WILEY | Hindawi

Research Article

Green Internet of Things and Big Data Application in Smart Cities Development

Zhai Yang,[1] **Liu Jianjun** (ID)**,**[1] **Humaira Faqiri,**[2] **Wasswa Shafik** (ID)**,**[3]
Alanazi Talal Abdulrahman (ID)**,**[4] **M. Yusuf,**[5] **and A.M. Sharawy**[6]

[1]*College of Landscape Architecture and Art, Northwest A & F University, Yangling, Shaanxi, China*
[2]*Education Faculty, Farah Institute of Higher Education, Farah, Afghanistan*
[3]*Computer Engineering Department, Intelligent Connectivity Research Laboratory, P.O. Box 89175-741, Yazd, Iran*
[4]*Department of Mathematics, College of Science, University of Ha'il, Ha'il, Saudi Arabia*
[5]*Department of Mathematics, Faculty of Science, Helwan University, Helwan, Egypt*
[6]*Department of Mathematical and Natural Sciences, Faculty of Engineering, Egyptian Russian University, Badr City, Egypt*

Correspondence should be addressed to Liu Jianjun; jianjun_liuu@outlook.com and Wasswa Shafik; wasswashafik@gmail.com

This study reveals that increases in the global population command an augmented demand for products and services that calls for more effective ways of using existing natural resources and materials. The recent development of information and communication technologies, which had a great impact on many areas, also had a damaging effect on the environment and human health. Therefore, societies are moving toward a greener future by reducing the consumption of nonrenewable materials, raw materials, and resources while at the same time decreasing energy pollution and consumption. Since information technology is considered a tool for solving ecological difficulties, the green Internet of things (G-IoT) is playing a vital role in creating a sustainable home. Extensive data analysis is required to obtain a valuable overview of the large and diverse data generated by the G-IoT. The gathered information will facilitate forecasting, decision-making, and other activities related to smart urban services and then contribute to the incessant development of G-IoT technology. Therefore, even if sustainable and smart cities become an actuality, the G-IoT approach and the knowledge gained through big data (BD) analysis will make cities more sustainable, safer, and smarter. The goal of this article is to combine innovation in technological development with the main focus on resource sharing in creating cities that improve the quality of life while reducing pollution and realizing more efficient use of the raw materials. In the practice of big data science, it is always of interest to provide the best description of the data under consideration. Recent studies have pointed out the applicability of the statistical distributions in modeling data in applied sciences. In this article, we introduce a new family of statistical models to provide the best description of the life span of the wireless sensors network's data. Based on the proposed approach, a special submodel called new exponent power-Weibull distribution is studied in detail. The applicability of the proposed model is shown by analyzing the life span of the wireless sensors network's data.

1. Introduction

Current works of literature predict that the world's population will reach over 11 billion by 2045 compared to the unknown population of 7.6 billion. Therefore, smart technologies must prevail, e.g., for a productive economy to thrive on both local and global levels, interconnectedness and innovation must prevail through environmental conversation, green house buildings, natural energy usage, and green urban planning, among others. For a smart government to thrive, open data analysis and availability, regulated information, communication, technology for the public domain, and electronic government to enable enough products and services for smart living, health, safety, cultural vibrancy, happiness, integrated information, and mixed-modal access must prevail. In addition, the general population must embrace creativity.

Modern medicines and improved living conditions are the main reasons for the increase in the world's population. There are 25 big cities in the world with a population of over 11 million, and 1.5 million people move to cities on a daily basis. It is projected that the Internet of things (IoT) will lead to smart cities and then to a smart world (see Figure 1).

1.1. Artificial Intelligence. Green cities with artificial intelligence are a notion that has already become a reality (see Figure 2). Significant enhancement in information and communication technology (ICT) has enhanced living standards, decreased the use of energy, products, and services, and limited dangerous pollution. The IoT utilizes various implanted sensors and intelligent devices that allow wireless connectivity of anything from anywhere, anytime. This connectivity concept plays an important role in the development of numerous fields such as health, transport, industry, smart education, and waste management. Hence, the IoT prompts a more intelligent, greener, and proficient world. With so many benefits of ICT and IoT that increase the daily use of smart devices, the use of energy that causes dangerous emissions has also increased [1].

Striving toward a greener future, the energy-restricted devices will be more productively utilized due to diminished contamination and guarantee an intelligent and green place for life. Currently, the G-IoT is playing a significant role in the development of smart cities. Due to the continuous development in smart sensors and gadgets, the G-IoT is also generating huge amounts of data. Therefore, there is a clear need for big data concepts here to deal with the data generated by the G-IoT. The study examines big data via a precise analysis of the rules that make a smart and feasible city through decreasing environmental pollution and contamination, diminishing energy requests, and effectively using resources [2].

1.2. Citizen Engagement. Recently, there has been a persistent need to develop a sustainable place to live. This should draw the attention of stakeholders, companies, and governments. Individuals also need to play a role in the development of a sustainable living place. Issues like climate change, reductions in natural resources, and vanishing biodiversity also need to be addressed. The main objective of the G-IoT revolution is to enhance the quality of life and secure the environment from these kinds of issues by using technological innovations [3]. Figure 3 portrays the total number of active device connections worldwide that Big Data uses as a big statistic in communications and projecting it by the year 2025.

