SN Computer Science (2020) 1:56
https://doi.org/10.1007/s42979-019-0056-2

SURVEY ARTICLE

G-Networks and the Performance of ICT with Renewable Energy

Yunxiao Zhang[1]

Received: 4 November 2019 / Accepted: 11 December 2019 / Published online: 11 January 2020

Abstract

As a consequence of developing information and communication technology that is playing a significant role in our society and has changed our life dramatically, we witnessed a significant increase in energy consumption in computer systems and networks. Subsequently, energy harvesting technologies with renewable energy are of great interest in the field of computer systems and networks, and thus lead to abundant research which has been carried out to address energy harvesting from different aspects. However, the majority of them focuses on wireless or small-scale networks, which left wired networks with a general structure neglected. We first present a comprehensive systemic review of the trends of overall energy consumption, and energy and quality of service optimization in computer systems and networks. Then, this paper reviews the recent research progress in G-networks and energy packet networks with renewable and intermittent energy from both the system paradigms and the performance optimization and energy reduction algorithms for the wired networks.

Keywords ICT · Energy harvesting · Energy packet networks · G-networks · QoS optimization · Renewable energy

Introduction

The recent growth of information and communication technology (ICT) has led to a rapid increase in energy consumption. The challenge of achieving energy-efficient ICT has, therefore, been extensively explored, including energy harvesting (EH) technologies with renewable sources of energy. The EH technologies are of great interest because they can potentially reduce the dependence on the electricity grid and use "clean" renewable energy in computer systems and networks. However, the intermittent nature of renewable energy makes it challenging to develop and evaluate the energy harvesting model. Thus, abundant research has been carried out to address challenges in energy harvesting from different aspects.

Recently, various communication networks powered by EH have been considered and investigated. Throughput maximization and transmission completion time minimization in single-user channel were studied in [73, 98, 114]. The model of energy cooperation, which can transmit energy between users wirelessly, was introduced in [68]. An optimal policy for multiple access channels with intermittent data and energy arrivals was investigated in [113]. Then, the model was extended to being energy cooperative in [67]. In [4], an online power scheduling policy was characterized to maximize a general utility function. The delay minimization problem was considered in [5]. In [15], a general framework has been developed to maximize the amount of transmitted data with battery capacity constraints and battery imperfections regarding energy leakage. Then, both energy costs in transmission and processing were considered in [96] to maximize throughput. The survey that summarized the recent advances and results of energy harvesting technologies could be found in [66, 88, 107].

Some of the previous work assumed that channel state information and energy harvesting profiles could be perfectly known [73, 98, 114]. Such systems can easily characterize the system state (both the channel state and the energy level), so as to derive the optimal distribution policies. However, the deviation between predicted energy profile and actual output power may enlarge model mismatch in the long term. Hence, recent work is focusing on the stochastic models in which the energy harvesting process is modeled as a random process. On the other hand, most of energy harvesting models consider optimization of quality of service (QoS) and energy efficiency for small-scale networks with one

✉ Yunxiao Zhang
yunxiao.zhang15@imperial.ac.uk

1 Intelligent Systems and Networks Group, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK

single node or two nodes [68, 73, 98, 114]. Although some work has also considered large-scale networks with energy harvesting, such as ad hoc networks [74] or cellular networks [75], they are mainly wireless networks. More attention should also be given to wired networks with general structures that can consume massive power but have been largely neglected.

The energy packet networks (EPN) [35, 39, 41, 63, 81, 84, 118] is a discrete state-space modeling framework that can analyze the interaction between discrete energy flows and job flows (or packets) in a single system. Energy packets (EPs), jobs and data packets are the customer classes in the EPN. One EP is a fixed amount of energy stored in batteries or energy stores (ESs), each of which is modeled as a "queue of energy", while jobs or data packets are normal customers (also called positive customers) circulating at workstations (WSs), servers or sensor nodes. Some EPN models also use a diffusion model with continuous flows, e.g., diffusion approximation, instead of discrete EPs or jobs moving at random times [2], whereas the two main approaches to the EPN are both discrete. One main approach to the EPN was first initiated in [38] to consider one single wireless node that can collect energy through energy harvesting and reap data through sensing. Subsequently, an analytical paradigm has been developed to analyze the performance of the energy harvesting wireless sensor node using a Markov chain representation [81, 84]. Additionally, a new product-form solution (PFS) has been found for tandem networks [84]. The work in [38, 81, 84] assumed that transmission is instantaneous, i.e., it takes zero time on the time scales to sense data, and harvest energy. This approach could be of great interest in autonomous digital devices operating with energy harvesting from intermittent sources.

Another approach to the EPN is based on G-networks [23, 31, 47, 70, 71, 94] that is able to consider more general network structures and the service time for both job processing and energy consumption. G-networks with positive and negative customers were first inspired by the research on biophysical neural networks that can communicate through impulse signals with random emitting intervals, known as random neural networks [26, 33]. Impulse signals traveling through the neural networks are analogous to customers traveling through queueing networks. At receiving neurons (servers), impulse signals are recognized as either "excitations" (positive customers) which add one potential of the neuron (the state of the server) or "inhibitions" (negative customers) which cancel one potential of the neuron (the state of the server), or have no effect if potential is zero. A remarkable property of G-networks is the PFS, which is the steady-state joint probability distribution of the number of positive customers at queues. The work of the EPN exploits G-networks to analyze and evaluate a multi-server system's performance or QoS, for instance, the average response time of jobs [41, 63], energy efficiency and average response time of jobs [118] or energy reserve [41].

The remaining paper is organized as follows. We first summarize the current state of energy usage in computer systems and networks in Sect. 2. Next, Sect. 3 reviews the recent development of energy-efficient ICT with various approaches. Then, we review the theoretical background of G-networks in Sect. 4; the research trends and challenges of the EPN are discussed in Sect. 5. The distinct approaches to the EPN are in Sect. 6. Finally, conclusions are in Sect. 7.

Energy Consumption in Computer Systems and Networks

The world's increasing usage of ICT inevitably led to a rise in energy consumption by ICT. Over the last decade, we have seen several reports that discuss how rapidly growing ICT will unsustainably consume a large amount of the overall available energy.

Some of the early estimations of the global carbon dioxide emissions and energy consumption [24, 25, 110] alerted us to the risk that energy consumption related to ICT would grow at a rapid rate. Recently, its carbon footprint has become comparable to that of air transportation, which is at approximately 1.4% of the total emission of carbon dioxide [92]. There are also more rigorous studies [93, 100] which have suggested that ICT consumed approximately 3.9% of global electricity and contributed to approximately 1.3% of global greenhouse gas emissions in 2007. It was also projected that ICT's electricity consumption would grow to about 3766 TWh by 2020, representing a 156% increase since 2008.

With these early estimations and studies from before 2010, one may easily conclude that ICT would consume a radically increasing amount of the overall world energy in the next decades, which will eventually hinder global energy sustainability. However, [8, 40, 92] evaluated the impact of different sectors of ICT on energy consumption and carbon dioxide emissions and concluded that growth of energy and carbon footprint of ICT has slowed down in the last decade and is now decreasing because of improved energy-efficient material used in devices and "smart system technology". According to some recent and detailed reports [89, 108] in 2015, ICT electricity consumption in 2012 was 920 TWh or 4.7% of the 19,500 TWh, representing global electricity consumption. In the latest report published in 2018 [92], ICT electricity consumption in 2015 reduced to 805 TWh out of 21,000 TWh global electricity consumption. Comparing 920 TWh with 3766 TWh or 805 TWh with 3766 TWh, we realize there is a significant deviation between the early projected consumption and actual data. In addition, it also suggested that ICT is playing an essential role in increasing the energy efficiency of a large number of activities. We should also consider ICT's positive impact on the reduction of overall energy consumption, for instance, in the house and transportation, or by smart management of the power grid.

Even though energy consumption of ICT is now decreasing, 805 TWh is a large quantity of electricity, and everything must be done to reduce energy consumption by ICT further.

Energy and QoS Optimization

Previous work [8, 34, 52] has discussed the development of energy-efficient ICT with various approaches for different system layers. One of the major challenges is to explore the relations among system components and the trade-off that can result in an optimal balance between performance, QoS, and energy consumption. [8].

This challenge has been explored in many aspects. In the rest of this section, we comprehensively review the energy reduction and QoS optimization in ICT from various perspectives: data centres, energy–QoS trade-off in cloud computing, QoS and energy-aware routing, and energy harvesting with intermittent energy sources.

In recent years, massive data centers have become the backbone of the Internet, leading to a huge amount of energy usage. In [86], the author mentioned that data centers used approximately 0.5% of total world electricity consumption in 2005, and that this number increased to 1.1% in 2015 [92]. Although the energy needs of ICT are being fed by the fast-rising demand for data and the increasing amount of installed hardware, data center electricity consumption has been growing moderately since increased Internet traffic, and loads are offset by much increased efficiency including ultra-efficient data centers and the development of "smart system technology", for instance, cloud computing that shares resources (processor and other hardware) to prevent the servers remaining idle [51, 101, 109].

