17}$$

Now using Lemma 5, we get

$$\begin{aligned}f(t) &= \sqrt{2\pi}\ \frac{\sigma}{t}\Big(1 + \sum_{k=0}^{\infty} e^{-2\pi^2k^2\sigma^2/t^2}\Big) \\ \Rightarrow f''(t) &= 2\sqrt{2\pi}\ \frac{\sigma}{t^3}\Big(1 + \frac{1}{x^4}(x^4 - 10\pi^2k^2\sigma^2x^2 \\ &\qquad + 8\pi^4k^4\sigma^4) e^{-2\pi^2k^2\sigma^2/t^2}\Big) \\ &\geq 0 \Leftarrow t \notin \Big(\pi\sigma\sqrt{5 - \sqrt{17}k},\ \pi\sigma\sqrt{5 + \sqrt{17}k}\Big), \ \forall k \geq 1 \ . \\ &\text{Therefore, } f(t) \text{ is convex if } t \leq 2.9\,\sigma \ . \end{aligned}\tag{18}$$

The proof is complete by combining the conditional convexity of $f(t)$ given in (17) and (18). □

Lemma 7. *The function $g(t) = tf(t)$ is convex in t .*

Proof. We prove the convexity in a way similar to Lemma 6.

$$\begin{aligned} f(t) &= t\Big(1+\sum_{k=1}^{\infty} e^{-k^2t^2/2\sigma^2}\Big) \\ \Rightarrow f''(t) &= \sum_{k=1}^{\infty}\frac{k^2 t}{\sigma^4}(k^2t^2-3\sigma^2)e^{-k^2t^2/2\sigma^2} \\ &\geq 0 \Leftarrow t \geq \sqrt{3}\,\frac{\sigma}{k} \quad ,\forall k \geq 1 \ . \\ &\text{Therefore, } f(t) \text{ is convex if } t \geq 1.74\,\sigma \ . \end{aligned} \tag{19}$$

Now using Lemma 5, we get

$$\begin{aligned} f(t) &= \sqrt{2\pi}\,\sigma\Big(1+\sum_{k=0}^{\infty} e^{-2\pi^2k^2\sigma^2/t^2}\Big) \\ \Rightarrow f''(t) &= \frac{4\sqrt{2\pi^5}\,k^2\sigma^3}{t^6}\big(4\pi^2k^2\sigma^2-3t^2\big)e^{-2\pi^2k^2\sigma^2/t^2} \\ &\geq 0 \Leftarrow t \leq \frac{2\pi\sigma}{\sqrt{3}}\cdot k \quad ,\forall k \geq 1 \ . \\ &\text{Therefore, } f(t) \text{ is convex if } t \leq 3.6\,\sigma \ . \end{aligned} \tag{20}$$

The proof is complete by combining the conditional convexity of $f(t)$ given in (19) and (20). □

Proposition 2. *The energy penalty $\mathcal{E}(T_s)$ is convex in T_s .*

Proof. The proof is straightforward from Lemma 6 and 7. □

With the convexity in hand, we can use known optimisation techniques [36] to solve $\hat{\mathcal{P}}$. Later in Section 6, we will use a simple bisection algorithm for this optimisation.

5 Solution to the Problem $\mathcal{P}$

In this section, we consider the optimisation problem $\mathcal{P}$ and find the optimal sampling interval and offset $\{T_\mathrm{s}^*, \delta^*\}$. Before going into the details, it is worth mentioning the motivation behind the introduction of the offset δ to the model. In practice, there is a minimum time threshold before which the event does not occur. For example, in a WCA, a human user takes a strictly non-zero minimum amount of time to finish a task and sampling before this minimum threshold simply adds to the energy wastage. The addition of an offset to the policy mitigates this energy wastage by not allowing the first sample to fall before a given time. However, this threshold is not necessary for benefiting from the offset. Even in the absence of a threshold, an offset in sampling is particularly useful in situations where the TTE distribution has a small variance and/or large mode. Here, the probability of the event occurring before the first sample is small enough so that the expected energy saving in delaying the sampling outweighs the expected cost of the encountered wait due to a (potential) event. As a result, the optimum value of an offset is always greater than or equal to the minimum threshold.

5.1 General TTE Distribution

As in the earlier section, we start by assuming a general TTE distribution and later extend it to particular distributions.

5.1.1 Expected number of samples

The probability that the number of samples taking any given integer value can be computed from the CDF and CCDF of of the TTE.

$$\begin{aligned} &\mathbb{P}(\mathcal{S}=1) = F_{\mathcal{T}}(\delta) \\ &\mathbb{P}(\mathcal{S}=k) = F_{\mathcal{T}}\big((k-1)T_\mathrm{s}+\delta\big) - F_{\mathcal{T}}\big((k-2)T_\mathrm{s}+\delta\big),\ \forall\ k \geq 2 \\ \Rightarrow\ &\mathbb{P}(\mathcal{S}\leq k) = F_{\mathcal{T}}\big((k-1)T_\mathrm{s}+\delta\big),\ \forall\ k\geq 1 \\ \Rightarrow\ &\mathbb{P}(\mathcal{S}> k) = \bar{F}_{\mathcal{T}}\big((k-1)T_\mathrm{s}+\delta\big),\ \forall\ k\geq 1 \ . \end{aligned} \tag{21}$$

The expected number of samples can be computed as

$$\begin{aligned} \mathbb{E}[\mathcal{S}] &= \sum_{k=0}^{\infty}\mathbb{P}(\mathcal{S}>k) \\ &= 1+\sum_{k=1}^{\infty}\bar{F}_{\mathcal{T}}\big((k-1)T_\mathrm{s}+\delta\big) \qquad \text{(from (21))} \\ \Rightarrow \mathbb{E}[\mathcal{S}] &= 1+\sum_{k=0}^{\infty}\bar{F}_{\mathcal{T}}\big(kT_\mathrm{s}+\delta\big) \ . \end{aligned} \tag{22}$$

5.1.2 Expected wait time

Recall that the offset $\delta \geq T_\mathrm{s}$ and that the maximum possible wait w is upper-bounded by the sampling interval (at the time of the event). As a result, a value of w such that $T_\mathrm{s} < w \leq \delta$ can arise only before the first sample while a value $w \leq T_\mathrm{s}$ can potentially arise before any sampling instant. The CDF of wait penalty $F_{\mathcal{W}}(w)$ can be obtained by taking the probability of the TTE to fall at most w short of any sampling instant.