1.3. Digital Service Delivery. Statistical analysis is the collection and interpretation of data to uncover patterns and trends. The electronic delivery of information includes portability of data and content across multiple platforms and devices, e.g., web or mobile. Information is presented in a way that is easy to use and understand and typically involves transactional services such as submitting forms for processing and receiving benefits.

1.4. Skills and Employment. Cities are developed with the most modern technologies to meet the daily life requirements of the population. In the green cities, stakeholders and governments deploy many advanced technologies. They also try to teach laborers to work with these technologies. However, the issue is that green smart cities also reduce the labor intensiveness within the company or organization. There is no doubt that modern technologies enhance increased product and service productivity but reduce employment.

1.5. Renewable Energy. While in transit to a savvy, green, and sustainable world, renewable energy must be prioritized. Energy production was founded on fossil fuels, resulting in huge amounts of carbon dioxide emissions, global warming, and extreme weather. Therefore, it is important to make some new technological changes in the energy sector, and this is now possible with the help of G-IoT. All G-IoT technologies are very easy to control, manage, and balance the demand and production of energy. This can be achieved by deploying intelligent appliances and smart actuators that can autonomously observe energy needs and adjust accordingly. Such energy sensors deployed in the environment can produce more products and services than the use of traditional methods by 30%–50%.

1.6. Clean Industry. Deployed sensors in a smart city monitor garbage, and when garbage exceeds a specific amount, the sensor informs the management system to collect the garbage and burn it to produce energy. The G-IoT and big data are employed for cost-effective and efficient energy production [4, 5].

1.7. Green Manufacturing. Green manufacturing involves environmentally friendly operations within manufacturing. Fundamentally, it is the "greening" of manufacturing, in which workers use fewer natural resources, reduce pollution and waste, recycle and reuse materials, and moderate emissions throughout their processes. Systematic approaches of the various engineering disciplines engaged in green engineering include sustainable designs, life cycle analysis, pollution prevention, design for the environment, design for disassembly, and design for recycling. As such, green engineering is a subset of sustainable engineering [6]. Figure 4 shows an increase in urbanization from 1990 to 2014 and a projection for 2050.

2. Analysis of Big Data

Back in 2001, Gartner analyst Doug Laney listed the three Vs of BD: variety, velocity, and volume. The name BD itself indicates an enormous size. The size of data plays a very crucial role in determining the data output. Additionally,
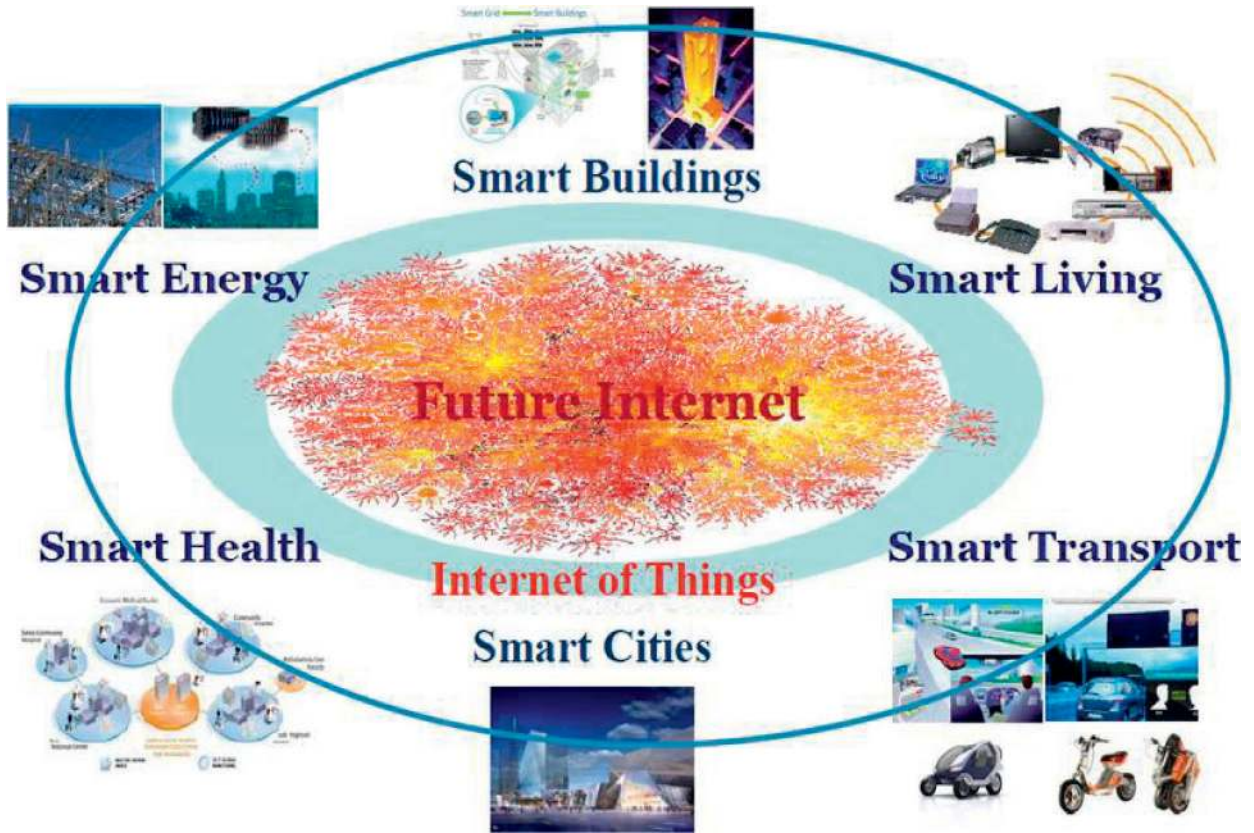
FIGURE 1: Future Internet and smart computing.