Regarding ultra-efficient data centers, the well-known Koomey's law [87] which describes a long-term trend in the development of computing hardware, concludes that the amount of computation per kilowatt-hour has doubled approximately every 1.52 years. This trend was remarkably stable and was faster than Moore's Law since the 1950s. Re-examined by Koomey in 2011, the doubling period has slowed to every 2.6 years since 2000, from which Moore's Law has also slowed as the rate of progress would reach saturation. As reported in [102], the energy usage of data centers in the United States can potentially decline by 25% if four-fifths of servers in small data centers are transferred to new hyper-scale ultra-efficient data centers. Presented in [8], ICT energy consumption depends primarily on CPU utilization. However, it is surprisingly hard to achieve high average levels of utilization of typical servers, and the utilization of CPUs is even worse with personal computers and laptops. Moreover, the most common operating mode observed in [7] (10–50% of the maximum utilization levels) unfortunately corresponds to the lowest energy-efficiency region in many current servers. Thus cloud computing, which has attracted much attention, is a promising approach to improve the utilization of data center resources and to save energy. Key ideas include moving services to those servers which operate at the most energy-efficient regime, reducing the number of redundant servers and turning unused servers off, or into "sleeping" mode to achieve energy efficiency. The hotspots generated at data centers should also be measured when a node or server is used excessively, and services or jobs can be transferred to some other nodes with a lower temperature or lower loads to avoid hotspots, and also reduce the overhead expense on cooling.

Because of the development of mobile networks and the IoT, in recent work, energy-aware routing protocols for wireless networks have also drawn much attention. The ad hoc cognitive packet networks in [48] develop an intelligent environment for seeking fast and energy-efficient routes with battery-powered nodes [49]. Topology control was considered in [78], to modify the network graph and optimize QoS. A search algorithm [1] for finding destinations in a random environment with a team of searching packets has explored the interaction of the number of searchers and energy consumption.

However, more attention should also be placed on wired networks that consume massive power and have been largely neglected. A software-based "Energy Management System" for wired packet networks was developed in [59, 60] to reduce energy consumption subject to QoS requirement. In [55], the optimization of a cost function that includes power consumption and QoS metrics was solved using the G-network theory. In other work [56, 57], optimization algorithms were developed to route packets and minimize composite cost functions combining overall network energy consumption and QoS.

Large numbers of heterogeneous digital devices and communication nodes are being incorporated into the Internet [6, 14, 65, 99, 111] to manage cities and various service activities [3, 9, 10, 13, 36, 62, 116], which also leads to more energy consumption by ICT. Such systems must operate autonomously over a long period, and can benefit from collecting energy physically from natural phenomena such as wind, liquid flows, photovoltaic energy, and other renewable energy sources. This has triggered a significant interest in investigations of EH for computer ommunication systems. Much work and research have been conducted to evaluate performance, optimize QoS and energy metrics regarding power consumption and sustainability for such systems [17, 50, 69, 85].

In [90], data transmission with energy harvesting nodes has been characterized by a continuous-time Markov chain, and the optimal online control policy for data was derived by dynamic programming. To reduce the computational complexity of the dynamic programming, in [98], the authors have investigated the simple optimal power transmission for a single channel with a rechargeable battery node. Their work has provided optimal policies to maximize the throughput and minimize the data transmission time. Related work from various perspectives can be found in [15, 73, 96, 114] which assumes, somewhat unrealistically, that channel state information and energy harvesting profiles can be perfectly known, such that statistical knowledge and causal information of energy and channel variation are available to drive the proposed optimal policies.

In other work, a two-user multiple access communication channel [68, 113] was considered, where energy has been discretized in a packet form which can be transferred between users through wireless energy transfer [76], while [67] considered energy harvesting in a two-user cooperative Gaussian multiple access channel. Some of the literature up to 2015 has been reviewed in [107], with emphasis on energy harvesting wireless communication and energy transfer from the perspective of communication and information theory.

Energy harvesting wireless sensor networks are different from general energy harvesting wireless communication systems,because devices need to meet the requirements of data transmission as well as source acquisition, which involves sensing, sampling, and compression, and entails an energy cost compared with that of data transmission. Thus, an optimal energy allocation for the limited energy resource to source acquisition and data transmission is required [64, 72, 79]. The optimal policies that only consider the energy harvesting process and data transmission should be revised to take resource acquisition, for instance, the quality of the measurement taken by the sensor into account. In [11], a single sensor with a signal receiver was considered, where energy harvesting processed in each time slot follows an ergodic stationary process. A case study of multiple sensors is in [106], and sensor reliability was studied in [91]. Moreover, in [97], the optimal resource allocation policy, which is subject to delay constraints, was derived. In [80], it used the energy packet network paradigm (a distinct approach from the EPN paradigms based on G-network theory) to model the wireless sensor networks with energy harvesting, in which a wireless sensor consumes one EP for data sensing and processing and another one EP for data transmission. In [83], The work was extended to the framework that data sensing and processing will consume a variable number of EPs, and data transmission also consumes a variable number of EPs.

The above work focused on point-to-point channels or small-scale systems powered by energy harvesting. Large-scale wireless networks powered by energy harvesting such as mobile ad hoc networks and cellular networks have received some attention recently. The spatial throughput of a mobile ad hoc network with energy harvesting was analyzed in [74]. A queueing model of an energy-efficient wireless base station is presented in [17]. In [75], the authors investigated the coverage of a cellular network, which is affected by the variations in energy sources. Moreover, heterogeneous cellular networks with energy harvesting were modeled in [16].

Recent work has developed the EPN paradigm based on G-network theory [21, 36, 37]. This framework is able to evaluate and analyze the interaction of energy flow and job flow for large-scale wired networks. We note that there is also a distinct approach to the EPN, which does not apply the G-network theory. We review the G-network theory and the EPN approach based on G-networks in the following sections.

G-Networks

G-networks that are advanced queueing networks with positive and negative customers were first introduced in [27, 29] and were inspired by the study of biophyiscal neural networks [26, 28, 32, 61]. This work started in 1990 and has continued until today. A remarkable and useful property of a G-network is the PFS which is the steady-state joint probability distribution of the number of customers at queues [29, 54]. In addition to ordinary customers or positive customers, G-networks have negative customers which can arrive from outside the networks or can be moved from one queue to another. When a negative customer arrives at a queue, the negative customer can "cancel" a positive customer at this queue. If the queue is empty, the negative customer disappears at which it arrives because negative customers cannot accumulate at queues. A single server system with negative and positive customers was studied in [43]. Stability conditions for G-networks were given in [58]. Multiple class models and relevant work were developed in a series of papers [12, 22, 23, 47, 95]. A special type of customer known as a "trigger" which is able to push an existing customer at a queue to another queue was introduced in [30]. Note that only positive customers can be pushed, because neither negative customers nor triggers can wait at a queue. The G-network model that allows to either trigger a customer from one queue to another or remove a batch of customers at this queue was considered in [31].

There are also other G-network models. For instance, G-networks with "resets" [42, 70], with "adders" [20] and with restarts [22, 23], which have been developed and adopted in a number of applications in computer systems modeling and other fields.

In the rest of this section, we briefly review G-networks with positive and negative customers. Finally, we review G-networks with batch removal which are the fundamental basis of the latest EPN paradigms.

Random Neural Networks and G-Networks with Positive and Negative Customers

Here we consider an open network with N servers with mutually independent service times which are modeled as independent and identically distributed (i.i.d.) random variables that follow an exponential distribution with parameter $r(i)$ where $i = 1, \dots, N$. External arrivals at the network can either be positive customers arriving at queue i according to a Poisson process with the rate of λ_i^+ or negative customers arriving at queue i according to a Poisson process with the rate of λ_i^-. We assume a customer who has been served leaves queue i and moves to queue j with probability P_{ij}^+ as a positive customer or with probability P_{ij}^- as a negative customer. The served customer also can be removed from the network with probability d_i. We note that $1 = d_i + \sum_{j=1}^N P_{ij}^+ + P_{ij}^-$ for all $i = 1, \dots, N$, because total probability is one. Thus, the transition probability of the Markov chain can be written as $P_{ij} = P_{ij}^+ + P_{ij}^-$ which represents the customer movements between servers. The queue length at servers is constituted only by positive customers, each of which adds one queue length at which it arrives, while one negative customer that does not require service cancels one positive customer waiting at the queue and reduces queue length by one. The negative customer does not affect the queue if the queue length is zero. Moreover, the queueing discipline is assumed to be the first-come-first-served (FCFS) basis, i.e., customers are served in the order of their arrival.

The network has a set of traffic equations, which is

$$\Lambda_i^+ = \lambda_i^+ + \sum_j^N q_j r(j) P_{ji}^+ \tag{1}$$

$$\Lambda_i^- = \lambda_i^- + \sum_j^N q_j r(j) P_{ji}^-. \tag{2}$$

where

$$q_i = \frac{\Lambda_i^+}{r(i) + \Lambda_i^-}, \quad i = 1, \dots, N, \tag{3}$$

is the utilization of queue i. Here we recall the PFS of G-networks, which is a steady-state joint probability distribution of the number of positive customers at queues. Let $\mathbf{K}(t) = (k_1(t), \dots, k_N(t))$ be the vector of queue lengths at time t, and $\mathbf{k} = (k_1, \dots, k_N)$ be an arbitrary vector of queue length.

