G-NETWORKS AND THEIR APPLICATIONS TO MACHINE LEARNING, ENERGY PACKET NETWORKS AND ROUTING: INTRODUCTION TO THE SPECIAL ISSUE

Mehmet Ufuk Caglayan
Department of Computer Engineering
Boğaziçi University
Bebek, Istanbul, Turkey
E-mail: caglayan@boun.edu.tr

This paper introduces a special issue of this journal (Probability in the Engineering and Informational Sciences) that is devoted to G(elenbe)-Networks and their Applications. The special issue is based on revised versions of some of the papers that were presented at a workshop held in early January 2017 at the Séminaire Saint-Paul in Nice (France). It includes contributions in several research directions that followed from the introduction of the G-Network in the late 1980s. The papers present original theoretical developments, as well as applications of G-Networks to Machine Learning, to the performance optimization of energy systems via the novel Energy Packet Networks formalism for systems that operate with renewable and intermittent energy sources, and to packet network routing and Cloud management over the Internet. We introduce these contributions from the perspective of an overview of recent work based on G-Networks.

Keywords: autonomic networks, cognitive packet networks, E-Networks, energy packet networks, G-Networks, random neural network

1. INTRODUCTION

This special issue spans theoretical developments in G-Networks (Gelenbe [48]), which are a very significant generalization of queueing networks (Gelenbe and Mitrani [125], Gelenbe and Muntz [130]), that was invented by Erol Gelenbe. It also presents applications and extensions of G-Networks to study probability models of energy flows, machine learning, network routing, and task allocation in the Cloud. As such, this special issue fits the scope of this journal on “Probability in the Engineering and Informational Sciences” which is edited by Professor Sheldon M. Ross, an eminent leader in Applied Probability and Operations Research. In addition to this introductory and review paper, the research papers published here are:

- “Finding non-stationary state probabilities of G-Network with signals and batch removal” by Mikhail Matulytski.
- “Analysis of multi-resource loss system with state-dependent arrival and service rate” by Valeriy Naumov and Konstantin Samouylov.
• “Processor sharing G-queues with inert customers and catastrophes: a model for server aging and rejuvenation” by Jean-Michel Fourneau and Youssef Ait Al Mahjoub.
• “Random Neural Network Learning Heuristics” by Abbas Javed, Hadi Larijani, Ali Ahmadiania and Rohnton Emmanuel.
• “Echo state queueing qetworks: a combination of reservoir computing and random neural networks” by Sebastián Basterrech and Gerardo Rubino.
• “Finite capacity Energy Packet Networks” by Yasin Murat Kadıoğlu.
• “A diffusion approach to Energy Packet Networks” by Omer H. Abdelrahman.
• “Optimum energy for Energy Packet Networks” by Yonghua Yin.
• “The Random Neural Network for cognitive traffic routing and task allocation in networks and the Cloud” by Lan Wang.

These papers illustrate new results on both theory and applications, spanning several directions, which have emerged the original work that started the field of G-Networks (Chabridon et al. [18], Fourneau and Gelenbe [31], Fourneau, Gelenbe, and Suros [32], Gelenbe, Glynn, and Sigman [86], Gelenbe and Labed [101], Gelenbe and Shachnai [137]). In these novel queuing networks, customers are either “positive” and enter a queue and then request service from a server, or they are “negative customers” which can be negative customers which destroy a customer (or a batch of customers) from a queue or they are “triggers” (Gelenbe [51]) which move an existing customer to another queue, or they are “resets” (Fourneau and Gelenbe [30]) which replace the state of the queue to its steady-state probability distribution. Normal customers that leave a queue can themselves become any of these types of new customers at a different queue.

1.1. Theoretical Developments

The paper “Finding non-stationary state probabilities of G-Network with signals and batch removal” deals with the time-dependent behavior of the G-Network model with batch removals or “catastrophes” (Gelenbe [50]) where some of the negative customers can remove customers in batches whose size is a random variable.

The second paper on “Analysis of multi-resource loss system with state-dependent arrival and service rate”, also takes on the “batch” idea to analyze a system where negative customers increase the amount of resources available to positive customers. The model is a loss system with multiple types of resources and state-dependent arrival and service rates, where customers require an amount of resources which may be zero, positive or negative for each resource.

The third paper on “Processor sharing G-queues with inert customers and catastrophes: a model for server aging and rejuvenation” innovates by discovering the product-form solution for a queueing network with processor sharing servers and batch removal when all customers are emptied in one batch (often called a “catastrophe”).