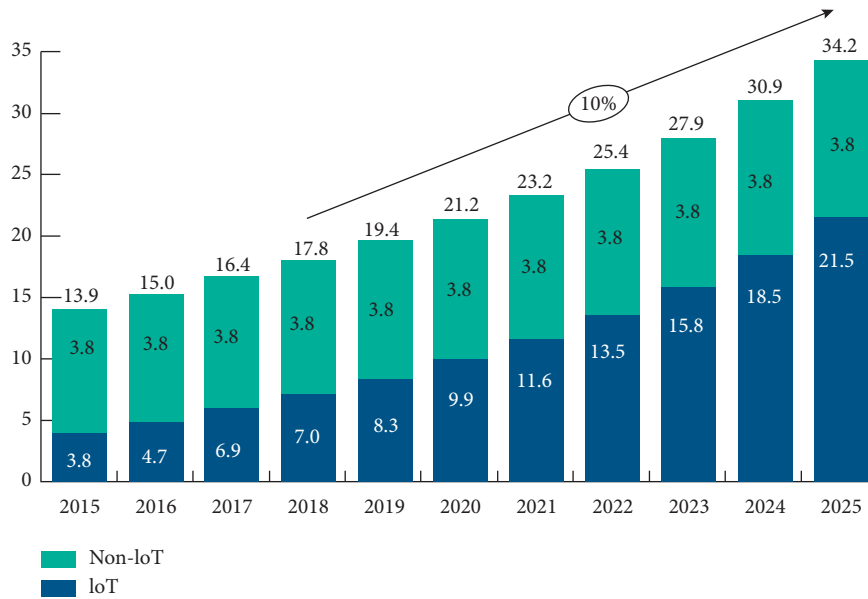FIGURE 2: Smart cities and related entities.

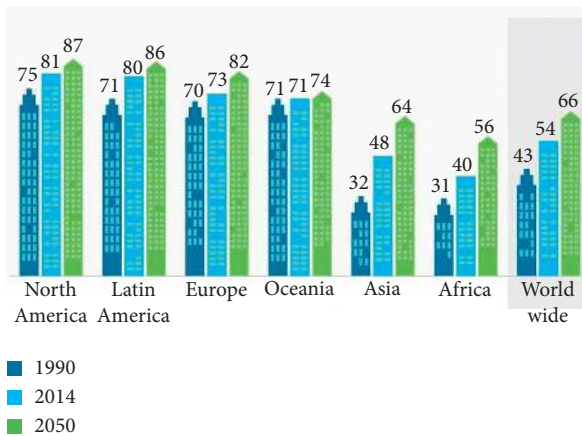FIGURE 3: Total number of active device (IoT and Non-IoT) connections worldwide.

FIGURE 4: Increase in population in cities.

whether a particular data can be considered as BD is dependent on its volume.

2.1. Characteristics of Big Data.
In this subsection, we present several characteristics of BD that are commonly used in computing.

2.1.1. Volume.
Volume is one of the characteristics of BD. We already know that big data indicate huge volumes of data that are being generated daily from different sources like social media platforms, business processes, machines, networks, and human interactions. Such large amounts of data are stored in data warehouses within tables, records, and transactions [7].

2.1.2. Variety.
BD comes in multiple varieties. Compared to the traditional data like phone numbers and addresses, data currently most often takes the form of an image, video, or

audio file. This makes about 80% of all data to be completely unstructured. The variety of BD refers to the structured, unstructured, and semistructured data that is gathered from multiple sources [8].

2.1.3. Velocity.
Velocity essentially refers to the speed at which data is being created in real time. In a broader prospect, it comprises the rate of change, linking incoming datasets at varying speeds and activity bursts. Since a major part of the data is unstructured and irrelevant, BD needs to find a way to filter or push out data. The more IoT devices are connected, like sensors using BD features in the world of computing, the greater the transmission of more bits of data to an ever-increasing degree. As the number of data units increases, data flow increases [9].

2.2. Types of Big Data Analytics.
In this subsection, various types of big data analytics are illustrated.

2.2.1. Prescriptive.
The most valuable and most underused big data analytics technique, prescriptive analytics, gives one a laser-like focus to answer a specific question. It helps us to determine the best solution among a variety of choices, given the known parameters and suggests options for how to take advantage of a future opportunity or mitigate a future risk [10].