Theorem 1 *If a unique non-negative solution of the traffic equations given in (1) and (2) exists such that* $0 < q_i < 1$ *for all* $i = 1, \dots, N$ *hold, then the following PFS:*

$$\lim_{t\to\infty} \Pr[\mathbf{K}(t) = \mathbf{k}] = \prod_{i=1}^N q_i^{k_i}(1 - q_i), \tag{4}$$

exists.

Proof Since $\{\mathbf{K}(t) : t \geq 0\}$ is a continues-time Markov chain that satisfies the Chapman–Kolmogorov Equation, we can verify that the PFS satisfies the global balance equations. Although it is similar to that in Jackson's Theorem [77], the proof is more complicated because of the negative customers canceling the customers waiting at queues and the nonlinearity of traffic equations. Since the proof has been given in many G-network papers, we do not provide the proof in this paper. Interested readers can find details about the proof in [29]. □

Remark 1 If the PFS exists, the marginal steady-state probability distribution of k_i customer at queue i is

$$\begin{aligned}\lim_{t\to\infty} \Pr[K_i(t) = k_i] &= \sum_{j=1, j\neq i}^N \sum_{k_i=1}^\infty \left(\prod_{i=1}^N q_i^{k_i}(1 - q_i) \right), \\ &= q_i^{k_i}(1 - q_i).\end{aligned} \tag{5}$$

Remark 2 If the PFS exists, the steady-state probability distribution of at least k_i customers at queue i is the product of the marginal probability distribution, which is

$$\lim_{t\to\infty} \Pr[K_i(t) \geq k_i] = q_i^{k_i}. \tag{6}$$

G-Networks with Batch Removals

We adapt the framework of G-networks with positive and negative customers to G-networks with batch removals by replacing negative customers with "signals". A signal has no effect on empty queues, while either one of the following two events happens if the signal arrives at a non-empty queue:

1. With probability $0 \leq 1 - D_i \leq 1$, the arriving signal is "trigger" which instantaneously moves one customer at the head of queue i to queue j with probability M_{ij} where $1 = \sum_{j=1}^N M_{ij}$ for $i = 1, \dots, N$.
2. With probability D_i, the arriving signal, which is "negative customer" instantaneously removes a batch of customers with size up to X_i. The size X_i is a random variable following a given probability distribution $\pi_i(s)$ where

$$\pi_i(s) = \Pr[X_i = s], \quad s = 1, 2, \ldots, \tag{7}$$

and

$$\sum_{s=1}^{\infty} \pi_i(s) = 1, \quad \forall i = 1, \ldots, N. \tag{8}$$

Thus, the average number of customers removed by one single "negative customer" signal is

$$E[X_i] = \sum_{s=1}^{\infty} s\pi_i(s), \tag{9}$$

If the queue length is less than X_i at time t, i.e., $X_i > k_i(t)$, the queue length becomes zero at the next time step t^+ after service. It can be expressed by

$$k_i(t^+) = \begin{cases} k_i(t) - X_i, & \text{if } k_i(t) \geq X_i, \\ 0, & \text{otherwise}. \end{cases} \tag{10}$$

The traffic equations of G-networks with batch removals are

$$\Lambda_i^+ = \lambda_i^+ + \left(\sum_{j=1}^{N} q_j r(j) \left[P_{ji}^+ + \sum_{m=1}^{N} P_{jm}^-(1 - D_m) q_m M_{mi} \right] + \lambda_j^-(1 - D_j) q_j M_{ji} \right), \tag{11}$$

$$\Lambda_i^- = \lambda_i^- + \sum_{j=1}^{N} q_j r(j) P_{ji}^-, \tag{12}$$

where

$$q_i = \frac{\Lambda_i^+}{r(i) + \Lambda_i^-[(1 - D_i) + D_i f_i(q_i)]}, \tag{13}$$

$$f_i(q_i) = \frac{1 - \sum_{s=1}^{\infty} \pi_i(s) q_i^s}{1 - q_i}, \quad \text{for } i = 1, \ldots, N, \tag{14}$$

Regarding Eq. (11), the left side of the equation represents the total arrival rate of positive customers, which is equal to the sum of the following terms:

- The external arrival rate of positive customers, represented by the first term of the right side of the equation.
- The arrival rate of positive customers from other nodes, some of which are served positive customer moving to node i according to the transition matrix $\mathbf{P}^+ = [P_{ji}^+]$. Other positive customers moving into node i are triggered by the trigger from node j according to the transition matrix $\mathbf{P}^- = [P_{jm}^-]$. This rate is represented by the second term of the right side of the equations.
- The arrival rate of positive customers from other nodes at which positive customers are triggered by trigger signals arriving from the external, represented by the third term of the right side of the equation.

Regarding Eq. (12), the left side of the equation represents the total arrival rate of signals which can be either "triggers" or "negative customers". This rate equals to the sum of the following terms:

- The external arrival rate of signals, represented by the first term of the right side of the equation.
- The total arrival rate of signals from other nodes where some served positive customers become signals. This rate is represented by the second term of the right side of the equation.

Since each server has a different ability to process customers, the probability distribution $\pi_i(s)$ for each server may not be the same. The q_i represents the utilization of queue i, and also the probability that queue i is not empty. Then we have the function $f_i(\cdot)$ which is related to the average number of customer that one "negative customer" signal can remove, and it depends on the state of the queue.

The PFS given in Theorem 1 is also valid for G-networks with batch removals. The uniqueness and existence of the PFS is guaranteed if a unique non-negative solution of (13) exists and we have $0 < q_i < 1$ for all $i = 1, \ldots, N$. Interested readers can find the proof in [31]

Energy Packet Networks Derived from G-Networks

The concept of the EPN begins with the integration of adaptive electrical ESs charged by renewable energy sources and distributed consumption systems such as WSs, servers, or sensors. In addition, the EPN paradigms use each EP to represent the fixed amount of energy in Joules, which can also be viewed as a pulse of power that lasts a certain time [36]. The amount of energy in such an EP can be small enough to be close to the smallest energy needs of consumers [36, 39, 41, 84], or large enough to be a significant quantity to power large energy consumers [115]. In the latest research [63, 117, 118], the number of jobs or data packets that can be executed by one single EP is a random variable. Thus if a WS is energy-efficient, it will execute more jobs with a single EP. This relaxed the previous assumption that an EP is the smallest amount of energy needs of consumers, i.e., an EP was used to process only a single job. Thus, the EPN offers a more explicit representation of the interaction between energy flow and job or data flows in computer networks powered by intermittent energy sources. Such EPN paradigms are developed for an environment where renewable sources are common, and energy storage facilities or devices are available [36].

The EPN was first inspired by investigating random arrivals of energy from intermittent energy sources, and random arrivals of jobs or data packets in one single system. Some classes of EPN can be described by G-network theory. Such EPN systems have an important and convenient property which is the existence of a product form of steady-state distribution of jobs in WSs and EPs in the ESs, namely the PFS. With the PFS, it is computationally conventional to evaluate, and so as to improve the QoS and energy efficiency.

It is worthy to notice that there is another approach to the EPN paradigm which is not related to the G-networks. These EPN paradigms, which are not based on G-networks, are associated with various stochastic models. Some EPN use diffusion approximation with continuous flows, instead of discrete EPs or jobs moving at random times [2]. A distinct approach to the discrete EPN in [38, 44, 45, 84] was developed for the devices with "zero service time". For instance, a device or sensor may process a data packet in nanoseconds or microseconds, and also consume an EP in milliseconds, which are negligibly small.

Note the EPN is not only a theoretical approach. Other independent researches [103–105] have proposed the "power packet" system, which is a practical hardware-based design for switching power and dispatching data packets simultaneously.

In the following sections, we review the recent development of the EPN paradigms comprehensively. First, we review the initial work of EPN based on G-networks where one EP can be used to process one single job, and its relevant applications and problems. Next, we review the latest model of the EPN, where the number of jobs that can be processed by one single EP is a random variable.

G-Networks with Negative Customers of the EPN and its Optimizations

The general structure of the EPN[1] has been considered in [41] which is modeled as a queuing network with a finite set of nodes. The networks has m ES nodes denoted as E_a where $1 \leq a \leq m$, and has n WS nodes denoted as W_i where $1 \leq i \leq n$. Jobs that must be executed in WSs are ordinary customers in the queueing network. They arrive at one of n WS nodes, say W_i according to the Poisson process at a rate of λ_i jobs/s. WS W_i may also receive a job from another WS, saying W_j with probability M_{ji} after WS W_j finishing process of that job. The energy from an intermittent external source is discretized to EPs which arrive at one of m ES nodes, say E_a according to Poisson process at a rate of γ_a EPs/s. Similarly, ES E_a may also transmit one EP to another ES E_b with probability P_{ab} to balance energy distribution.