2. EXTENSIONS AND APPLICATIONS OF THE RANDOM NEURAL NETWORK (RNN)

The RNN was introduced in (Gelenbe [37,45]) as a biologically inspired model for natural computation (Gelenbe [65]), while the link between spiking neural models and queueing networks was also presented in (Gelenbe and Stafylopatis [140]). Questions regarding
the existence and uniqueness of its equilibrium solution were raised in (Gelenbe [46]), and sufficient conditions for feed-forward networks were given in (Gelenbe and Schassberger [135]). The RNN was applied to the detailed modeling of somato-sensory cortex of rats in (Gelenbe and Cramer [78]).

The RNN is at the origin of G-Networks. Because it was developed to model biological neural networks and Machine Learning, the first known polynomial (in this case of computational complexity $O(N^3)$ for a network with $N$ neurons) gradient descent learning algorithm for arbitrarily interconnected “recurrent” neural networks was based on the RNN (Gelenbe [52]). It was extended to learning in “multiple class” models that can represent multi-sensory inputs or color images in (Gelenbe and Hussain [91]). Other learning paradigms were studied in (Gelenbe [47]). The RNN learning algorithm was applied to video compression (Cramer and Gelenbe [22], Gelenbe et al. [141]), to recognize textures (Gelenbe and Feng [79]), and tumors in magnetic resonance images of the human brain (Gelenbe, Feng, and Krishnan [80]). Other applications of the random neural network that do not require learning include function optimization (Gelenbe, Koubi, and Pekergin [99]) and texture generation (Atalay and Gelenbe [9], Atalay, Gelenbe, and Yalabik [10]). Applications of the RNN were published for video compression (Cramer, Gelenbe, and Bakircioglu [20,21]), complex recognition tasks (Abdelbaki, Gelenbe, and El-Khamy [1], Abdelbaki, Gelenbe, and Kocak [2], Abdelbaki et al. [3], Aguilar and Gelenbe [8], Gelenbe, Ghanwani, and Srinivasan [85], Hocaoglu et al. [155]), and to the sensory search of patterns and objects (Gelenbe and Cao [74], Gelenbe and Kocak [97], Gelenbe, Kocak, and Wang [98]). A polynomial time-complexity learning algorithm for RNNs having soma-to-soma interactions was first presented in (Gelenbe and Timotheou [142]) and is further developed in (Wang and Gelenbe [184]). The fact that a finite RNN is a universal approximator for bounded and continuous function was shown in (Gelenbe, Mao, and Li [40,121,122]) and the RNN was exploited for deep learning in (Gelenbe and Yin [150]).

G-Networks have been used to model Gene Regulatory Networks (Gelenbe [58]) leading to an interesting technique to detect anomalies in genetic data (Kim and Gelenbe [158,159,161], Kim et al. [163], Kim, Park, and Gelenbe [165]). In this work, a known set of gene interactions and their time constants are modeled by a G-Network whose parameters are first estimated from “normal” micro-array data (Kim and Gelenbe [160,162]). Then the model is used for comparison with other micro-array data, to determine possible anomalies (Kim, Atalay, and Gelenbe [157], Kim, Park, and Gelenbe [164]). The RNN has also been used (Phan, Stemberg, and Gelenbe [174]) to identify alignments in protein interaction networks. In other work, some of the underlying chemistry of gene regulation is analyzed in (Gelenbe [59]).

In a quite different area, a G-Network based approach (Filippoupolitis and Gelenbe [27], Gelenbe, Hussain, and Kaptan [90], Gelenbe, Kaptan, and Wang [95]) that represents the fast changing conditions during an emergency were used in a simulator (Dimakis, Filippoupolitis, and Gelenbe [25]) for emergency evacuation planning, and decentralized algorithms that do not require permanent infrastructures were considered in (Filippoupolitis, Gorbil, and Gelenbe [29]). Here, the RNN was used to design fast decision algorithms based on learning a wide range of problem instances (Gelenbe and Timotheou [143]), and then using it for rapidly selecting the best match to the current need.

Thus, the paper “Random Neural Network Learning Heuristics” by Abbas Javed, Hadi Larijani, Ali Ahmadinia and Rohnton Emmanuel, describes the various methods that can be used to select the parameters or “weights” of the RNN so that learning of a given set of data relationships is carried out optimally and efficiently, and the authors comprehensively compare a wide variety of methods, including Gelenbe’s original gradient descent algorithm (Gelenbe [52]) both with regard to learning the original data and to generalization.
with unknown data. On the other hand, the work on “Echo state queueing qetworks: a combination of reservoir computing and random neural networks” combines a recurrent RNN as the means to store data, with an efficient learning sub-network that is connected to the RNN. The approach is illustrated as a means to “learn” several well-known dynamical system models.