2.2.2. Predictive.
Predictive analytics is the most commonly used technique. It uses models to forecast what might happen in specific scenarios, such as the next best offers, churn risk, and renewal risk analysis. Moreover, it is focused on nondiscrete predictions of future states, relationship, and patterns, the description of prediction result set probability

distributions and likelihoods, model application, and non-discrete forecasting [11].

2.2.3. Diagnostic. When trying to determine why anything has happened, data scientists turn to diagnostic methodology. It is helpful to consult the most loyal customers when studying leading churn metrics and consumption patterns. Churn reason analysis and consumer health score analysis are examples of diagnostic analytics, which includes a backward-looking focus on causal associations and sequences.

2.2.4. Descriptive. This technique is the most time-intensive and often produces the least value; however, it is useful for uncovering patterns within a certain segment of customers. Descriptive analytics provide insight into what has happened historically and will provide trends that can be dug into in more detail. For example, descriptive analytics include summary statistics, clustering, and association rules used in market basket analysis and in pattern detection and descriptions [12].

2.3. Big Data Contribution to Smart Cities with G-IoT. Due to the wireless sensor network support of connectivity of devices from any place anywhere and the increasing number of people using the Internet and smart gadgets, huge amounts of data are being created. Processing the data collected from a smart city is not possible with the current traditional data processing techniques such as SQL queries relational database management systems.

2.3.1. Location Tracking. Location tracking allows us to record where users are and where they have been. When a user enables location tracking, licensed users can use the Tracker mobile app to upload their locations to the location-tracking layer from G-IoT's perspective. The tracks are secure in the location-tracking layer: users only see their tracks, and additional permissions are hidden. An example of this category is the use of the Internet and mobile map applications on smartphones that can calculate a user's current position using the global positioning system and show where they are on a map [13].

2.3.2. Precision Medicine. This tactic to patient care allows doctors to select treatments that are most likely to help patients based on a genetic understanding of their disease. Apparatuses employed in precision medicine include molecular diagnostics, imaging, and analytics. In explaining the distinction from a similar term, i.e., personalized medicine, the National Research Council explains that "precision medicine refers to the tailoring of medical treatment to the individual characteristics of each patient."

2.3.3. Advertising. Many new small businesses in G-IoT owners focus so much on their products and service that they do not recognize that success comes not just from what the business offers. Success comes from the ability to reach customers and get them to buy. Marketing is a crucial element to small business success. Before the Internet, marketing, especially to the masses, was difficult and expensive [14].

2.3.4. Fraud Detection. A variety of specialized activities are conducted to avoid the acquisition of access or property under false pretenses. Fraud detection is applied to many sectors that use technology, such as banking or insurance, as a routine way of working. In banking, fraud can involve forging checks or using stolen credit cards as a case study. Handling traffic from this point will help to more efficiently route various network layer traffic to where it needs to go before entering any back haul networks for telecommunications and definitely before direct ingestion at the data center or colocation.

2.3.5. Fraud Handling. Fraud handling aims to prevent a client from committing a crime or fraud that is reasonably certain to result in substantial injury to financial interests or property of another, and in furtherance of which the client has used or is using the G-IoT services, i.e., to secure legal advice about the G-IoT compliance with these rules [15]. There exist a number of ways of ensuring the security of devices, e.g., streamlining G-IoT investigations with intelligent case management; treating fraud, money laundering, and cyber events together; using artificial intelligence to bring greater precision and reliability to fraud detection. From an application aspect, device security can be improved by religiously removing sensitive data files from systems when they are no longer needed. Also recommended is the continuous use of encryption when storing or transmitting sensitive data and using different fraud-detection techniques to secure systems.

2.3.6. Entertainment and Media. The cloud enables G-IoT and BD for the storage, processing, and mining of data. In other words, cloud computing effectively maintains the massive amount of a city's data. For the G-IoT and BD, there are many techniques proposed by the research community to deal with the data, such as data mining, machine learning, clustering, and fuzzy-based decision-making concepts, and novel algorithms, which are used for issue solution, determining the similarity of the informational collection, and obtaining desired results, of which the following are the notable contributions of BD [16].

2.3.7. Knowledge Discovery Tools. The planning and development of green computing devices have improved networking infrastructures, communication protocols, and maximized bandwidth through less use of energy. The deployment of these G-IoT devices and innovative technologies can decrease hazardous gas emissions and improve a living space. The G-IoT contains sensors, fog nodes that act like entrenched processing nodes, cloud-based nodes, and software for task automation.

3. G-IoT's Impact on Smart City Development

The G-IoT supports smarter cities by connecting smart sensors, devices, vehicles, and infrastructure at all places in the city. The G-IoT enables stakeholders to decrease chemical emissions and water consumption while simultaneously increasing the quality of life. For this smart city concept, the G-IoT has two characteristics [17]. Hence, the G-IoT's implementation in smart cities considers the insertion of valuable ICTs; for example, Wireless Sensor Networks (WSN), cellular networks, radio-frequency identification (RFID), communication devices, big data analytics, and data handling [18, 19].