Here, the number of EPs at E_a is denoted as $L_a(t)$, and the number of jobs at W_i is denoted as $K_i(t)$ at time t. It is assumed that both WS nodes and ES nodes have a queue with infinity capacity. EPs at ES E_a are expanded because of energy leakage, or moved to WSs for job processing. When ES E_a is not empty at time t (i.e., $L_a(t) > 0$), it has the following manner:

1. The successive EP leakage times at ES E_a are modeled as i.i.d. random variables following a common exponential distribution with parameter δ_a.
2. or ES E_a can forward one EP either to another ES to balance energy distribution, or to one corresponding WS to power job processing. The successive EP forwarding times at ES E_a are also i.i.d. random variables that follows a common exponential distribution with parameter w_a.
   (a) With probability F_{ai}, the EP stored in ES E_a was forwarded to WS W_i for job processing.
   (b) As mentioned P_{ab} is the probability that ES E_a sends energy to another ES E_b to balance energy distribution. Thus, we have $1 = \sum_{i=1}^{n} F_{ai} + \sum_{b=1}^{m} P_{ab}$.
3. In summary, energy leaks at a rate of δ_a EPs/s, or is forwarded to another ES or WSs at a rate of w_a EPs/s. Thus, the number of EP at E_i is one less (i.e., $L_a(t^+) = L_a(t) - 1$) after the average time of $(\delta_a + w_a)^{-1}$ seconds.

Jobs stored at the WS W_i may be lost at time out without the need for energy. The successive jobs time-out times are also modeled as i.i.d. random variables following a common exponential distribution with parameter β_i. However, we note that processing jobs at WS W_i needs energy arrival from one of the ES nodes. When an EP arrives at WS W_a which is not empty at time t (i.e., $K_i(t) > 0$), the following happens:

1. With probability d_i, the job that has completed processing is removed from WS W_a without forwarding to other WSs because the job terminates at W_a.
2. With probability M_{ij}, the job that has completed processing at WS W_i is forwarded to WS W_j for further more processing, where $1 = d_i + \sum_{j=1}^{n} M_{ij}$.

[1] We slightly abuse the term of "EPN" to refer "EPN based on G-networks" in the following sections.

The EPN model discussed above corresponds to a multi-class G-network with negative customers and $(m+n)$ queues (m queues are ESs and n queues are WSs). The network has two classes customers ($C=2$) where Class 1 refers to the jobs to be processed in the WSs, and Class 2 refers to the EPs to power job processing. The jobs are positive customers at WS queues, and the EPs are positive customers at ES queues. However, one EP becomes a negative customer or a trigger when it arrives at a WS queue. The EP is expended to finish one job processing at one WS, saying W_i. After the job has completed its processing, the job will be forwarded to another WS node, for instance W_j, for more processing with probability M_{ij} where the EP behaves as a trigger, or the job will be removed from the network with probability $d_i = 1 - \sum_{i=j}^{n} M_{ij}$ where the EP behaves as a negative customer.

Specifically, traffic equations of network for each queue can be represented using G-network theory:

$$\Lambda^+_{1,i} = \lambda_i + \sum_{j=1}^{n}\sum_{a=1}^{n} q_{2,a} w_a F_{aj} q_{1,j} M_{ji}, \quad i = 1, \ldots, n, \tag{15}$$

$$\Lambda^-_{1,i} = \sum_{a=1}^{m} q_{2,a} w_a F_{ai}; \quad r(1,i) = \beta_i \quad i = 1, \ldots, n, \tag{16}$$

$$\Lambda^+_{2,a} = \gamma_a + \sum_{b=1}^{m} q_{2,b} w_b P_{ba}, \quad a = 1, \ldots, m, \tag{17}$$

$$\Lambda^-_{2,a} = 0; \quad r(2,i) = w_a + \delta_a \quad a = 1, \ldots, m, \tag{18}$$

where

$$q_{1,i} = \frac{\Lambda^+_{1,i}}{r(1,i) + \Lambda^-_{1,i}}, \quad i = 1, \ldots, n, \tag{19}$$

$$q_{2,i} = \frac{\Lambda^+_{2,i}}{r(2,i) + \Lambda^-_{2,i}} \quad a = 1, \ldots, m. \tag{20}$$

are the utilization of WS queues and ES queues. Moreover, $q_{1,i}$ denotes the probability that WS W_i has at least one job in its queue, and $q_{2,a}$ denotes the probability that ES E_a has at least one EP in its queue.

1. The $\Lambda^+_{1,i}$ denotes the total effective arrival rate of jobs at WS W_i. The first term of right-hand side of (15) refers to the job arriving from external. The second term refers to the arrival rate of jobs from other WSs. These jobs have completed processing at other WS, and require more process at WS W_i.
2. The $\Lambda^-_{1,i}$ and $r(1, i)$ denotes the total effective rate of job leaving from WS W_i. The term $r(1, i)$ refers to the rate of job lost at time out without the need for energy. The term $\Lambda^-_{1,i}$ refers to the job processing rate, which requires the arrival of EPs from ES nodes. These jobs that have completed processing at WS W_i are to be forwarded to other WS nodes for more processing, or to be removed without forwarding to other WS nodes.
3. The $\Lambda^+_{2,a}$ denotes the total effective arrival rate of EPs at ES E_a. The first term of right-hand side of (17) refers to the rate of EPs arrives from the external intermittent energy sources. The second term refers to the arrival rate of EPs from other ES nodes for energy balance.
4. The $\Lambda^-_{2,a}$ and $r(2, i)$ denotes the total effective rate of EPs left from ES E_a. The first term of $r(2, i)$ refers to the rate of EPs that is sent to either other ES nodes or WS nodes. The second term refers to the rate of energy lost because of leakage. Since energy cannot be destroyed, the term $\Lambda^-_{2,a}$ is null.

The above traffic equations have the following PFS result:

Result 1 *Let $k_1, \ldots, k_n$ represent the backlogs of jobs to be processed at the WS nodes, and $l_1, \ldots, l_m$ represent the number of EPs stored at the ES nodes. If a unique non-negative solution of the traffic equations in (15) to (18) exists such that $q_{1,i}$, $q_{2,a} \in (0, 1)$ for all $i = 1, \ldots, n$ and $a = 1, \ldots, m$ hold, then the following PFS*

$$\lim_{t\to\infty} \Pr[K_1(t) = k_1, \ldots, K_n(t) = k_n, L_1(t) = l_1, \ldots, \\ L_m(t) = l_m] = \prod_{i=1}^{n} q_{1,i}^{k_i}(1 - q_{1,i}) \prod_{a=1}^{m} q_{2,a}^{l_i}(1 - q_{2,a}), \tag{21}$$

exists.

Relevant Applications and Optimization Problems

Utility Functions and Optimizations

With the PFS which gives the explicit expression of the steady-state distribution of jobs in WS queues and EPs in ES queues, it is computationally conventient to evaluate the performance and the energy efficiency of the system. In [41], the authors proposed the relevant utility functions that can evaluate the performance and the energy efficiency:

1. We may wish to limit the backlog of job at WS queues, and we also want to reserve energy as much as possible in ES queues for unpredictable needs. Thus, a sensible utility function is

$$U_1 = \sum_{i=1}^{n} \alpha_i q_{1,i}^{k_i} + \sum_{a=1}^{m} \beta_a \left(1 - q_{2,a}^{l_a}\right), \tag{22}$$

where $\alpha_i > 0$ and $\beta_a > 0$ are weights. This is a sum of the weighted probability that the backlog of jobs exceeds the value k_i and the weighted probability that the number of EPs is less than l_a. By minimizing this utility function, it is able to reduce the probability of backlog of jobs and increase the probability of more energy reserved at ES nodes. When $k_i = 1$ and $l_a = 1$, the utility function becomes

$$U_1^0 = \sum_{i=1}^{n} \alpha_i q_{1,i} + \sum_{a=1}^{m} \beta_a (1 - q_{2,a}). \tag{23}$$

To minimize this special function, we want to reserve energy as much as possible and reduce the backlog of jobs as many as possible.

2. A very similar utility function considers the average response time of jobs waiting at WS nodes rather than the backlog of jobs. It is

$$U_2 = \sum_{i=1}^{n} \alpha_i \frac{(\Lambda_{1,i}^{-})^{-1}}{1 - q_{1,i}} + \sum_{a=1}^{m} \beta_a \left(1 - q_{2,a}^{l_a}\right), \tag{24}$$

3. The third utility function considering the throughput of the system and reserved energy needs to be maximized:

$$U_3 = \sum_{a=1}^{m} \left(\sum_{i=1}^{n} \alpha_i q_{2,a} F_{ai} q_{1,i} + \beta_a q_{2,a}^{l_a} \right), \tag{25}$$

The optimization of the EPN can be expressed as the minimization or maximization of these utility functions with respect to the control variables that are P_{ab}, M_{ij} and F_{ai}. By selecting the optimal value of these control variables within the constrains (i.e., $1 = \sum_{i=1}^{n} F_{a,i} + \sum_{b+1}^{m} P_{ab}$, $1 = d_i + \sum_{j=1}^{n} M_{ij}$ and $0 < q_{1,i} < 1$, $0 < q_{2,a} < 1$ for all $i = 1, \ldots, n$ and $a = 1, \ldots, m$), the minimization or maximization of the utility function can be found. In [41], the gradient descent is a useful tool to solve the optimization problems, because these utility function are continuous and differentiable. Readers who are interested in how to use the gradient descent to solve the optimization problems please refer to [41] for the details.