3. E-NETWORKS AND ENERGY PACKET NETWORKS (EPNS)

Energy consumption is obviously a critical issue in battery powered mobile communications (Gelenbe and Lent [104]), but it is also critical when massive energy is being used in data centers (Berl et al. [12]) and by information technology as a whole (Gelenbe and Caseau [75]). Thus, G-Networks have been used to analyze and optimize, for instance via dynamic routing, the energy consumption in packet networks (Gelenbe and Morfopoulou [126,128,129], Morfopoulou, Sakellari, and Gelenbe [169], Sakellari, Morfopoulou, and Gelenbe [178]). Indeed, the energy-per-packet or energy-per-task consumption of networking equipment such as routers, or of servers, depends in a non-linear manner on the overall load that is being carried by the equipment Gelenbe and Lent [103, 106, 107]. Dynamic routing, both for packets and tasks, determines how the different system elements are being loaded (Brun, Wang, and Gelenbe [16], Wang and Gelenbe [180]), so that it will have a significant impact not only on performance, but also on energy consumption.

The work on energy consumption (Gelenbe [56]) in Wireless Ad-Hoc Networks contributed a technique to extend the battery lifetime of a multi-hop network by using paths whose nodes have the fullest batteries. It was pursued by work that uses network routing and connection admission as a means to save energy (Gelenbe and Mahmoodi [118], Gelenbe, Mahmoodi, and Morfopoulou [120], Gelenbe and Morfopoulou [127], Gelenbe and Silvestri [138,139], Gelenbe and Wu [145], Sakellari et al. [179]). Energy optimization of software systems was discussed in (Pernici et al. [173]) and its effect on performance trade-offs in remote accesses to Cloud services is analyzed in (Gelenbe, Lent, and Douratsos [109]). In wireless communications, the power level for transmissions can also be chosen so as to minimize the energy consumed per correctly received packet or bit, depending on the interference from other similar transmitters, and on noise (Gelenbe and Gunduz [89], Gelenbe and Lent [108], Gelenbe and Oklander [132], Oklander and Gelenbe [171]).

3.1. E-Networks

As with mobile communications, wireless sensors have a critical need for power and it is often impossible to connect them to the power grid, or to change batteries at regular intervals when the sensors are in remote locations (Gelenbe et al. [84]). In such cases, they need to be powered through renewable sources of energy such as photovoltaic, vibrations, or the ambient electromagnetic transmissions coming from other wireless devices such as mobile base stations, which is the subject of the approach called E-Networks (Gelenbe [68]). In this type of model, both the data collection process and the energy collection process are random, and energy queues are batteries or capacitors, while data buffers are used to store the data. A node in an E-Network can only function if there is both at least one unit of data (a packet) and a unit of stored energy (Gelenbe [70]) leading to interesting stability problems related to the build-up of backlogs when one or the other type of resource (energy and data) is lacking.

In (Gelenbe and Marin [123]), such systems with multiple hops for the data are considered (but not for the energy, which must be harvested separately at each hop), while in
(Gelenbe and Kadioglu [93,94], Kadioglu and Gelenbe [156]) data transmission errors due to noise and wireless interference, repeated transmissions, finite batteries and finite data buffers are also taken into consideration.

At a nano-scopic level, when single particles or packets can carry both energy or data – as with electrons whose spin is used to encode one or more bits – the stochastic nature of the sources of particles, the physical non-homogeneities of the medium in which the particles move, the noise that can affect their spin, as well as mutual influences between spins, need to be modeled Gelenbe [67], and despite several steps forward (Gelenbe [69,71]) much work remains to be done.

In this special issue, the paper “Finite capacity Energy Packet Networks” surveys results that were recently obtained on E-Networks to compute the performance characteristics, including bit error probabilities, of a single wireless sensor that operates with harvested energy, where the amount of energy required per transmission, the effect of wireless interference from multiple identically operating devices, the effect of noise, energy leakage from the battery, and the finite capacity of data buffers and of the battery, are all represented.

The next paper on “A diffusion approach to Energy Packet Networks” replaces the discrete modeling approach by one where the energy flow into the sensor device, and the energy consumption, are modeled as a continuous process, while the data packets (DP) are represented as discrete quantities.

### 3.2. Energy Packet Networks

As with E-Networks, an EPN represents the flow of energy in discrete units, together with the discrete items of work, for example, data packets in a communication network, tasks in a Cloud computer system, within the same queueing network model. The network that will then contain at least two types of customers: for instance, energy packets (EP), and DP or tasks as first suggested in (Gelenbe [62,63]). A continuous space approximation for such models was discussed in (Abdelrahman and Gelenbe [7]).