3.1. Reduction of Hazardous Gas Emissions. Different sensors, gadgets, and devices are deployed in the city for monitoring and storing information. There are also some meters installed in the environment to measure energy use, waste levels, and gas emissions.

In the G-IoT, WSN and RFID are the most important technologies that allow each device to connect and collaborate with others. WSN is a collection of smart sensors and devices that can be deployed in any environment to monitor and store information on the various activities performed in that environment. The innovative developments in the WSN domain enable it to improve efficient energy use; moreover, enriched networking protocols have made WSN greener [20].

Both WSN and RFID support the G-IoT to achieve its objective of smart cities. Owing to the huge number of people moving to cities and the greater number of industries that cater to a growing population's demand for products and services by the stakeholders, the main objective of the G-IoT is the management of energy use and gas emissions [21]. Many protocols, technologies, infrastructures, and communication strategies have been proposed to this end.

The G-IoT is set up in the green industry, green correspondence, green preparation, green energy use, and green waste management, allowing it to be collaborative, automatic, and intelligent. The use of ecological policy and implication of the bioproducts for making producers of G-IoT mechanisms serves as an alternative way to use energy like solar, water, air, and biogases. Reducing energy utilization and expenses while delivering energy-related G-IoT segments is vital; shutting down the G-IoT hardware at whatever point is conceivable, which will allow the rest of planning calculations to contribute appropriately as well. The reduction in the size of wireless data paths for data delivery and use of advanced data communication technologies are key factors [22].

As a result, the deployment of all these technologies in the environment will support the G-IoT. The G-IoT improves human health because of its green considerations.

All G-IoT technologies are very easy to control, manage, and balance the demand and production of energy. This can be achieved by deploying intelligent appliances and smart actuators that can autonomously observe energy needs and adjust accordingly. This involves the use of a smart grid (SG).

Due to continuous and large developments in the SG (especially related to its energy usage), it is compulsory to implement G-IoT concepts such as green power consumption. Then, the SG will become Green SG (G-Sg). G-Sg is a system that is automated, self-healing, safe, qualitative, cost-effective, and environmentally safe. G-Sg also provides alternative means of energy production, such as through solar, wind, and water resources, and controls the distribution of energy between sources. The G-IoT will surely contribute to the reduction of gases by 25%, energy consumption by 30%, and efficiently distribute the resources between sectors by 25%; see [23].

In smart cities, water management through ICT can be achieved. Here, the G-IoT's main goal is to save water and enhance its management. In the G-IoT, this can be achieved by the use of intelligent measuring devices and autonomous monitoring [24]. The G-IoT is not only involved with water safety but also in air vitalization, etc. Furthermore, the G-IoT is transforming modern lifestyles.

Environmental resources produce a massive amount of data that needs to be handled as big data. The G-IoT store all generated data in the cloud, and then some prediction techniques are applied so that the system can take appropriate actions. By using advanced controlled and automated systems, energy use and wastage can be decreased by 40% to 70% [25].

Due to the migration of people to cities, waste has increased, and this challenges the G-IoT in maintaining the smart city. Now the G-IoT needs some technological solutions to manage waste and use it for energy production. In using the G-IoT concept, many countries, e.g., China, are producing energy, and the most developed countries are using their waste for many other purposes as well, in addition to producing several resources as alternatives to natural resources [26–35]. This strategy has been realized through the adoption of G-IoT concepts and innovative technologies such as sensors, smart waste collection, and waste burning systems.

The energy using the sensors deployed in the environment can be increased over that produced using traditional methods by 30%–50%. There is also a need for surveillance in smart cities to reduce the crime rate and enhance public security. The safety can be monitored by the installation of smart cameras and sensors. The G-IoT using big data can also help to manage traffic and help a driver choose the shortest route to their destination. This can be achieved by the continuous analysis of road data using big data mining techniques. The G-IoT using big data can also support society in solving parking issues. Due to the huge population moving toward the urban areas, there is an issue of available parking. Therefore, the G-IoT using big data can suggest an empty parking space in busy parking areas.

Alternative network logistics (virtual networks) dealing with heterogeneous QoS requirements of network traffic require fundamental deviations from traditional Internet architectures. The development and advancement of G-IoT facilitate innovation of VNs including, sensor networks, autonomous driving that includes the efficient utilization of the wireless links between the autonomous vehicle and the

cloud; having a request-based system that enables autonomous cars to gather information about different locations at various resolutions, and microgrids; aggregation of data on imported or exported electricity from the distribution node to the microgrid by means of a concentrator together with isolated sensor networks toward networked (ubiquitous) sensor networks.

4. Interaction of G-IoT Virtual Networks in All-IP Networks

Allocation problems arise on two levels: network infrastructure capacities (QoS for All-IP network) and virtual service network providers: "Thus, network virtualization should provide the capability of regulating the upper limit bandwidth usage by each LINP in order to maintain the overall throughput and performance. Virtual or logistic networks focus on end-to-end QoS. Current Internet architecture is not enough to support quality of service from a user or application perspective. It is still unclear how and where to integrate different levels of quality of service in the architecture.