The EPN Used for Mobile Networks with Energy Harvesting

The EPN approach has been used for system analysis of a backhaul multi-hop connection of a wireless mobile network with energy harvesting [39]. The data or other traffic are carried out as data packets (DPs) rather than jobs. These DPs arrive from outside the network in the form of a random process representing the data or other traffic generated by users. DPs stored in data buffers can enter the next node or leave the system. Thus, this EPN has a tandem structure that considers a multi-hop connection traversing the nodes $N_1, \ldots, N_n$, which can be base stations, routers, or WiFi nodes operating with intermittent harvested energy.

It is assumed that node N_1 receives end-to-end user traffic (voice or SMS) which travels over n node to its destination. Thus, this is a mobile connection transmitting DPs along the full path of nodes $N_1, \ldots, N_n$. If a DP cannot be forwarded from node N_i to node N_{i+1}, $1 \le i \le n-1$ or node N_n cannot transmit the DP to the external network, the DP is lost and needs to be retransmitted from the first node N_1 with probability p. End user generates fresh user traffic at an aggregate rate of λ DPs/s at node N_1 according to a Poisson process. Each node N_i, $1 \le i \le n$ also receives a cross-traffic Poisson flow Λ_i DPs/s according to Poisson process. The external arrival processes are assumed to be independent of each other.

Each node N_i is powered by an ES charged from intermittent energy sources at rate of γ_i^H EPs/s, or also from the grid at rate γ_i^G EPs/s, so that the total energy supply rate at the nodes is $\gamma_1, \ldots, \gamma_n$, where $\lambda_i = \lambda_i^H + \lambda_i^G$, $i = 1, \ldots, n$. The energy (EPs) received by a node is first stored in an ES of unlimited capacity. The ES at node N_i leaks at rate δ_i, so that the number of EPs stored at the battery is one less after the exponentially distributed time of an average of δ_i^{-1}. Note that the nodes are efficient that they only use energy when they process and forward a DP, and one EP is the amount of energy consumed by the node to process and forward one DP.

Using the approach in EPN paradigms, the probability q_i that the battery of node N_i has at least one EP is

$$q_i = \frac{\gamma_i}{\delta_i + \Lambda_i + \lambda_i}. \tag{26}$$

Note that this formula does not require that the arrival of EPs follows a Poisson process, nor that the EP inter-leakage and depletion times are exponentially distributed. Depletions of the battery from leakage and energy consumption can be analyzed without limiting assumptions of Poisson arrivals, and using stationary point process theory with the general distribution of inter-arrival and service times.

Three cases of practical interests regarding the impact of intermittent energy on the packet loss and delay have been analyzed using this model:

1. The first case considered the type of systems which has been suggested in [112]. The system is very fast so that data buffering is not needed, and the equipment's power consumption is directly proportional to the load, which goes to "sleep" mode as there is no work to do. Thus, it

is assumed that the backhaul network is so fast that no buffering of DPs is needed. The system is also so efficient such that each node only consumes energy when DPs are processing or forwarding, while no energy is consumed when a node is "sleeping". It assumes one DP is lost when the DP attempts to transit through a node that has run out of energy (i.e., battery is depleted, and its energy source cannot provide power).
2. The second case assumes that nodes store packets until they can be forwarded, and the nodes are kept on even when they have no packets to forward. Thus, only DPs in transit may be lost, while DPs in the memory of a node is preserved.
3. The third case is the "worst case". It is similar to the second case, while DPs stored in memory will be lost when that node has run out of energy.

G-Network with Batch Removals of the EPN and Its Optimization Problems

The latest work modified the EPN model by adapting G-networks with batch removals to model a multi-server system consisting of servers of WSs which are powered by an ES that is charged from an intermittent source. It uses energy flow and job flow together to optimize a multi-server system's performance or QoS, for instance, the average response time of jobs. Throughout the work, it relaxed the assumption of the previous EPN paradigms that one EP can only be used to execute one single job. The number of jobs can be executed by one EP is a random variable of X_i following general probability distribution.

The EPN consists of N WSs and N ESs. Jobs that must be executed in the system are modeled as ordinary customers in a queueing network. They arrive at one of the N WSs, say WS i, at a rate of λ_i jobs/s. Each WS is represented as a queue containing jobs. Jobs first arrive at a given WS i. Each WS i has an energy storage battery denoted by ES i, and there are a total of N ESs. EPs arrive from an external intermittent energy source at rate γ_i EPs/s to the ES i which can be viewed as a "queue of EPs".

In the EPN model, the EPs in ES i either can be forwarded to the corresponding WS i on demand with probability d_i, or moved to another ES j with probability P_{ij} to balance the energy distribution. The jobs in WS i can be processed locally with probability D_i or forwarded to some other WS j with probability M_{ij} for further steps of execution. In this model, w_i is the rate at which EPs are forwarded from ES i to one of the ESs or the corresponding WS. On the other hand, δ_i is the loss rate of energy (i.e., leakage) from ES i.

It denotes the number of jobs at WS i by $K_i(t)$ and the number of EPs at ES i by $B_i(t)$ at time t. It assumes that both the WS queues, and the ES queues (i.e., batteries) are unbounded, i.e., infinite capacity. EPs at ES i are expended in the following manner:

If ES i is not empty, i.e., $B_i(t) > 0$, ES i will:

1. Either leak energy at some rate $\delta_i \geq 0$ EPs/s, and after a time of average value δ_i^{-1}, we will have one less EP at ES i due to energy leakage. The successive EP leakage times for ES i are modeled as i.i.d. random variables having a common exponential distribution with parameter δ_i.
2. Or ES i will forward one EP at rate w_i EPs/s to WS i or other ESs. The successive times (that one EP is forwarded) for ES i are also modeled as i.i.d. random variables having a common exponential distribution with parameter w_i.

When an EP is forwarded to WS i, this EP is used locally by WS i as follows:

1. With probability $0 \leq D_i \leq 1$, if $K_i(t) > 0$, one EP will be expended to serve a batch of up to X_i jobs at the WS i in one step. After service, we end up with
$$K_i(t^+) = \begin{cases} K_i(t) - X_i, & \text{if } K_i(t) \geq X_i, \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$
Since each job may have different energy requirements at WS i, we assume that the number of jobs that can be processed with a single EP at WS i is a random variable.
2. Since our purpose is to model different WSs that have different levels of energy efficiency, a single EP is used to process one or more jobs, if there are jobs waiting in the WS queue.
3. With probability $1 - D_i$, if $K_i(t) > 0$, one EP will be used to serve just one job, and then forward that job to another WS j according to the transition probability matrix $\mathbf{M} = [M_{ij}]$. As a result, we will have $K_i(t^+) = K_i(t) - 1$ and $K_j(t^+) = K_j(t) + 1$. In the mathematical model, the transition matrix allows the job to return to the same workstation, i.e., the diagonal entries are not null. However, it is not physically meaningful because it wastes one EP to move the job, which is at the head of the WS queue to the tail of the WS queue. Thus, we assume that the diagonal entries of the transition probability matrix are null, such that the jobs cannot return to the same WS.
4. If an EP arrives at an empty WS i, i.e., $K_i(t) = 0$, the EP will just be expended to keep the WS in working order (i.e., to keep it on), and no jobs will be processed or moved.

Since the EPN model discussed above is a special case of the G-network with batch removals and multiple classes of customers, it can be analyzed with regard to G-network theory.

The corresponding expressions for the EPN model for Class 1 customers (jobs) and Class 2 customers (EPs) are given by the following expressions, for $i = 1, \dots, N$:

$$q_{1,i} = \frac{\Lambda^+_{1,i}}{q_{2,i+N} w_i d_i[(1-D_i) + D_i \frac{1-\sum_{s=1}^{\infty} q^s_{1,i}\pi_i(s)}{1-q_{1,i}}]}, \tag{28}$$

$$q_{2,i+N} = \frac{\gamma_i + \sum_{j=1}^{N} w_j q_{2,j+N} P_{ji}}{w_i + \delta_i}, \tag{29}$$

where

$$\Lambda^+_{1,i} = \lambda_i + \sum_{j=1}^{N} q_{1,j}(1-D_j) d_j w_j M_{j,i} q_{2,j+N}.$$

Note that the $q_{c,i}$ denotes the steady-state probability that queue i has at least one job (if $c = 1$) or one EP (if $c = 2$). For $i = 1, \dots, N$, we have $q_{1,i} \geq 0$ and $q_{2,i} = 0$ because EPs cannot wait at WSs; and we have $q_{2,i+N} \geq 0$ and $q_{1,i+N} = 0$ because jobs cannot wait at ESs.