$$F_{\mathcal{W}}(w) = \begin{cases} \sum\limits_{k=0}^{\infty}\big(F_{\mathcal{T}}(kT_\mathrm{s}+\delta)-F_{\mathcal{T}}(kT_\mathrm{s}+\delta-w)\big) & \text{if } w\leq T_\mathrm{s} \\ 1-F_{\mathcal{T}}\big(\delta-w\big) & \text{if } T_\mathrm{s}<w\leq\delta \\ 1 & \text{if } w>\delta \end{cases}$$

Since $\mathcal{W}$ is a non-negative random variable, we have

$$\begin{aligned} \mathbb{E}[\mathcal{W}] &= \int_0^{\infty}\big(1-F_{\mathcal{W}}(w)\big)\ \mathrm{d}w \\ &= \int_0^{T_\mathrm{s}}\Big(1-\sum_{k=0}^{\infty}\big(F_{\mathcal{T}}(kT_\mathrm{s}+\delta)-F_{\mathcal{T}}(kT_\mathrm{s}+\delta-w)\big)\Big)\ \mathrm{d}w \\ &\qquad + \int_{T_\mathrm{s}}^{\delta} F_{\mathcal{T}}\big(\delta-w\big)\ \mathrm{d}w \ . \end{aligned}$$

Since integral limits and infinite sum are finite, and the summand is non-negative, we get

$$\begin{aligned} \mathbb{E}[\mathcal{W}] = T_\mathrm{s} - \sum_{k=0}^{\infty}\int_0^{T_\mathrm{s}}\big(F_{\mathcal{T}}(kT_\mathrm{s}+\delta)-F_{\mathcal{T}}(kT_\mathrm{s}+\delta-w)\big)\ \mathrm{d}w \\ + \int_{T_\mathrm{s}}^{\delta} F_{\mathcal{T}}\big(\delta-w\big)\ \mathrm{d}w \ . \end{aligned} \tag{23}$$

As expected, a substitution of $\delta = T_\mathrm{s}$ in (22) or (23) under the policy Π returns the same result that we obtained under the policy $\hat{\Pi}$ discussed in Section 4.

5.2 Exponentially Distributed TTE

Due to the memory-less property of the distribution, it is predictable that the optimum offset under an exponentially distributed TTE takes the default value of $\delta = T_\mathrm{s}$; thus making the solution of $\mathcal{P}$ and $\hat{\mathcal{P}}$ one and the same. Before formally proving this in Proposition 3, we will first compute the energy penalty $\mathcal{E}(T_\mathrm{s},\delta)$ by substituting the CCDF and CDF of the exponential distribution in (22) and (23), respectively. We get

$$\begin{aligned}
\mathbb{E}[\mathcal{S}] &= 1 + \frac{e^{-\lambda\delta}}{1-e^{-\lambda T_\mathrm{s}}} \\
\mathbb{E}[\mathcal{W}] &= \frac{(1-\lambda\delta)e^{-\lambda T_\mathrm{s}} + \lambda T_\mathrm{s}\, e^{-\lambda\delta} + \lambda\delta - 1}{\lambda(1-e^{-\lambda T_\mathrm{s}})} \\
\Rightarrow \mathcal{E}(T_\mathrm{s},\delta) &= \alpha\mathbb{E}[\mathcal{S}] + \beta\mathbb{E}[\mathcal{S}] \\
&= \frac{e^{-\lambda T_\mathrm{s}}\big(\beta - \lambda(\beta\delta+\alpha)\big) + \lambda\big(\beta\delta+\alpha+e^{-\lambda\delta}(\beta T_\mathrm{s}+\alpha)\big) - \beta}{\lambda(1-e^{-\lambda T_\mathrm{s}})} \ . \qquad (24)
\end{aligned}$$

Proposition 3. *Given that the TTEs are exponentially distributed, $\delta^* = T_\mathrm{s}^*$ under an optimal policy for $\mathcal{P}$; therefore, $\mathcal{P}$ and $\hat{\mathcal{P}}$ are equivalent.*

Proof. Let $\pi^{(t_1,t_2\dots)}$ be a sampling policy with sampling intervals given by t_1, $t_2,\dots$ and let $\mathcal{E}\big(\pi^{(t_1,t_2\dots)}\big)$ be the associated energy penalty. Let $T_\mathrm{s}^\#$ be the optimum sampling interval obtained by solving $\hat{\mathcal{P}}$, where there is no offset. That is

$$\mathcal{E}\big(\pi^{(t,t,t,\dots)}\big) \geq \mathcal{E}\big(\pi^{(T_\mathrm{s}^\#,T_\mathrm{s}^\#,T_\mathrm{s}^\#,\dots)}\big),\ \forall t \geq 0 \ . \qquad (25)$$

Now, let T_s^* and δ^* be the solution to $\mathcal{P}$ resulting in an optimal policy $\pi^{(\delta^*,T_\mathrm{s}{}^*,T_\mathrm{s}{}^*,T_\mathrm{s}{}^*,\dots)}$ and let $F_{\mathcal{T}|\delta^*}$ be the conditional CDF of the TTE given $\mathcal{T} > \delta^*$. If the problem to compute the optimum sampling interval and phase is repeated at time δ^* using the conditional CDF, we get a new solution containing an optimum sampling interval $\tilde{T}_\mathrm{s}{}^*$ and an optimum (second) offset $\tilde{\delta}^*$. As this updated set of sampling policy denoted by $\pi^{(\delta^*,\tilde{\delta}^*,\tilde{T}_\mathrm{s}{}^*,\tilde{T}_\mathrm{s}{}^*,\dots)}$ is a superset of the previously obtained sampling policy $\pi^{(\delta^*,T_\mathrm{s}{}^*,T_\mathrm{s}{}^*,T_\mathrm{s}{}^*,\dots)}$, the new solution provides us with an energy penalty that is not worse than the one obtained before. However, due to the memory-less property of the exponential distribution we know that

$$F_{\mathcal{T}}(t) = F_{\mathcal{T}|\mathcal{T}>k\,\delta^*}(t + k\delta^*),\ \forall k \in \mathbb{N} \ .$$

Therefore by using the conditional CDF, optimisation problem is essentially unchanged and the newly obtained solution will be numerically equal to the corresponding values obtained in the previous step; i.e., $\tilde{T}_\mathrm{s}{}^* = T_\mathrm{s}{}^*$ and $\tilde{\delta}^* = \delta^*$. By repeating this process supported by similar arguments, we get a sequence of non-increasing energy penalties. That is

$$\begin{aligned}
\mathcal{E}\big(\pi^{(T_\mathrm{s}^\#,T_\mathrm{s}^\#,\dots)}\big) &\geq \mathcal{E}\big(\pi^{(\delta^*,T_\mathrm{s}^*,T_\mathrm{s}^*,\dots)}\big) && (\text{as } \hat{\Pi} \subseteq \Pi) \\
&\geq \mathcal{E}\big(\pi^{(\delta^*,\delta^*,T_\mathrm{s}^*,T_\mathrm{s}^*,\dots)}\big) \\
&\ \ \vdots \\
&\geq \mathcal{E}\big(\pi^{(\delta^*,\delta^*,\dots)}\big) \\
&\geq \mathcal{E}\big(\pi^{(T_\mathrm{s}^\#,T_\mathrm{s}^\#,\dots)}\big) && (\text{from } (25)) \\
\Rightarrow \mathcal{E}\big(\pi^{(T_\mathrm{s}^\#,T_\mathrm{s}^\#,\dots)}\big) &= \mathcal{E}\big(\pi^{(\delta^*,\delta^*,\dots)}\big) \ .
\end{aligned}$$