A multichannel congestion pricing model without channel priority does not lead to quality-of-service differentiation with quality guarantees. Since there is no channel priority with QoS guarantees provided for the users, only the degree of congestion within the different channels is relevant and thus users would immediately switch to an alternative traffic service provider offering the same degree of congestion at a lower price. Therefore, in equilibrium, if bandwidth capacity differs among channels, the split of the number of packets on the different channels results endogenously (without admission control) such that the level of congestion and the congestion prices are identical.

According to analyses, pricing levels in other traffic classes were so low as to be negligible.

Under competition on the markets for traffic services, each traffic service provider makes his own autonomous decisions. The choice of the number of channels $i = 1, \ldots, n$ with monotone decreasing channel quality ψ_i _$i = 1, \ldots, n$, the bandwidths, and usage dependent prices for the different channels are derived simultaneously. There is no social planner to globally optimize capacities and prices for all traffic service networks. In the following, the allocation problems of an arbitrarily chosen traffic service provider under free entry are analyzed.

In the absence of channel prioritization, channel separation could be motivated by reducing congestion externalities by splitting the traffic over several channels. However, the strategy of channel separation can be particularly motivated by the advantages of channel-tailored deterministic traffic quality control and investment incentives for a hierarchy of deterministic QoS guarantees the multichannel model must therefore be generalized for a hierarchy of traffic classes with strict admission control.

The hierarchy of traffic classes based on a monotone relation of more strictly defined QoS parameters is based on a monotone relation of opportunity costs of bandwidth capacity usage due to different QoS classes and restoration priority parameters. Network capacity (bandwidth) is allocated to each quality class (channel) separately, including the required reserve capacity due to the required restoration priority parameter to guarantee the deterministic parameters (delay, jitter, packet loss). There are quantifiable requirements with respect to delay, jitter, and packet loss associated with the forwarding of data packets, and there is transparent priority implied among channels.

Only the deterministic guarantee of traffic quality parameters is relevant, resulting in higher opportunity costs for more restrictive deterministic traffic guarantees. Even if two channels are equally congested within a considered time period, the users would not be indifferent between the channels. Channel numbering does indicate the quality of service hierarchy. Class 1 would be preferred for high quality sensitive applications due to the absolute highest traffic quality guarantee, even if at some time congestion.

It is assumed that the network of a typical traffic service provider consists of a hierarchy of n logically separated channels. For the sake of simplicity, we assume that the hierarchy of channels regarding QoS guarantees and the hierarchy of service restoration priority levels are identical, avoiding an evaluation conflict between lower channel QoS class versus higher service restoration priority. Channel 1 is the channel with the highest priority guarantee, channel n, the channel with the lowest quality guarantees (or absence of any quality guarantees).

The opportunity costs of channel usage depend on the traffic quality class, with increasing bandwidth capacity usage due to different QoS classes and restoration priority parameters. Let $P_{it}(Q_{it}, \psi_i)$ denote the inverse demand function for packet transmission in channel i with traffic quality ψ_i in period t with traffic flow Q_{it}. We assume that demand is independent across time periods because we do not aim to analyze intertemporal demand interdependencies. There is no reshifting of capacity from one time period to another between channels.

Bandwidth capacity allocated to channel i with quality ψ_i consists of a share of reserved bandwidth w_i and restorations spare (reserve) capacity $r_i(\psi_i, w_i)$. Capacity costs of channel i are denoted $p_i(w_i) + p_i(p_i(\psi_i, w_i))$, consisting of the capacity costs of allocated bandwidth and the bandwidth costs of restoration reserve requirements. Note that the bandwidth costs are chosen ex-ante and are fixed costs for the whole "lifetime" including all time periods $t = 1, \ldots, T$.

Due to channel separation, restoration reserve capacity is class-specific and depends only on the channel of class i. With quality ψ_i restoration reserve capacity requirements are monotonously decreasing with the increasing number of the quality class. Thus, costs of restoration spare capacity are increasing with quality parameter ψ_i.

Let $k_{it}(Q_i, w_i, \psi_i)$ be the private (average) variable costs of a packet transmission within channel i within quality class ψ_i and channel bandwidth $w_i = i = 1, \ldots, n$ quality classes.

$$\frac{\partial k_{it}}{\partial Q_{it}} > 0. \tag{1}$$

If bandwidth w_i remains constant, additional traffic within channel i will slow down every packet within this channel, thereby raising intrachannel externalities:

$$\frac{\partial k_{it}}{\partial Q_{it}} > Q_{it}. \tag{2}$$

Admission control and restoration reserve capacity guarantee that intrachannel externalities do not conflict with the deterministic traffic quality guarantees.

$$\frac{\partial k_{it}}{\partial w_{it}} < 0. \tag{3}$$

If traffic remains constant, the additional bandwidth capacity of channel i will allow to speed up every packet in this channel:

$$\frac{\partial k_{it}}{\partial \psi_{it}} < 0, \tag{4}$$

where i represents an increment in traffic quality; the maximum channel capacity decreases, resulting in an upward shift of the variable cost functions. Thus, in case the quality level increases, variable costs also increase.