Because the EPN model we have described is a special case of a G-network with two classes of customers, namely jobs for Class 1 and EPs for Class 2, we can directly apply the PFS given in G-network theory to the EPN model. For this case, i.e., where we model an EPN, each of the queues is either a WS or an ES. WSs only contain Class 1 customers, and ESs only contain Class 2 customers. Therefore, the PFS of the EPN models is

Result 2 *If the traffic equations given in (28) and (29) have a unique solution such that all $q_{1,i}$ and $q_{2,i+N}$ lie between 0 and 1 for $i = 1, \dots, N$, the following PFS holds:*

$$\begin{aligned} &\lim_{t\to\infty} \Pr[\mathbf{K}(t) = (k_{1,1}, \dots, k_{1,N}, k_{2,N+1}, \dots, k_{2,2N})] \\ &= \prod_{i=1}^{N} (q_{1,i})^{k_{1,i}}(1-q_{1,i})(q_{2,i+N})^{k_{2,i+N}}(1-q_{2,i+N}). \end{aligned} \tag{30}$$

From Remarks 1 and 2, we can have the marginal queue length probability distribution for queue i which are, for $c = 1, 2$:

$$\lim_{t\to\infty} \Pr[K_{c,i}(t) = k_{c,i}] = (q_{c,i})^{k_{c,i}}(1-q_{c,i}), \tag{31}$$

$$\lim_{t\to\infty} \Pr[K_{c,i}(t) \geq k_{c,i}] = (q_{c,i})^{k_{c,i}}. \tag{32}$$

Optimization Problems

Specifically, the work uses the EPN paradigm to address relevant problems of practical interest to optimize performance and energy efficiency.

1. Problem 1 investigates how to select the optimal fraction of power that is shared between heterogeneous servers, so as to minimize the average response time of jobs. Problem 1 was solved analytically using the Lagrange multiplier method. In addition, a physically meaningful condition was obtained to guarantee system stability and optimality.
2. Problem 2 investigates how to select the optimal fraction of power that is shared between heterogeneous servers, so as to minimize the average response time of jobs. Problem 2 minimizes the average response time of jobs by dynamically deciding whether to move jobs between servers, so as to balance the workload at each server. Problem 2 is solved numerically through gradient descent.
3. In Problem 3, it considers a cost function that combines both the average response time and the rate of energy loss. It selects the optimal fraction of power shared between heterogeneous servers to minimize the cost function. The optimal solution was obtained by solving a system of equations simultaneously.
4. Problem 4 investigates how to match energy flow into energy buffers and job flow into the corresponding WSs, so as to minimize the average response time of jobs. Through the investigation, the optimal solution must follow a necessary condition that the fraction of energy flow to ES i must align with the fraction of jobs flows WS i.

Distinct Approaches to the Energy Packet Network

Energy Packet Network with Zero Service Time

A distinct approach to the EPN was first initiated in [38] in which the author considered a single wireless node that can collect energy through energy harvesting and reap data through sensing. In the EPN based on G-networks, we consider "service time" for both energy consumption and job processing. However, in the distinct approach, it considers devices with "zero service time". For instance, a sensor may process a packet of data in nanoseconds or microseconds, and also consume an EP in milliseconds, which are negligibly small. Such a system is unstable if the data buffer and energy storage have infinite capacities, while an explicit expression for the joint probability distribution of the number of data packets and EPs was derived in [38] when both buffers are finite. Moreover, with respect to such a system, the author provided an analytical analysis for the mathematical model in which data flows and energy flows can be viewed as instantaneously synchronized flows. Developed in [53], the analysis was extended for two-node systems.

Based on the early work of [44, 45], an analytical model was used to analyze the performance of the energy harvesting wireless sensor nodes using a Markov chain representation. A generalized model of a data transmission system with energy harvesting was considered in [82] where a data packet transmission needs more than one EP. It also investigated the probability that a receiver correctly receives a packet in the presence of noise. The work of [80] models the wireless sensor networks that source acquisition (data sensing and processing) consumes one EP, and data transmission also consumes exactly one EP. The work was extended in [83] where a variable number of EPs will be consumed by data sensing and processing, and data transmission, respectively. In [84], the authors derived a new PFS, which is a distinct approach from G-network, for a tandem network of N nodes using harvested energy stored in batteries. The effect of battery attacks at the nodes of energy provisioning systems was investigated in [46]. A simple mitigation technique was proposed for a wireless node with a renewable or a replaceable battery to drop a fraction of the traffic to prolong battery lifetime.

Energy Packet Networks with a Multiple Class Extension and Service Disciplines

Another distinct approach to the EPN considers multiple classes of DPs, where each class of a DP has an independent routing matrix and determines the number of EPs needed to be sent [18, 19]. In this approach, it assumes the DP queue (or WS) is the initiator of the transfer, which means the arrival of a DP at the corresponding ES triggers the movement of EPs (which are consumed). If the energy level at the ES is not enough to power the arrived DP, the DP is lost. Moreover, one of the three following types: the FCFS, the last-come-first-serve with preempt and resume (LCFS-PR) and the processor sharing (PS) is the service discipline for DP queues.

The sufficient conditions for the stability of this EPN approach with an arbitrary topology have been provided. Moreover, such an EPN also possesses the product form of the steady-state distribution of DPs in the queues provided that a stable solution of a fixed point problem exists. Therefore, the distinct approach can be used for a new type of optimization problems for the use of energy based on the fragmentation of data with several routes.

Conclusion

In the paper, we have briefly surveyed the current state of energy usage in ICT, and also challenges and opportunities corresponding to developing EH technologies on the computer and network systems when EH can be used to replace or complement power supply from the grid, and relevant problems and applications.

Although the growth of energy consumption and the carbon footprint of ICT has decreased, it is still a large quantity of electricity consumption. Thus, a series of researches were carried out to achieve energy-efficient ICT. However, the majority of them focused on wireless or small-scale networks because of the boom of IoT, which resulted in wired networks being largely neglected.

Then, we reviewed a novel EPN-based approach to analyze the interaction of energy flow and job flow in multi-server systems. The approach to the EPN we reviewed in the paper is based on the G-network theory, which is an advanced queueing network with negative customers. We first review the theoretical background of G-network, then we discussed the initial and latest development, challenges and applications of the EPN.

We expect that there will be further work on the development and analysis of energy-efficient ICT, especially the wired networks that operate with renewable and intermittent energy sources. We hope this present paper can provide some insight into the operation and design of such systems.

Compliance with ethical standards

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Abdelrahman OH, Gelenbe E. Time and energy in team-based search. Phys Rev E. 2013;87(3):032125.
2. Abdelrahman OH, Gelenbe E. A diffusion model for energy harvesting sensor nodes. In: 2016 IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS), IEEE; 2016. pp. 154–158.
3. Al-Ali A, Aburukba R. Role of internet of things in the smart grid technology. J Comput Commun. 2015;3(05):229.
4. Arafa A, Baknina A, Ulukus S. Online fixed fraction policies in energy harvesting communication systems. IEEE Trans Wireless Commun. 2018;17(5):2975–86. https://doi.org/10.1109/TWC.2018.2805336.