Therefore, $\delta^* = T_\mathrm{s}^\#$ as a result of the convexity of $\mathcal{E}(T_\mathrm{s})$ from Proposition 2. Thus, the optimum set of sampling instants under Π is $\{t = kT_\mathrm{s}^*,\ k \in \mathcal{N}^+\}$, which is essentially the solution to $\hat{\mathcal{P}}$, proving that $\mathcal{P} \equiv \hat{\mathcal{P}}$. □

As a result of the equivalence of $\mathcal{P}$ and $\hat{\mathcal{P}}$, we do not benefit from proceeding with the relatively complex optimisation of (24) and we can use the solution from Proposition 1 to find the optimum energy penalty.

Algorithm 1 Algorithm to find optimum sampling interval T_s^* and optimum offset δ^* with the constraint $\delta = nT_\mathrm{s}$, $n \in \mathbb{N}^+$.

```
Initialise the upper limit of optimising variables, $T_\mathrm{s}^{(\mathrm{max})}$ and $n^{(\mathrm{max})}$;
Initialise stopping criterion $\xi$;
$n^{(\mathrm{L})} \leftarrow 1$; $n^{(\mathrm{H})} \leftarrow n^{(\mathrm{max})}$;
for $i \leftarrow 0$ to $\lceil \log_2(n^{(max)} - 1) \rceil$ do
    $n \leftarrow (n^{(\mathrm{L})} + n^{(\mathrm{H})})/2$;
    $T_\mathrm{s}^{(\mathrm{L})} = 0$; $T_\mathrm{s}^{(\mathrm{H})} = T_\mathrm{s}^{(\mathrm{max})}$;
    while $T_\mathrm{s}^{(\mathrm{H})} - T_\mathrm{s}^{(\mathrm{L})} > \xi$ do
        $T_\mathrm{s} \leftarrow (T_\mathrm{s}^{(\mathrm{L})} + T_\mathrm{s}^{(\mathrm{H})})/2$;
        if $\frac{\partial}{\partial T_\mathrm{s}}\mathcal{E}(T_\mathrm{s}, nT_s) \geq 0$ then
            $T_\mathrm{s}^{(\mathrm{H})} \leftarrow T_\mathrm{s}$;
        else
            $T_\mathrm{s}^{(\mathrm{L})} \leftarrow T_\mathrm{s}$;
        end
    end
    if $\mathcal{E}(T_s, nT_s) - \mathcal{E}(T_s, (n-1)T_s) \geq 0$ then
        $n^{(\mathrm{H})} \leftarrow n$;
    else
        $n^{(\mathrm{L})} \leftarrow n$;
    end
end
$T_\mathrm{s}^* = T_\mathrm{s}$; $\delta^* = nT_\mathrm{s}^*$;
```

5.3 Rayleigh Distributed TTE

Lemma 8. *The energy penalty $\mathcal{E}$ under Rayleigh distributed TTE is given by*

$$\mathcal{E}(T_s,\delta) = (\alpha + \beta T_s)\sum_{k=0}^{\infty} e^{-(kT_\mathrm{s}+\delta)^2/2\sigma^2} - \beta\sigma\sqrt{\tfrac{\pi}{2}} + \beta\delta + \alpha \ . \quad (26)$$

Proof. Substituting the CDF of the Rayleigh distribution, $F_{\mathcal{T}}(t) = 1 - e^{-t^2/2\sigma^2}$ in (22) or (23) gives

$$\mathbb{E}[\mathcal{S}] = 1 + \sum_{k=0}^{\infty} e^{-(kT_\mathrm{s}+\delta)^2/2\sigma^2} \qquad (27)$$

$$\mathbb{E}[\mathcal{W}] = \delta + T_\mathrm{s}\sum_{k=0}^{\infty} e^{-(kT_\mathrm{s}+\delta)^2/2\sigma^2} - \sigma\sqrt{\frac{\pi}{2}} \ . \qquad (28)$$

The energy penalty $\mathcal{E}(T_\mathrm{s},\delta)$ can be computed by adding $\mathbb{E}[\mathcal{S}]$ and $\mathbb{E}[\mathcal{W}]$ using the weights α and β, respectively. □

Proving the convexity of $\mathcal{E}(T_\mathrm{s},\delta)$ in the two variables T_s and δ is hard. The difficulty arises as we need to prove that the Hessian is non-negative and showing this for the infinite sum is highly non-trivial. Furthermore, the proof that we used in Lemma 5 cannot be extended as the function is not even in T_s after the addition of the second variable δ. Instead, we will consider a sub-optimal solution by assuming that the offset is an integer multiple of the sampling interval. We will now show that the energy penalty is convex in T_s for a fixed offset of the form $\delta = nT_\mathrm{s}$.

Lemma 9. *The function $f_n(t) = (a + bt)\sum_{k=0}^{\infty} e^{-((k+n)t)^2/2\sigma^2}$ is convex in t for $n \in \mathbb{N}$.*

Proof. See that $f_n(t)$ for any n is obtained by removing first $n-1$ terms from $f_0(t)$. From Proposition 2, we know that $f_0(t)$ is convex.

$$\Rightarrow 0 \leq f_0''(t) = \sum_{k=0}^{\infty} \frac{k^2t^2(a+bt) - s^2(a+3bt)}{\sigma^2} k^2 e^{-(k+n)^2t^2/2\sigma^2} .$$

As the terms in $f_0''(t)$ are non-decreasing with k, removing first $n-1$ terms will not affect the sign of the sum and therefore $f_n(t)$ is also convex. □

We can see with the help of (16) and (26) that the energy penalty is the product of a linearly increasing function and a continuously decreasing and convex Jacobi theta function. This shows the presence of a single minima with respect to δ for any fixed T_{s}. With this information and the convexity in T_{s} for any fixed δ (Lemma 9), we use Algorithm 1 to energy minimising sampling interval and offset while satisfying the constraint $\delta = nT_{\mathrm{s}},\ n \in \mathbb{N}$. It can be easily seen that this algorithm is an adapted version of the bisection algorithm. We use one bisection algorithm to find the optimum sampling interval for a given integer n and another one to find the optimum n, in a nested manner. The former uses $\left\lceil \log\left((T_{\mathrm{s}}^{(\mathrm{H})} - T_{\mathrm{s}}^{(\mathrm{L})})/\xi\right) \right\rceil$ iterations and the latter uses $\lceil \log(n^{\max} - 1) \rceil$ iterations, where $T_{\mathrm{s}}^{(\mathrm{H})}, T_{\mathrm{s}}^{(\mathrm{L})}, \xi$ and $n^{\max}$ are an upper and lower bounds for the sampling interval, a small enough stopping criterion and an upper bound for n, respectively. As we will see later, this is typically in the order of a few tens of iterations (bisections) to get to millisecond precision. Note that this algorithm is run one time as part of the admission control and hence the complexity and energy usage to execute this algorithm is negligible.