Under competition, each traffic service provider chooses channel quality classes, channel capacities, and packet prices in each channel in such a way as to maximize profit. The profit maximization problem is defined by the following:

$$\prod (Q_{1t}, \ldots, Q_{nt}, w_i, \ldots, w_n, \psi_1, \ldots, \psi_n) = \sum_{t=1}^{T} [P_{1t}(Q_{1t}, \psi_1)Q_{1t} + \cdots + P_{nt}(Q_{nt}, \psi n)Q_{nt}]$$

$$- \sum_{t=1}^{T} \left(\sum_{i=1}^{n} k_{it}(Q_{1t}, w_i, \psi_i)Q_{1t} \right) - \sum_{i=1}^{n} p_i(w_i) - \sum_{i=1}^{n} p_i(r_i(\psi_i, w_i)). \tag{5}$$

Necessary conditions for the maximum are derived by differentiating (5) with respect to $Q_{1t}, \ldots, Q_{nt}$ for each $t = 1, \ldots, T$ and with respect to $w_i, \ldots, w_n$ and with respect to $\psi_i, \ldots, \psi_n$ setting each derivative to zero.

The optimal pricing rule concerning the congestion fee for a packet transmission on channel i is given by the following:

$$\tau_{it} = P_{it} - k_{it} = \frac{\partial k_{it}(w_i, \psi_i)}{\partial Q_{it}} Q_{it}. \tag{6}$$

The first-best optimal rule for (bandwidth) capacity in channel i is given by the following:

$$\frac{\partial p_i(w_i)}{\partial w_i} + \frac{\partial p_i(r_i(\psi_1, w_i))}{\partial w_i} = \sum_{t=1}^{T} Q_{it} \cdot \frac{\partial k_{it}(Q_{it}, w_1, \psi_i)}{\partial w_i}, \quad i = 1, \ldots, n. \tag{7}$$

The first-best optimal rule for quality choice ψ_i is given by the following:

$$p_{it}(Q_{it}, \psi_i) = \frac{\partial k_{it}(Q_{it}, wi, \psi_i)}{\partial \psi_i}, \quad t = 1, \ldots, T, i = 1, \ldots, n. \tag{8}$$

Simultaneous solutions of equations (6)–(8) provide the first-best allocation of traffic flows $Q_{1t}^*, \ldots, Q_{nt}^*$ $t = 1, \ldots, T$ between the different channels and first-best capacity dimensions for each individual channel and optimal traffic classes $\psi_1^*, \ldots, \psi_n^*$.

The following conclusions can be drawn:

$$\tau_{it} = \frac{\partial k_{it}(w_i, \psi_i)}{\partial Q_{it}}. \tag{9}$$

Congestion externalities in channel i increase with the number of packets transmitted in this channel (within the limits of deterministic traffic quality guarantees). Due to the bandwidth reservation for each channel, only externalities within separate channels are relevant, whereas externalities between channels (interchannel externalities) do not occur as expressed in (9).

Optimal congestion fees increase with the level of congestion as well as with increasing traffic quality. Congestion fees within each channel $i\tau_{it}$ may vary over time, in case demand varies over time.

$$\frac{\partial p_{it}(w_i)}{\partial (w_i)} + \frac{\partial p_i(r_i(\psi_i, wi))}{\partial w_i}\Big|_{wi=w_i^*} = \sum_{t=1}^{T} Q_{it} \cdot \frac{\partial k_{it}(Q_{it}, w_i^*, \psi_i)}{\partial w_i}. \tag{10}$$

Capacity (bandwidth) in each channel i should be extended to the point where the marginal cost of increasing share of bandwidth and the marginal cost of restoration capacity is equal to its marginal benefits of reduced congestion (for the packet flow) within channel i with traffic quality ψ_i (10).

$$p_{it} = Q_{it}, Q_i^* = \frac{\partial k_{it}(Q_{it}, w_i, \psi_i^*)}{\partial w_i}, \quad t = 1, \ldots, T; i = 1, 2, \ldots, n. \tag{11}$$

For higher quality classes, marginal cost functions are shifted upwards, resulting (ceteris paribus) in higher package charges. Thus, the hierarchy of traffic quality classes results (ceteris paribus) in higher congestion charges (11). Therefore, the price for packet transmission increases with required traffic quality because variable costs increase with traffic quality. For more details about these measures, we refer to [36].

5. Statistical Methodology and Lifetime Analysis

This section deals with the introduction of a new statistical model and lifetime analysis of the wireless sensor networks (WSNs), which play a prominent role in IoT.

5.1. Statistical Methodology.

The WSNs have driven an increasing interest from research and industry. This can be driven by the availability of cheap, low-powered miniature elements like processors, radios, and sensors that are typically integrated into one chip. The Internet is slickly migrating from an Internet of people toward an IoT. By 2021, it is expected that around fifty (50) billion things will be connected to the Internet.

The lifetime of WSNs is a period of time during which the network may, as appropriate, keep enough connectivity, cover all the considered areas, or keep the data loss rate below a specified level. Several interpretations of the WSNs lifetime exist in the literature. It can be outlined by the waiting time till the primary node sleeps and awakes. It may also be outlined as the time till a mere proportion of nodes dies.