5. Arafa A, Tong T, Fu M, Ulukus S, Chen W. Delay minimal policies in energy harvesting communication systems. IEEE Trans Commun. 2018;66(7):2918–30. https://doi.org/10.1109/TCOMM.2018.2805357.
6. Atzori L, Iera A, Morabito G. The internet of things: a survey. Comput Netw. 2010;54(15):2787–805.
7. Barroso LA, Hölzle U. The case for energy-proportional computing. Computer. 2007;40(12):33–7. https://doi.org/10.1109/MC.2007.443.
8. Berl A, Gelenbe E, Di Girolamo M, Giuliani G, De Meer H, Dang MQ, Pentikousis K. Energy-efficient cloud computing. Comput J. 2010;53(7):1045–51.
9. Bi H, Abdelrahman OH. Energy-aware navigation in large-scale evacuation using g-networks. Probab Eng Inf Sci. 2018;32(3):340–52. https://doi.org/10.1017/S0269964816000115.
10. Brun O, Yin Y, Gelenbe E. Deep learning with dense random neural network for detecting attacks against iot-connected home environments. Procedia Comput Sci. 2018;134:458–63.
11. Castiglione P, Simeone O, Erkip E, Zemen T. Energy management policies for energy-neutral source-channel coding. IEEE Trans Commun. 2012;60(9):2668–78.
12. Chao X, Pinedo M. On generalized networks of queues with positive and negative arrivals. Probab Eng Inf Sci. 1993;7(3):301–34.
13. Collen A, Nijdam NA, Augusto-Gonzalez J, Katsikas SK, Giannoutakis KM, Spathoulas G, Gelenbe E, Votis K, Tzovaras D, Ghavami N, Volkamer M. Ghost-safe-guarding home IoT environments with personalised real-time risk control. Cham: Springer; 2018. p. 68–78.
14. Da Xu L, He W, Li S. Internet of things in industries: a survey. IEEE Trans Ind Inf. 2014;10(4):2233–43.
15. Devillers B, Gündüz D. A general framework for the optimization of energy harvesting communication systems with battery imperfections. J Commun Netw. 2012;14(2):130–9. https://doi.org/10.1109/JCN.2012.6253061.
16. Dhillon HS, Li Y, Nuggehalli P, Pi Z, Andrews JG. Fundamentals of heterogeneous cellular networks with energy harvesting. IEEE Trans Wirel Commun. 2014;13(5):2782–97.
17. Dimitriou I, Alouf S, Jean-Marie A. A Markovian queueing system for modeling a smart green base station. In: European Workshop on Performance Engineering. Springer, New York; 2015. pp. 3–18.
18. Doncel J, Fourneau JM. Balancing energy consumption and losses with energy packet network models. In: 2019 IEEE International Conference on Fog Computing, ICFC 2019, Prague, Czech Republic, June 24–26. IEEE; 2019. pp. 59–68.
19. Doncel J, Fourneau JM. Energy packet networks with multiple energy packet requirements. Probab Eng Inf Sci. 2019:1–19. https://doi.org/10.1017/S0269964819000226
20. Fourneau JM, Gelenbe E. G-networks with adders. Future Internet. 2017;9(3):34.
21. Fourneau JM, Marin A, Balsamo S. Modeling energy packets networks in the presence of failures. In: 2016 IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS); 2016. pp. 144–153 . https://doi.org/10.1109/MASCOTS.2016.44
22. Fourneau JM, Wolter K. Some applications of multiple classes g-networks with restart. In: Computer and Information Sciences—31st International Symposium, ISCIS 2016, Kraków, Poland, October 27–28, 2016, Proceedings; 2016. pp. 126–133 . https://doi.org/10.1007/978-3-319-47217-1_14
23. Fourneau JM, Wolter K, Reinecke P, Krauß T, Danilkina A. Multiple class g-networks with restart. In: ACM/SPEC International Conference on Performance Engineering, ICPE'13, Prague, Czech Republic—April 21–24, 2013; 2013. pp. 39–50 https://doi.org/10.1145/2479871.2479880
24. Gartner GI. The new industry shockwave. In: Gartner Symposium, ITxpo. 2007.
25. Gartner Estimates ICT. Industry accounts for 2 percent of global CO2 emissions. Press release. 2007.
26. Gelenbe E. Random neural networks with negative and positive signals and product form solution. Neural Comput. 1989;1(4):502–10.
27. Gelenbe E. Réseaux neuronaux aléatoires stables. Comptes rendus de l'Académie des sciences. Série 2 Mécanique Phys Chim Sci de l'univers Sci de la Terre. 1990;310(3):177–180
28. Gelenbe E. Stability of the random neural network model. Neural Comput. 1990;2(2):239–47.
29. Gelenbe E. Product-form queueing networks with negative and positive customers. J Appl Probab. 1991;28(3):656–63. https://doi.org/10.2307/3214499.
30. Gelenbe E. G-networks by triggered customer movement. J Appl Probab. 1993;30(3):742–8.
31. Gelenbe E. G-networks with signals and batch removal. Probab Eng Inf Sci. 1993;7(3):335–42.
32. Gelenbe E. Learning in the recurrent random neural network. Neural Comput. 1993;5(1):154–64.
33. Gelenbe E. G-networks: a unifying model for neural and queueing networks. Ann Oper Res. 1994;48(5):433–61.
34. Gelenbe E. Steps toward self-aware networks. Commun ACM. 2009;52(7):66–75. https://doi.org/10.1145/1538788.1538809.
35. Gelenbe E. Energy packet networks: ICT based energy allocation and storage. In: International Conference on Green Communications and Networking. Springer, Berlin, Heidelberg; 2011. pp. 186–195.
36. Gelenbe E. Energy packet networks: adaptive energy management for the cloud. In: Proceedings of the 2nd International Workshop on Cloud Computing Platforms. ACM, New York; 2012.
37. Gelenbe E. Energy packet networks: smart electricity storage to meet surges in demand. In: Proceedings of the 5th International ICST Conference on Simulation Tools and Techniques, pp. 1–7. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). 2012.
38. Gelenbe E. Synchronising energy harvesting and data packets in a wireless sensor. Energies. 2015;8(1):356–69. https://doi.org/10.3390/en8010356. http://www.mdpi.com/1996-1073/8/1/356.
39. Gelenbe E, Abdelrahman OH. An energy packet network model for mobile networks with energy harvesting. Nonlinear Theory Appl IEICE. 2018;9(3):1–15.
40. Gelenbe E, Caseau Y. The impact of information technology on energy consumption and carbon emissions. Ubiquity. 2015;2015(June):1.
41. Gelenbe E, Ceran ET. Energy packet networks with energy harvesting. IEEE Access. 2016;4:1321–31.
42. Gelenbe E, Fourneau JM. G-networks with resets. Perform Eval. 2002;49(1):179–91.
43. Gelenbe E, Glynn P, Sigman K. Queues with negative arrivals. J Appl Probab. 1991;28(1):245–50.
44. Gelenbe E, Kadioglu YM. Energy loss through standby and leakage in energy harvesting wireless sensors. In: 2015 IEEE 20th International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD). IEEE; 2015. pp. 231–236.
45. Gelenbe E, Kadioglu YM. Performance of an autonomous energy harvesting wireless sensor. In: Information Sciences and Systems 2015. Springer, Cham; 2016. pp. 35–43.
46. Gelenbe E, Kadioglu YM. Energy life-time of wireless nodes with network attacks and mitigation. In: 2018 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE; 2018. pp. 1–6.

47. Gelenbe E, Labed A. G-networks with multiple classes of signals and positive customers. Eur J Oper Res. 1998;108:293–305.
48. Gelenbe E, Lent R. A power-aware routing algorithm. Simul Ser. 2003;35(4):502–7.
49. Gelenbe E, Lent R. Power-aware ad hoc cognitive packet networks. Ad Hoc Netw. 2004;2(3):205–16.
50. Gelenbe E, Lent R. Energy-qos trade-offs in mobile service selection. Future Internet. 2013;5(2):128–39.
51. Gelenbe E, Lent R, Douratsos M. Choosing a local or remote cloud. In: 2012 Second Symposium on Network Cloud Computing and Applications. IEEE; 2012. pp. 25–30.
52. Gelenbe E, Lent R, Nunez A. Self-aware networks and qos. Proc IEEE. 2004;92(9):1478–89.
53. Gelenbe E, Marin A. Interconnected wireless sensors with energy harvesting. In: International Conference on Analytical and Stochastic Modeling Techniques and Applications; Springer, Cham; 2015. pp. 87–99.
54. Gelenbe E, Mitrani I. Analysis and synthesis of computer systems, vol. 4. Singapore: World Scientific; 2010.
55. Gelenbe E, Morfopoulou C. Routing and G-networks to optimise energy and quality of service in packet networks. In: International Conference on Energy-Efficient Computing and Networking, Springer, Berlin, Heidelberg; 2010. pp. 163–173.
56. Gelenbe E, Morfopoulou C. A framework for energy-aware routing in packet networks. Comput J. 2011;54(6):850–9.
57. Gelenbe E, Morfopoulou C. Power savings in packet networks via optimised routing. Mobile Netw Appl. 2012;17(1):152–9.
58. Gelenbe E, Schassberger R. Stability of product form g-networks. Probab Eng Inf Sci. 1992;6(3):271–6.
59. Gelenbe E, Silvestri S. Optimisation of power consumption in wired packet networks. In: International conference on heterogeneous networking for quality, reliability, security and robustness. Springer, Berlin, Heidelberg; 2009. pp. 717–729.
60. Gelenbe E, Silvestri S. Reducing power consumption in wired networks. In: 2009 24th International symposium on computer and information sciences. IEEE; 2009. pp. 292–297.
61. Gelenbe E, Timotheou S. Random neural networks with synchronized interactions. Neural Comput. 2008;20(9):2308–24.
62. Gelenbe E, Wu FJ. Future research on cyber-physical emergency management systems. Future Internet. 2013;5(3):336–54.
63. Gelenbe E, Zhang Y. Performance optimization with energy packets. IEEE Syst J. 2019. https://doi.org/10.1109/JSYST.2019.2912013
64. Georgiadis L, Neely MJ, Tassiulas L. Resource allocation and cross-layer control in wireless networks. Found Trends Netw. 2006;1(1):1–144. https://doi.org/10.1561/1300000001.
65. Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of things (iot): a vision, architectural elements, and future directions. Future Gen Comput Syst. 2013;29(7):1645–60.
66. Gündüz D, Stamatiou K, Michelusi N, Zorzi M. Designing intelligent energy harvesting communication systems. IEEE Commun Mag. 2014;52(1):210–6. https://doi.org/10.1109/MCOM.2014.6710085.
67. Gurakan B, Kaya O, Ulukus S. Energy and data cooperative multiple access channel with intermittent data arrivals. IEEE Trans Wirel Commun. 2018;17(3):2016–28. https://doi.org/10.1109/TWC.2017.2787744.
68. Gurakan B, Ozel O, Yang J, Ulukus S. Energy cooperation in energy harvesting communications. IEEE Trans Commun. 2013;61(12):4884–98.
69. Guruacharya S, Hossain E. Self-sustainability of energy harvesting systems: concept, analysis, and design. IEEE Trans Green Commun Netw. 2018;2(1):175–92.
70. Harrison PG. G-networks with propagating resets via RCAT. SIGMETRICS Perform Eval Rev. 2003;31(2):3–5. https://doi.org/10.1145/959143.959144.
71. Harrison PG, Pitel E. Response time distributions in tandem G-networks. J Appl Probab. 1995;32(1):224–46. https://doi.org/10.2307/3214932.
72. He Z, Wu D. Resource allocation and performance analysis of wireless video sensors. IEEE Trans Circuits Syst Video Technol. 2006;16(5):590–9.
73. Ho CK, Zhang R. Optimal energy allocation for wireless communications with energy harvesting constraints. IEEE Trans Signal Process. 2012;60(9):4808–18. https://doi.org/10.1109/TSP.2012.2199984.
74. Huang K. Spatial throughput of mobile ad hoc networks powered by energy harvesting. IEEE Trans Inf Theory. 2013;59(11):7597–612.
75. Huang K, Kountouris M, Li VO. Renewable powered cellular networks: Energy field modeling and network coverage. IEEE Trans Wirel Commun. 2015;14(8):4234–47.
76. Huang K, Lau VK. Enabling wireless power transfer in cellular networks: architecture, modeling and deployment. IEEE Trans Wirel Commun. 2014;13(2):902–12. https://doi.org/10.1109/TWC.2013.122313.130727.
77. Jackson JR. Jobshop-like queueing systems. Manage Sci. 1963;10(1):131–42.
78. Jia X, Li D, Du D. Qos topology control in ad hoc wireless networks. In: IEEE INFOCOM 2004, IEEE; 2004. vol. 2, pp. 1264–1272
79. Joseph V, Sharma V, Mukherji U, Kashyap M. Joint power control, scheduling and routing for multicast in multihop energy harvesting sensor networks. In: 2009 International Conference on Ultra Modern Telecommunications & Workshops. IEEE; 2009. pp. 1–8.
80. Kadioglu YM. Energy consumption model for data processing and transmission in energy harvesting wireless sensors. In: International symposium on computer and information sciences. Springer; 2016. pp. 117–125.
81. Kadioglu YM. Finite capacity energy packet networks. Probab Eng Inf Sci. 2017;31(4):477–504.
82. Kadioglu YM, Gelenbe E: Packet transmission with k energy packets in an energy harvesting sensor. In: Proceedings of the 2nd international workshop on energy-aware simulation. ACM; 2016. p. 1.
83. Kadioglu YM, Gelenbe E. Wireless sensor with data and energy packets. In: 2017 IEEE International Conference on Communications Workshops (ICC Workshops); 2017. pp. 564–569. https://doi.org/10.1109/ICCW.2017.7962718
84. Kadioglu YM, Gelenbe E. Product-form solution for cascade networks with intermittent energy. IEEE Syst J. 2019;13(1):918–27. https://doi.org/10.1109/JSYST.2018.2854838.
85. Kansal A, Hsu J, Zahedi S, Srivastava MB. Power management in energy harvesting sensor networks. ACM Trans Embed Comput Syst. 2007. https://doi.org/10.1145/1274858.1274870.
86. Koomey JG. Worldwide electricity used in data centers. Environ Res Lett. 2008;3(3):034008.
87. Koomey JG, Berard S, Sanchez M, Wong H. Implications of historical trends in the electrical efficiency of computing. IEEE Ann Hist Comput. 2011;33(3):46–54. https://doi.org/10.1109/MAHC.2010.28.
88. Ku ML, Li W, Chen Y, Liu KR. Advances in energy harvesting communications: Past, present, and future challenges. IEEE Commun Surv Tutor. 2016;18(2):1384–412. https://doi.org/10.1109/COMST.2015.2497324.
89. Lannoo B, Lambert S, Van Heddeghem W, Pickavet M, Kuipers F, Koutitas G, Niavis H, Satsiou A, Beck MT, Fischer A, de Meer