In the next section, we will compare this energy penalty to $\mathcal{E}^*$ obtained using a brute force search and show that Algorithm 1 attains near-optimality. As a direct result of this near-optimality, we will reuse the same notations $(\cdot)^*$ to represent the solution obtained using Algorithm 1, unless otherwise necessary.

6 Numerical Results

In this section, we present the numerical results to illustrate the behaviour of the energy penalty, the gain achieved by the proposed design, and their dependency on the system parameters. We consider two parameter settings that are motivated by the characteristics of the CPS and the VAS. We start with describing the characterisation of both of these systems and the underlying distribution of the observed process. Afterwards, we compare the proposed design(s) with a baseline sampling policy that is used in practice. Furthermore, to verify the near-optimality of Algorithm 1 in solving $\mathcal{P}$, we compare its solution with that of a brute force method. For the CPS with an exponentially distributed TTE, we already saw in 3 that the offset is irrelevant and hence we are not using Algorithm 1. As a result, this comparison with a brute-force solution is required only for the VAS but not for the CPS.

6.1 Parameter Settings

For the CPS, we consider a failure detection system where the TTE is exponentially distributed with a mean of 10s. To detect the events (i.e., failures for a CPS), the terminal samples the process and sends the sample to the back-end via a low-power, low-throughput transmitter. For the device, we refer to the wireless sensor network characterisation in [38] with the data size fixed at 127 Bytes. For communication technology, we choose 802.15.4 operating in the Sub1GHz ISM band with an approximate throughput of 250kbps, resulting in a communication delay of 4ms. The device is assumed to work at 3V drawing a current of 15mA during communication and 5mA during idling. We also assume that the processing time of the successful sample $\tau_{\mathrm{s}} = 5$ms.

On the other hand, a VAS typically need to transmit video frames, thus requiring a larger throughput. We refer to the WCA Lego experiment [33] where a set of tasks performed by a human user are monitored to detect the task completion by taking snapshots of the progress. The frame size used in the experiment is roughly 300kB (640×480 resolution) and the mean task time observed is 4.846s. The processing time observed at the back end is approximately 525ms. For the terminal device, we consider a Google Glass using an 802.11ax transmitter providing a data rate of 400Mbps thus resulting in a 5.85ms communication delay for each snapshot. The Google Glass typically takes 334mW during *active/screen-off* and 2960mW during *video chat* [39]. Thus, we assume these power figures as the idle power and the communication power, respectively. On top of this, we also assume a minimum possible task time $t_{\min}$ (response time of the human user) of 0.5s. The parameters used for the CPS and the VAS are presented in TABLE 2.

	P_0	P_{c}	τ_{c}	τ_{s}	β/α	$\mathbb{E}[\mathcal{T}]$	$t_{\min}$
CPS	15mW	45mW	4ms	5ms	125	10s	–
VAS	334mW	2.96W	5.85ms	525ms	21.7	4.84s	0.5s

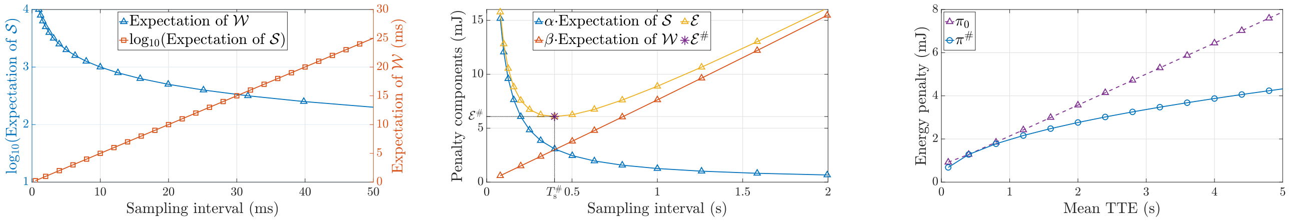
TABLE 2: Parameters used for the cyber-physical system (CPS) and the video analytics system (VAS).

Even though our solution does not depend on the type of energy source, we assume that the CPS and the VAS are powered by a fixed battery pack of arbitrary capacity. As battery life improvement is one of the directly observable results of energy saving, we evaluate the performance of the proposed solution in terms of percentage increase in battery life. Note that as we use a percentage increase, the exact capacity of the battery is irrelevant.

For the performance evaluation of the proposed policies $\pi^{\#}$ and π^*, we consider a baseline periodic sampling policy denoted by π_0 as shown in TABLE 3. For selecting the baseline sampling rate for the VAS, we take a hint from the WCA system [35] that motivated the system model. The authors use a video camera to capture the monitored process that runs at a frame rate of 24fps which ideally results in a sampling interval of 41.6ms. However, whenever the processing of the discarded samples takes more than

		π_0	$\pi^{\#}$	π^*	π_{b}
CPS	T_{s}	83.3 ms	$T_{\mathrm{s}}^{\#}$	-	-
VAS	T_{s}	83.3 ms	$T_{\mathrm{s}}^{\#}$	T_{s}^*	Brute Force
	δ	-	-	δ^*	

TABLE 3: Various sampling policies considered for comparison and their corresponding sampling interval and offset.

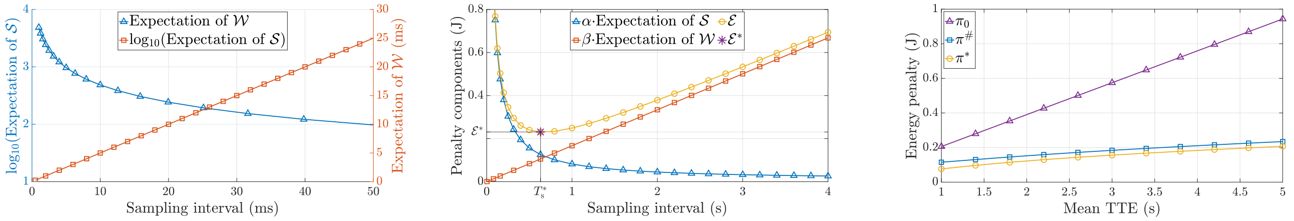


(a) $\log_{10}(\mathbb{E}[\mathcal{S}])$, $\mathbb{E}[\mathcal{W}]$ vs. T_s.