Here electrical energy storage devices such as batteries or capacitors are also represented as energy queues. Similarly the model can represent other forms of energy storage, for instance reservoirs of petrol, or bins containing compressed air to drive generator turbines, also as queues within a large system. The approach can be readily generalized: for instance we may model different forms of energy within the same system, as well as electricity originating from different sources (e.g., the power grid or photovoltaic). Thus, we must distinguish them with respect to economic cost and origin: photovoltaic electricity may be cheaper than energy from the grid, and it is more valuable because it is cleaner. Here models descended from G-Networks (Gelenbe [49,53]) become useful Abdelrahman and Gelenbe [5], Gelenbe [64].

This approach is adopted in (Gelenbe [66]) where a framework for adaptively allocating these different sources of electricity to a set of computer servers was suggested. Recent work has considered how an electricity dispatching system can be designed to deal with different computer architectures (Gelenbe and Ceran [76]), while in (Gelenbe and Ceran [77]) an approach using G-Networks was developed to select the proportion of different sources of electricity (such as wind and photovoltaic) to different consumers (e.g., security sensors, computational resources, communication devices) so as to minimize a composite utility function that includes the relative importance of the consumers and the availability of the different sources of energy. In (Ceran and Gelenbe [17]) an approximate computational algorithm for such problems is discussed. Thus the work on “Optimum energy for Energy Packet Networks” develops fast algorithms that fit the profile of energy needs and costs for...
different energy sources, to the consumers needs and profiles, using a variety of non-linear optimization techniques that have also been used in Machine Learning.

4. AUTONOMIC SYSTEMS AND COGNITIVE PACKET NETWORKS (CPNS)

Simplified analytical models of subsystems, such as memory management (Gelenbe [41], Gelenbe and Iasnogorodski [92]), link-control procedures (Boguslavskij and Gelenbe [15], Gelenbe, Labetoulle, and Pujolle [102]), or database optimization (Coffman, Gelenbe, and Plateau [19]), had seemed possible (Gelenbe and Pujolle [133]). However, the sheer complexity of computer systems and networks even early on in the 1970s and 1980s has forced us to consider the adaptive control of computer systems and networks (Badel et al. [11], Gelenbe [44], Gelenbe and Kurinckx [100], Gelenbe and Mitrani [124], Potier, Gelenbe, and Lenfant [175]), where the challenge is to deal both with the very large size of the systems, the lack of accurate dynamic models to describe the way they work, and the strong time dependence of their workloads. Thus, in the face of their complexity, industry has encouraged the study of Autonomic or Self-Organized Computer Systems and Networks (Dobson et al. [26], Gelenbe [57]).

The CPN is a patented (Gelenbe [55]) scheme to implement Self-Organized routing for packet networks. It was discussed in (Gelenbe and Lent [39], Gelenbe, Xu, and Seref [149]) as an autonomic approach packet routing algorithm (Gelenbe, Gellman, and Su [83], Gelenbe and Núñez [131]) for wired (Gelenbe and Kazhmaganbetova [38], Gelenbe et al. [81]) and wireless ad hoc networks (Gelenbe and Lent [105]). It exploits the RNN in a reinforcement learning algorithm to provide network quality of service (QoS) in an automatic manner. It was also recently tested for Software Defined Networks (François and Gelenbe [34]) and for intercontinental overlay networks over the Internet (Brun, Wang, and Gelenbe [16], François and Gelenbe [33]). While traditional network research has relied on mathematical modeling and discrete event simulation (Gelenbe [42,43]), CPN is based on an autonomic or self-organized approach (Gelenbe, Seref, and Xu [136]) that suggests that decisions based on “naturally inspired heuristics” (Gelenbe [54]) can be effective even in environments where adversarial agents may operate (Gelenbe and Cao [74]).

The basic idea of CPN was tested in several experiments (Gelenbe et al. [110], Gelenbe, Lent, and Nunez [111], Gelenbe, Lent, and Xu [112,113], Gelenbe and Liu [115]). It uses cognitive packets (CPs) that measure the network’s performance while it operates, and search for paths via Reinforcement Learning using a RNN, based on the QoS objective pursued by the end user. CPs furnish information to the end user, which then selects a given path to the destination(s) required by the connection it is managing (Gelenbe [60], Gelenbe and Lent [104], Gelenbe, Lent, and Xu [114]).