H. Overview of ICT energy consumption. Network of Excellence in Internet Science; 2013. pp. 1–59
90. Lei J, Yates R, Greenstein L. A generic model for optimizing single-hop transmission policy of replenishable sensors. IEEE Trans Wirel Commun. 2009;8(2):547–51.
91. Lei L, Kuang Y, Shen XS, Yang K, Qiao J, Zhong Z. Optimal reliability in energy harvesting industrial wireless sensor networks. IEEE Trans Wirel Commun. 2016;15(8):5399–413.
92. Malmodin J, Lundén D. The energy and carbon footprint of the global ICT and e&m sectors 2010–2015. Sustainability. 2018;10(9):3027.
93. Malmodin J, Moberg Å, Lundén D, Finnveden G, Lövehagen N. Greenhouse gas emissions and operational electricity use in the ICT and entertainment & media sectors. J Ind Ecol. 2010;14(5):770–90.
94. Marin A. Product-form in G-networks. Probab Eng Inf Sci. 2016;30(3):345–60. https://doi.org/10.1017/S0269964816000048.
95. Matalytski M. Finding expected revenues in g-network with multiple classes of positive and negative customers. Probab Eng Inf Sci. 2018. https://doi.org/10.1017/S0269964818000013
96. Orhan O, Gündüz D, Erkip E. Throughput maximization for an energy harvesting communication system with processing cost. In: 2012 IEEE Information Theory Workshop; 2012. pp. 84–88. https://doi.org/10.1109/ITW.2012.6404770
97. Orhan O, Gündüz D, Erkip E. Delay-constrained distortion minimization for energy harvesting transmission over a fading channel. In: 2013 IEEE International Symposium on Information Theory. IEEE; 2013. pp. 1794–1798.
98. Ozel O, Tutuncuoglu K, Yang J, Ulukus S, Yener A. Transmission with energy harvesting nodes in fading wireless channels: optimal policies. IEEE J Sel Areas Commun. 2011;29(8):1732–43.
99. Perera C, Liu CH, Jayawardena S, Chen M. A survey on internet of things from industrial market perspective. IEEE Access. 2014;2:1660–79. https://doi.org/10.1109/ACCESS.2015.2389854.
100. Pickavet M, Vereecken W, Demeyer S, Audenaert P, Vermeulen B, Develder C, Colle D, Dhoedt B, Demeester P. Worldwide energy needs for ICT: The rise of power-aware networking. In: 2008 2nd International Symposium on Advanced Networks and Telecommunication Systems, IEEE; 2008. pp. 1–3.
101. Quan DM, Basmadjian R, De Meer H, Lent R, Mahmoodi T, Sannelli D, Mezza F, Telesca L, Dupont C. Energy efficient resource allocation strategy for cloud data centres. In: Computer and information sciences II, Springer, London; 2011. pp. 133–141.
102. Shehabi A, Smith SJ, Sartor DA, Brown RE, Herrlin M, Koomey JG, Masanet ER, Horner N, Azevedo IL, Lintner W, United states data center energy usage report. Lawrence Berkeley National Laboratory; 2016.
103. Takahashi R, Azuma Si, Tashiro K, Hikihara T. Design and experimental verification of power packet generation system for power packet dispatching system. In: 2013 American Control Conference. IEEE; 2013. pp. 4368–4373.
104. Takahashi R, Takuno T, Hikihara T. Estimation of power packet transfer properties on indoor power line channel. Energies. 2012;5(7):2141–9.
105. Takahashi R, Tashiro K, Hikihara T. Router for power packet distribution network: design and experimental verification. IEEE Trans Smart Grid. 2015;6(2):618–26.
106. Tapparello C, Simeone O, Rossi M. Dynamic compression-transmission for energy-harvesting multihop networks with correlated sources. IEEE/ACM Trans Netw. 2014;22(6):1729–41.
107. Ulukus S, Yener A, Erkip E, Simeone O, Zorzi M, Grover P, Huang K. Energy harvesting wireless communications: a review of recent advances. IEEE J Sel Areas Commun. 2015;33(3):360–81. https://doi.org/10.1109/JSAC.2015.2391531.
108. Van Heddeghem W, Lambert S, Lannoo B, Colle D, Pickavet M, Demeester P. Trends in worldwide ict electricity consumption from 2007 to 2012. Comput Commun. 2014;50:64–76.
109. Wang L, Gelenbe E. Adaptive dispatching of tasks in the cloud. IEEE Trans Cloud Comput. 2018;6(1):33–45. https://doi.org/10.1109/TCC.2015.2474406.
110. Webb M. Smart 2020: Enabling the low carbon economy in the information age. The Climate Group, London; 2008. pp. 1–1.
111. Whitmore A, Agarwal A, Da Xu L. The internet of things: a survey of topics and trends. Inf Syst Front. 2015;17(2):261–74.
112. Wu J, Zhang Y, Zukerman M, Yung EK. Energy-efficient base-stations sleep-mode techniques in green cellular networks: a survey. IEEE Commun Surv Tutor. 2015;17(2):803–26. https://doi.org/10.1109/COMST.2015.2403395.
113. Yang J, Ulukus S. Optimal packet scheduling in a multiple access channel with energy harvesting transmitters. J Commun Netw. 2012;14(2):140–50. https://doi.org/10.1109/JCN.2012.6253062.
114. Yang J, Ulukus S. Optimal packet scheduling in an energy harvesting communication system. IEEE Trans Commun. 2012;60(1):220–30. https://doi.org/10.1109/TCOMM.2011.112811.100349.
115. Yin Y. Optiaml energy for energy packet networks. Probab Eng Inf Sci. 2017;31(4):516–39.
116. Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M. Internet of things for smart cities. IEEE Internet Things J. 2014;1(1):22–32.
117. Zhang Y. Energy and performance optimisation with energy packet networks (Doctoral dissertation, Imperial College London). 2019. https://doi.org/10.25560/73892
118. Zhang Y. Optimal energy distribution with energy packet networks. Probab Eng Inf Sci. 2019. https://doi.org/10.1017/S0269964818000566

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.