(b) Penalty components, $\mathcal{E}$ vs. T_s.

(c) $\mathcal{E}$ vs. $\mathbb{E}[\mathcal{T}]$.

Fig. 2: Expected number of samples $\mathbb{E}[\mathcal{S}]$, expected wait $\mathbb{E}[\mathcal{W}]$, energy penalty $\mathcal{E}$ and its components plotted against the sampling interval T_s and mean TTE $\mathbb{E}[\mathcal{T}]$ for the CPS. Proposed policy $\pi^{\#}$ is compared with the baseline policy π_0.



(a) $\log_{10}(\mathbb{E}[\mathcal{S}])$, $\mathbb{E}[\mathcal{W}]$ vs. T_s.

(b) Penalty components, $\mathcal{E}$ vs. T_s.

(c) $\mathcal{E}$ vs. $\mathbb{E}[\mathcal{T}]$.

Fig. 3: Expected number of samples $\mathbb{E}[\mathcal{S}]$, expected wait $\mathbb{E}[\mathcal{W}]$, energy penalty $\mathcal{E}$ and its components vs. the sampling interval T_s and mean TTE $\mathbb{E}[\mathcal{T}]$ for the VAS. Proposed policies $\pi^{\#}$ and π^{*} are compared with the baseline policy π_0.

41.6ms, the next sample is delayed accordingly. As a result, the mean sampling interval is observed to be around 83.3ms thus justifying a baseline with sampling interval anywhere in between 41.6ms and 83.3ms. That being said, we will see in the next subsection that the optimum sampling interval for the considered VAS is around 300ms. As a result, even though our model does not take samples by considering the processing time, we take our baseline sampling to be 83.3ms to show the minimum performance improvement when the baseline is closest to the optimum We use the same baseline sampling interval for the CPS for simplicity in comparison.

To verify the near-optimality of π^* obtained using Algorithm 1, we compare it with that of a brute force solution of $\mathcal{P}$. To find this brute force solution, we divide the practically feasible domain of the two-dimensional plane formed by T_s and δ into grids of size $10^{-3} \times 10^{-3}$ s^2 and search for the optimum solution (also known as grid search).

As the explanations to some of the figures of the CPS and the VAS involve similar arguments, we explain them concurrently whenever it is convenient. Also, note that the y-axes of some of these figures are kept different intentionally for better visualisation. The parameters for all the plots are taken from TABLE 2 and TABLE 3, with a potential exception to the variable parameter under discussion in a particular plot.

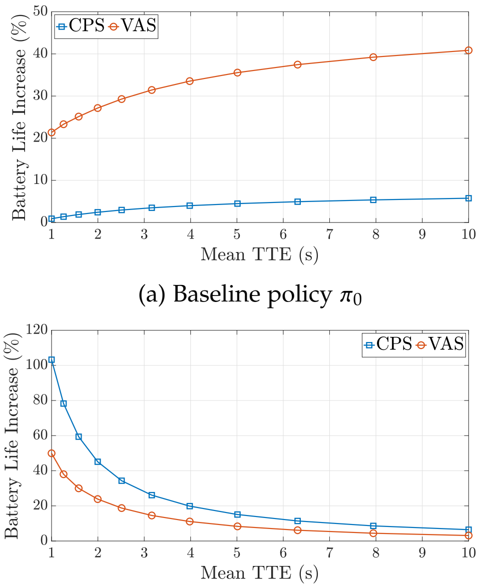
6.2 Energy Penalty

In Fig. 2a and Fig. 3a we present the the expected number of samples $\mathbb{E}[\mathcal{S}]$ and the expected wait $\mathbb{E}[\mathcal{W}]$ by varying T_s for the CPS and the VAS, respectively. For comparison, at $T_\text{s} = 10$ms, the CPS and the VAS expect 1000 and 400 samples per event, respectively. The expected wait however is approximately the same at around 5ms. As noted in Sections 4 and 5, $\mathbb{E}[\mathcal{S}]$ and $\mathbb{E}[\mathcal{W}]$ show opposing behaviour with an increase in T_s. After a very rapid decrease, $\mathbb{E}[\mathcal{S}]$ goes asymptotically to 1 with increasing T_s. This points to the single sample that is ideally required for the event detection. On the other hand, $\mathbb{E}[\mathcal{W}]$ shows a gradual (approximately linear) increase with T_s. For instance, doubling the sampling interval doubles the wait. For a VAS operating at a sampling frequency of 1s, this corresponds to approximately 50% increase in the TTF. Recall that TTF is the effective delay experienced by the human user.

In Fig. 2b and Fig. 3b, we show the energy penalty and its components as a function of T_s. Observe $\mathbb{E}[\mathcal{S}]$ and $\mathbb{E}[\mathcal{W}]$ exhibit an opposing behaviour with a change in T_s resulting in an energy penalty minima in their weighted sum. This minima and the corresponding T_s is also marked in the figures. The qualitative behaviour of the energy penalty and its components is similar for the CPS and the VAS with a minimum value of 6mJ and 230mJ, respectively. Another important observation is the change in $\mathcal{E}$ as a result of a deviation of T_s from the optimum. As $\mathcal{E}$ increases much more rapidly with a negative deviation than with a positive deviation, it is particularly important to avoid any negative errors in calculating this optimum. That is, an oversampling can cost much more than an undersampling by a similar amount. These errors can arise due to an incorrect estimation of the TTE statistics or insufficient convergence of the optimisation algorithm.

6.3 Energy Saving

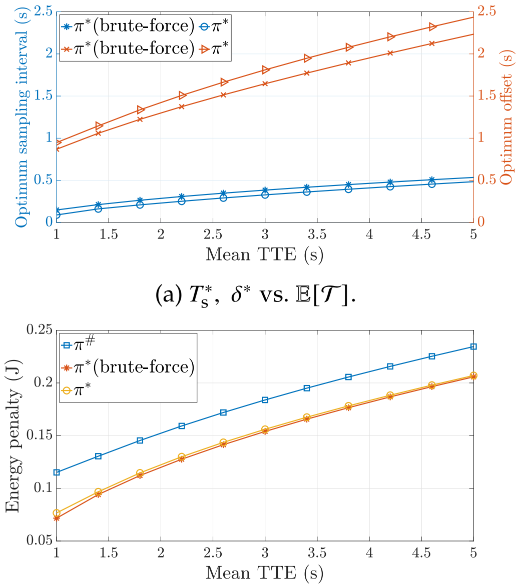
In Fig. 2c and Fig. 3c we show energy penalty incurred for the CPS and the VAS as a function of the mean TTE for various sampling policies from TABLE 3. Since the mean TTE value is application-specific, plotting energy penalty across mean TTE values provides an insight into how the behaviour of the particular CPS and VAS scales across



(a) Baseline policy π_0

(b) Baseline policy with $T_s = 2$s.