The theoretical foundations of why CPN works well can be found in (Gelenbe [61]). Here it is shown that if a searcher proceeds from a source to a destination, then despite an infinitely large search space it will reach the destination in a finite amount of time provided two conditions are satisfied: (a) a finite time-out is used to abort and then relaunch the search if the searcher has attained the time-out; (b) when a search is repeated it is randomized independently of past searches, and past errors are not systematically repeated. It is also shown that there is also an optimum time-out that minimizes the search time. This is exactly what CPN does, since it also uses a time-out (although as a maximum number of hops rather than a maximum amount of time), and a fraction of CPs (of the order of 10% in many implementations) are routed at random. A remarkable aspect of this theoretical result is that it proves that the searcher will reach the destination in a finite amount of time even if at each step, the search heads in the wrong direction on average. The energy
consumption for such searches is discussed in (Abdelrahman and Gelenbe [6]) where it is shown how the amount of energy consumed can be minimized by an appropriate choice of the time-out. These results were further extended and detailed in the setting of big networks and big data in (Abdelrahman and Gelenbe [4,5], Gelenbe [72]).

An extension of the CPN using genetic algorithms was proposed in (Gelenbe, Liu, and Lainé [116], Liu and Gelenbe [168]). In (Gelenbe and Kazhmaganbetova [96]), web applications where the uplink requires short response time, while the download requires high bandwidth and low packet loss are considered, and a system that supports these needs is designed. Other work deals with the QoS needs of Voice over the Internet (Wang and Gelenbe [182]). Energy aware routing with CPN is also discussed in (Gelenbe and Mahmoodi [119]). Other extensions concern admission control (Gelenbe, Sakellari, and D’arienzo [134]) and distributed denial of service (DDoS) defense (Gelenbe, Gillman, and Loukas [82], Gelenbe and Loukas [117]). In DDoS, network attacks can be detected using CPN as QoS violations, and CPN can be modified to automatically counter-attack by tracing the attacking traffic and using CPN ACK packets to give “drop” orders regarding the attacking traffic to routers upstream (Oke, Loukas, and Gelenbe [170]). Network worms were also considered, and CPN was used to reroute the users’ traffic to avoid the infected nodes (Sakellari and Gelenbe [176], Sakellari, Hey, and Gelenbe [177]). Further research on network security can be found in (Gelenbe et al. [87], Gorbil et al. [151], Yu et al. [185]).

Low-cost disruption-tolerant techniques for robust communications in emergencies (Filippoupolitis, Gorbil, and Gelenbe [28], Gorbil, Filippoupolitis, and Gelenbe [152], Gorbil and Gelenbe [154]), such as opportunistic communication systems were evaluated in (Gorbil and Gelenbe [153], Lent et al. [167]). The use of fast autonomic RNN-based routing of evacues using CPN (Bi, Desmet, and Gelenbe [13]) and directional navigation were also developed (Gelenbe and Bi [73], Kokuti and Gelenbe [166], Gelenbe, Akinwande, and Bi [172]). There have been several applications of this approach to emergency management (Gelenbe and Wu [146,148]), and routing algorithms specifically for this area are discussed in (Desmet and Gelenbe [23], Gelenbe, Gorbil, and Wu [88], Gelenbe and Wu [147]). Proactive and reactive congestion management is introduced in (Desmet and Gelenbe [24]), while designing the routing using CPN for evacuees who have diverse capabilities regarding their mobility and speed is discussed in (Bi and Gelenbe [14]).

Turning to research regarding the Internet, smart traffic routing (Wang and Gelenbe [183]) in conjunction with smart task allocation in Clouds using the CPN approach are discussed in (Gelenbe and Wang [144], Wang, Brun, and Gelenbe [181]). Thus, the paper on “The Random Neural Network for cognitive traffic routing and task allocation in networks and the Cloud” summarizes recent research in three directions: traffic routing when there are significant real-time QoS constraints, overlay routing in the Internet, and the dispatching of tasks from end users to Cloud servers when end-to-end QoS (including delays through the Internet) must be considered.

5. CONCLUSIONS

This paper has reviewed both theoretical developments and applications of research on G-Networks that has been active for the last 25 years. For the sake of brevity, we have not tried to be comprehensive in all the theoretical developments, nor have we been complete regarding the applications.

However, as we have reviewed the different areas, we have linked the publications in this special issue to the previous work that has given rise to the new developments. We therefore hope that our review will help those who are seeking new applications of probability models
in Information Science and Engineering, and also motivate novel mathematical work to be pursued in G-Networks to be able to model more accurately the many different and exciting real systems that have been discussed in this paper. We are also aware that fundamental developments of G-Network models are still being introduced (Furneau and Gelenbe [35,36]) and that they will give rise to further theoretical developments and applications.

References