Fig. 4: Percentage increase in battery life achieved by the policy π^* vs. mean TTE $\mathbb{E}[\mathcal{T}]$ for the CPS and the VAS. The increase is calculated by considering two baselines policies: (a) π_0, and (b) a policy with sampling interval $T_s = 2$s.



(a) T_s^*, δ^* vs. $\mathbb{E}[\mathcal{T}]$.

(b) $\mathcal{E}$ vs. $\mathbb{E}[\mathcal{T}]$.

Fig. 5: Optimum sampling interval T_s^*, optimum offset δ^* and the corresponding energy penalty $\mathcal{E}^*$ for various sampling policies vs. the mean TTE $\mathcal{E}[\mathcal{T}]$ for the VAS.

application scenarios with different parameterisation. The energy penalty reduction achieved by the proposed policies is clear from these figures. Note that when the sampling interval of 83.3ms is smaller than $T_s^{\#}$ or T_s^*, the penalty is dominated by the discarded samples. As a result, the difference in $\mathcal{E}$ increases with an increase in mean TTE. This increase is much more prominent for the VAS.

The increase in battery life is a more direct measure of energy saving from an application perspective. In Fig. 4 we present this by plotting the percentage increase in the expected battery life of the terminal. In Fig. 4a we choose the baseline π_0 whereas in Fig. 4b we choose a different baseline sampling interval of 2s. We can see that irrespective of the baseline or the application considered, there is an observable increase in battery life. Particularly for the VAS, the proposed sampling policy improves the battery life by 36% over π_0 at the considered mean TTE of 4.84s.

The motivation to consider two baselines is the value of their sampling interval relative to the optimum sampling interval (for the TTE range considered). The sampling interval of 0.83ms is much smaller, and the sampling interval 2s is much larger than the optimal sampling intervals. Therefore, the former results in oversampling and the latter results in undersampling which results in shifting the dominating component of the energy penalty from $\mathbb{E}[\mathcal{S}]$ to $\mathbb{E}[\mathcal{W}]$. As a result, the increase in battery life is a increasing function of the mean TTE in Fig. 4a, whereas it is a decreasing function in Fig. 4b. Note that the optimum sampling interval is positively correlated with the mean task time and the battery life is inversely proportional to the energy. Therefore, this change of percentage increase in battery life is in alignment with the variation of $\mathbb{E}[\mathcal{S}]$ and $\mathbb{E}[\mathcal{W}]$ with T_s observed earlier in Fig. 2 and Fig. 3.

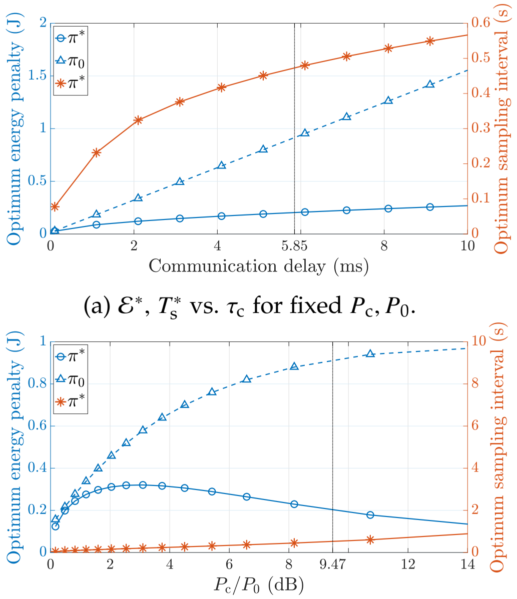
6.4 Comparison of Sampling Policies

We saw the battery life improvement provided by π^* over π_0 earlier in Fig. 4. Even though we have two solutions – $\pi^{\#}$ and π^* – for the VAS, we considered only the better performing π^* in that figure. Now in Fig. 5, we compare the performance of the two policies and see the benefit provided by adding the offset δ. We also compare the proposed π^* with the brute force solution discussed in Section 6.1. In Fig. 5a, we compare the optimum sampling interval and offset obtained by Algorithm 1 and brute-force, whereas Fig. 5b compares the optimum energy penalty attained by $\pi^{\#}$ and π^*. From Fig. 5a, we can see that there is an observable (albeit small) difference in T_s^* (and δ^*) obtained by the algorithm and the brute-force. Nonetheless, Fig. 5b shows that the difference in energy penalty generated as a result of this difference is very minimal, thus confirming the near-optimality of Algorithm 1. We also observe that the additional penalty reduction offered by $\mathcal{P}$ over $\hat{\mathcal{P}}$ amounts to roughly 51% at a mean task time of 1s and 14% at a mean task time of 5s, thus demonstrating the advantage of introducing the offset.

6.5 Variation of $\mathcal{E}$ with System Parameters

Recall that the optimum sampling interval depends only on the ratio $\beta/\alpha = \tau_c^{-1}(\frac{P_c}{P_0}-1)^{-1}$. Fig. 6a and Fig. 6b illustrates the variation of the energy penalty of the VAS with respect to the system parameters τ_c and P_c/P_0. Observe from the figures that the energy penalty obtained by solving $\mathcal{P}$ varies much less with the system parameters than the baseline policy. This shows the relative independence of the optimum energy penalty with the system parameters. Since these results are qualitatively similar for the CPS, we omit them due to space constraints.

For fixed power, communication delay τ_c is dictated by the data size per sample and the transmission schemes.



(a) $\mathcal{E}^*$, T_s^* vs. τ_c for fixed P_c, P_0.

(b) $\mathcal{E}^*$, T_s^* vs. $\log_{10}(P_c/P_0)$ for fixed τ_c, P_c.

Fig. 6: Variation of optimum sampling interval T_s^* and the corresponding energy penalty $\mathcal{E}^*$ for various sampling policies with variation in communication delay τ_c and idle power P_0 for the VAS. The vertical lines denotes τ_c and $\log_{10}(P_c/P_0)$ of the VAS from TABLE 2.

Fig. 6a illustrates the dependency of $\mathcal{E}$ with respect to these aspects of communication. We can see that a larger communication delay results in larger T_s^*. However, the corresponding increase in the incurred penalty is much smaller for the proposed optimum. In Fig. 6b we show the change in T_s^* and $\mathcal{E}$ with respect to P_c/P_0 for constant τ_c and P_c. We can see that the optimum sampling interval increases with an increase in the difference between the communication power and idle power. However, a more interesting observation is the corresponding change in the energy penalty. With $P_0 << P_c$, we obtain a significant penalty reduction by using π^*. On the other hand, as the P_0 approaches P_c (i.e., 0dB in the figure), the penalty reduction goes to zero.

With advancements in semiconductor technology that reduces P_0 and the emergence of communication standards like the mmWave, P_c/P_0 ratio is expected to increase from the 9.47dB that we modelled. Similarly, increasing image processing capabilities increase τ_c as well. As a result, it is evident from the figures that the gain offered by the proposed algorithm is more likely to increase in the future.

7 Conclusion and Future Scope

We considered an edge-based feedback system that captures essential events via sampling and proposed an optimisation framework with which the sampling interval that minimises the energy consumption can be computed. Apart from the generic approach to solving the optimisation problem for an arbitrary task time distribution, we also considered two particular examples of interests – a VAS (video analytics system) and a CPS (cyber physical system). The TTE (time to event) of the VAS follows a Rayleigh distribution while that of the CPS follows an exponential distribution. These two systems are used to illustrate the behaviour of the variable components of energy. We discussed the energy savings enabled by the optimum sampling interval, the benefit provided by offsetting the first sample and the near-optimality of the proposed algorithm. We also discussed the dependency of the optimum to the system parameters like communication delay and idle power. Finally, we also discussed the additional energy expended as a result of a computation error and concluded that this expense is relatively steeper for a negative error.

From a mathematical perspective, the basic difference between a CPS and a VAS is the underlying TTE distribution. However, from a design perspective, the one important distinction is the benefit of adding the offset δ. As explained before, offset is particularly useful when the statistical mode of the TTE distribution is large enough and the variance of the distribution is small enough so that the probability of events occurring in the early stages of a monitoring cycle is small. As a result, offset is not useful for an application with an exponentially distributed TTE (see Proposition 3) but is useful when TTE is Rayleigh distributed. However, there are other distributions where the offset is much more useful and the proposed solution can provide a higher performance improvement. For instance, the exponentially modified Gaussian distribution can model the human response times in some similar applications.

There are aperiodic policies that can potentially outperform the optimum periodic policy and there is a need for a better understanding of such policies. The characterisation of an optimal aperiodic policy is a primary direction to extend this work. We have created the model with fixed communication and processing delays. Even though this assumption is fairly accurate in our model, addressing the possibility of random communication and processing delays will be another interesting research direction. Furthermore, we considered the problem with only one category of event. However, systems might need to capture different categories of events, and thus might have different severity or costs associated with them. In a VAS, the responsiveness of the system need to be taken into consideration and any minimisation of energy usage that affects the responsiveness over a certain limit should be avoided. Such demands for an additional characterisation that adds a variable weight to the required energy savings is yet another important direction that this work can be extended.

8 Acknowledgement

This research was supported in part by VINNOVA Competence Center for Trustworthy Edge Computing Systems and Applications (TECoSA) at KTH Royal Institute of Technology, and in part by the Swedish Foundation for Strategic Research (SSF) under grant ITM17-0246 (ExPECA).

References

[1] F. Alriksson, L. Boström, J. Sachs, Y. Wang, and A. Zaidi, "Critical iot connectivity: Ideal for time-crit-ical communications," *Ericsson technology review*, 2020.

[2] 3GPP, "Digital cellular telecommunications system (Phase 2+) (GSM;Universal Mobile Telecommunications System (UMTS;LTE;5G," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 21.915, 10 2019, version 15.0.0.

[3] ——, "Digital cellular telecommunications system (Phase 2+) (GSM;Universal Mobile Telecommunications System (UMTS;LTE;5G," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 21.916, 09 2021, version 16.0.1.

[4] B. Lu, Y. Li, X. Wu, and Z. Yang, "A review of recent advances in wind turbine condition monitoring and fault diagnosis," in *2009 IEEE Power Electronics and Machines in Wind Applications*, 2009, pp. 1–7.

[5] W. A. Shewhart, "The application of statistics as an aid in maintaining quality of a manufactured product," *Journal of the American Statistical Association*, vol. 20, no. 152, pp. 546–548, 1925.

[6] J. Gertler, "Survey of model-based failure detection and isolation in complex plants," *IEEE Control Systems Magazine*, vol. 8, no. 6, pp. 3–11, 1988.

[7] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[8] V. V. Veeravalli and T. Banerjee, "Quickest change detection," 2012.

[9] A. Mahmood, K. Shi, S. Khatoon, and M. Xiao, "Data mining techniques for wireless sensor networks: A survey," *International Journal of Distributed Sensor Networks*, vol. 9, no. 7, p. 406316, 2013.

[10] G. Ananthanarayanan, P. Bahl, P. Bodík, K. Chintalapudi, M. Philipose, L. Ravindranath, and S. Sinha, "Real-time video analytics: The killer app for edge computing," *Computer*, vol. 50, no. 10, pp. 58–67, 2017.

[11] B. Luo, S. Tan, Z. Yu, and W. Shi, "Edgebox: Live edge video analytics for near real-time event detection," in *2018 IEEE/ACM Symposium on Edge Computing (SEC)*, 2018, pp. 347–348.

[12] Y. Ye, S. Ci, A. K. Katsaggelos, Y. Liu, and Y. Qian, "Wireless video surveillance: A survey," *IEEE Access*, vol. 1, pp. 646–660, 2013.

[13] X. Wang, S. Wang, and D. Bi, "Distributed visual-target-surveillance system in wireless sensor networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 5, pp. 1134–1146, 2009.

[14] H. Wang, F. Zhai, Y. Eisenberg, and A. K. Katsaggelos, "Cost-distortion optimized unequal error protection for object-based video communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1505–1516, 2005.

[15] S. Cui and A. J. Goldsmith, "Cross-layer optimization of sensor networks based on cooperative mimo techniques with rate adaptation," in *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications*, 2005, pp. 960–964.

[16] M. Casares and S. Velipasalar, "Adaptive methodologies for energy-efficient object detection and tracking with battery-powered embedded smart cameras," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 10, pp. 1438–1452, 2011.

[17] J. A. Fuemmeler and V. V. Veeravalli, *Smart Sleeping Policies for Energy-Efficient Tracking in Sensor Networks*. Boston, MA: Springer US, 2008, pp. 267–287.

[18] B. Shi, J. Yang, Z. Huang, and P. Hui, "Offloading guidelines for augmented reality applications on wearable devices," in *Proc. of ACM International Conference on Multimedia*, 2015, p. 1271–1274.

[19] R. Kemp, N. Palmer, T. Kielmann, F. Seinstra, N. Drost, J. Maassen, and H. Bal, "eyeDentify: Multimedia cyber foraging from a smartphone," in *Proc. IEEE International Symposium on Multimedia*, 2009, pp. 392–399.

[20] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[21] J. P. Champati and B. Liang, "Semi-online algorithms for computational task offloading with communication delay," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 4, pp. 1189–1201, 2017.

[22] ——, "Single restart with time stamps for parallel task processing with known and unknown processors," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 1, pp. 187–200, 2020.

[23] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4569–4581, 2013.

[24] C. You, K. Huang, and H. Chae, "Energy efficient mobile cloud computing powered by wireless energy transfer," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1757–1771, 2016.

[25] K. Kumar and Y. Lu, "Cloud computing for mobile users: Can offloading computation save energy?" *Computer*, vol. 43, no. 4, pp. 51–56, 2010.

[26] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 45–55, 2014.

[27] M. Zhao, J.-J. Yu, W.-T. Li, D. Liu, S. Yao, W. Feng, C. She, and T. Q. S. Quek, "Energy-aware offloading in time-sensitive networks with mobile edge computing," *CoRR*, vol. abs/2003.12719, 2020.

[28] S. E. Mahmoodi, R. N. Uma, and K. P. Subbalakshmi, "Optimal joint scheduling and cloud offloading for mobile applications," *IEEE Transactions on Cloud Computing*, vol. 7, no. 2, pp. 301–313, 2019.

[29] W. Zhang, Y. Wen, and D. O. Wu, "Collaborative task execution in mobile cloud computing under a stochastic wireless channel," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 81–93, 2015.

[30] D. Huang, P. Wang, and D. Niyato, "A dynamic offloading algorithm for mobile computing," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 1991–1995, 2012.

[31] W. Heemels, K. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 2012, pp. 3270–3285.

[32] V. N. Moothedath, J. P. Champati, and J. Gross, "Energy-optimal sampling of edge-based feedback systems," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021, pp. 1–6.

[33] K. Ha, Z. Chen, W. Hu, W. Richter, P. Pillai, and M. Satyanarayanan, "Towards wearable cognitive assistance," in *Proc. of the 12th Annual International Conference on Mobile Systems, Applications, and Services*. Association for Computing Machinery, 2014, p. 68–81.

[34] Z. Chen, W. Hu, J. Wang, S. Zhao, B. Amos, G. Wu, K. Ha, K. Elgazzar, P. Pillai, R. Klatzky, D. Siewiorek, and M. Satyanarayanan, "An empirical study of latency in an emerging class of edge computing applications for wearable cognitive assistance," in *Proc. ACM/IEEE Symposium on Edge Computing*, 2017.

[35] M. O. Muñoz, R. Klatzky, J. Wang, P. Pillai, M. Satyanarayanan, and J. Gross, "Impact of delayed response on wearable cognitive assistance," *CoRR*, vol. abs/2011.02555, 2020.

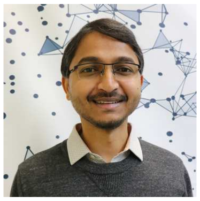
[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[37] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th ed., ser. Cambridge Mathematical Library. Cambridge University Press, 1996.

[38] A. Moschitta and I. Neri, "Power consumption assessment in wireless sensor networks," *ICT-Energy-Concepts Towards Zero-Power Information and Communication Technology*, 02 2014.

[39] R. LiKamWa, Z. Wang, A. Carroll, F. X. Lin, and L. Zhong, "Draining our glass: An energy and heat characterization of google glass," in *Proceedings of 5th Asia-Pacific Workshop on Systems*, ser. APSys '14. New York, NY, USA: Association for Computing Machinery, 2014.

Vishnu Narayanan Moothedath is a doctoral student in the department of Intelligent Systems under the School of Electrical Engineering and Computer Science (EECS) at KTH Royal Institute of Technology from 2021. He completed his master's degree in Communication Systems from the Indian Institute of Technology (IIT) Madras in 2016 and his bachelor's degree in Electrics and Communication Engineering from the National Institute of Technology (NIT) Calicut in 2012. He has also worked for five years in the cellular industry with Intel India Pvt. Ltd. and Apple India Pvt. Ltd. in their respective LTE/5G-NR base-band modem group. His current research is in the area of edge computing and performance optimisation with a specific focus on improving the energy efficiency and responsiveness of edge computing systems through optimised sampling.



Jaya Prakash Champati (IEEE member) received his bachelor of technology degree from the National Institute of Technology Warangal, India in 2008, and master of technology degree from the Indian Institute of Technology (IIT) Bombay, India in 2010. He received his PhD in Electrical and Computer Engineering from the University of Toronto, Canada in 2017. From 2017-2020, he was a post-doctoral researcher with the division of Information Science and Engineering, EECS, KTH Royal Institute of Technology, Sweden. He is currently a Research Assistant Professor at IMDEA Networks Institute, Madrid, Spain. His general research interest is in the design and analysis of algorithms for scheduling problems that arise in networking and information systems. Currently, his research focus is in Edge Computing/Intelligence, Age of Information, Cyber-Physical Systems (CPS), and Internet of Things (IoT). Prior to joining PhD he worked at Broadcom Communications, where he was involved in developing the LTE MAC layer. He was a recipient of the best paper award at IEEE National Conference on Communications, India, 2011.

James Gross received his Ph.D. degree from TU Berlin in 2006. From 2008-2012 he was with RWTH Aachen University as assistant professor and research associate of RWTH's center of excellence on Ultra-high speed Mobile Information and Communication (UMIC). Since November 2012, he has been with the Electrical Engineering and Computer Science School, KTH Royal Institute of Technology, Stockholm, where he is professor for machine-to-machine communications. At KTH, James served as director for the ACCESS Linnaeus Centre from 2016 to 2019, while he is currently associate director in the newly formed KTH Digital Futures research center, as well as co-director in the newly formed VINNOVA competence center on Trustworthy Edge Computing Systems and Applications (TECoSA). His research interests are in the area of mobile systems and networks, with a focus on critical machine-to-machine communications, edge computing, resource allocation, as well as performance evaluation. He has authored over 150 (peer-reviewed) papers in international journals and conferences. His work has been awarded multiple times, including the best paper awards at ACM MSWiM 2015, IEEE WoWMoM 2009 and European Wireless 2009. In 2007, he was the recipient of the ITG/KuVS dissertation award for his Ph.D. thesis.