In the practice of big data science and other applied fields, the modeling of a lifetime is an interesting research topic. For modeling lifetime phenomena, the statistical models have received considerable attention [37]. Among the applied fields, modeling lifetime data in network computing, particularly the life span of the WSNs, is the most prominent one.

5.1.1. Development of the Proposed Family.

In this subsection, we propose a new family of statistical models to provide the best description of the life span of the WSNs. Let $k(t)$ be the pdf (probability density function) random variable (r.v) say T, where $T \in [r_1, r_2]$, $-\infty < r_1 < r_2 < \infty$ and let suppose $Q[F(x)]$ be a function of cdf (cumulative distribution function) $F(x)$ of a r.v X, fulfilling the conditions given below:

(i) $Q[F(x)] \in [r_1, r_2]$,

(ii) $Q[F(x)]$ is differentiable and monotonically increasing, and

(iii) $Q[F(x)] \longrightarrow r_1$ as $x \longrightarrow -\infty$ and $Q[F(x)] \longrightarrow r_2$ as $x \longrightarrow \infty$

The of the T-X family has the form

$$G(x) = \int_{r_1}^{Q[F(x)]} k(t)dt, \quad x \in \mathbb{R}, \quad (12)$$

where $Q[F(x)]$ satisfies the conditions stated above. Pdf corresponding to (12) is

$$g(x) = \frac{\partial}{\partial x}\{Q[F(x)]\} k\{Q[F(x)]\}, \quad x \in \mathbb{R}. \quad (13)$$

For the contributed work based on the T-X approach, we refer to [38]. Inspired by (12), we propose a new class of statistical models suitable for modeling the lifetime of the WSNs data. Henceforth, we call the proposed method as the new exponent power-X (NEP-X) family of distributions. Let T has the exponential distribution with rate parameter $\lambda = 1$, then its cdf is given by

$$K(t) = e^{-t}, \quad t \geq 0. \quad (14)$$

The pdf corresponding to (14) reduces to

$$k(t) = e^{-t}, \quad t > 0. \quad (15)$$

If $k(t)$ follows (15) and setting $-\log(1 - \{(\exp(1 - (1 - F(x)/e^{F(x)})) - 1/e - 1)\})$ in (12), we obtain the cdf of the NEP-X family as follows:

$$G(x) = 1 - \left(\frac{\exp\left(1 - \left(1 - F(x)/e^{F(x)}\right)\right) - 1}{e - 1} \right) - 1, \quad x \in \mathbb{R},$$

$$G(x) = \frac{\exp\left(1 - \left(1 - F(x)/e^{F(x)}\right)\right) - 1}{e - 1}, \quad x \in \mathbb{R}. \quad (16)$$

The pdf corresponding to (16) is given by

$$g(x) = \frac{f(x)[2 - F(x)]}{(e - 1)e^{F(x)}} \exp\left(1 - \frac{1 - F(x)}{e^{F(x)}}\right), \quad x \in \mathbb{R}. \quad (17)$$

5.1.2. A New Exponent Power-Weibull: A Special Submodel.

In this sub-section, we introduce a sub model of the of the NEP-X family, called a new exponent power-Weibull (NEP-Weibull) distribution. A r.v X is said to have the two-parameter Weibull distribution, if its cdf and pdf are respectively given by.

$F(x) = 1 - e^{-\theta x^\alpha}$, $x \geq 0, \alpha, \gamma > 0$, and $f(x) = \alpha \theta x^{\alpha-1} e^{-\theta x^\alpha}$, $x > 0$. Using the cdf of the Weibull distribution in (16), we get the cdf of the NEP-Weibull distribution given by

$$G(x) = \frac{\exp\left(1 - \left(e^{-\theta x^\alpha}/e^{\left(1-e^{-\theta x^\alpha}\right)}\right)\right) - 1}{e - 1}, \quad x \geq 0, \alpha, \theta > 0, \quad (18)$$

with survival function (sf)

$$S(x) = \frac{e - \exp\left(1 - \left(e^{-\theta x^\alpha}/e^{\left(1-e^{-\theta x^\alpha}\right)}\right)\right)}{e - 1}, \quad x > 0. \quad (19)$$

For different values of the model parameters, the plots of cdf and sf of the NEP-Weibull distribution are presented in Figure 5.

The pdf corresponding to (18) is given by

$$g(x) = \frac{\alpha \theta x^{\alpha-1} e^{-\theta x^\alpha}\left[1 + e^{-\theta x^\alpha}\right]}{(e - 1)e^{\left(1-e^{-\theta x^\alpha}\right)}} \exp\left(1 - \frac{e^{-\theta x^\alpha}}{e^{\left(1-e^{-\theta x^\alpha}\right)}}\right), \quad x > 0. \quad (20)$$

For different values of the model parameters, plots of the pdf of the NEP-Weibull distribution are sketched in Figure 6.

5.1.3. Maximum Likelihood Estimation.

In this section, we obtain the maximum likelihood estimators (MLEs) of the NEP-Weibull distribution. Let $x_1, x_2, ..., x_n$ be the sample observations taken from pdf (11) with parameters α and θ. Corresponding to equation (11), the log-likelihood function denoted by $\ell(\alpha, \theta)$, is given by
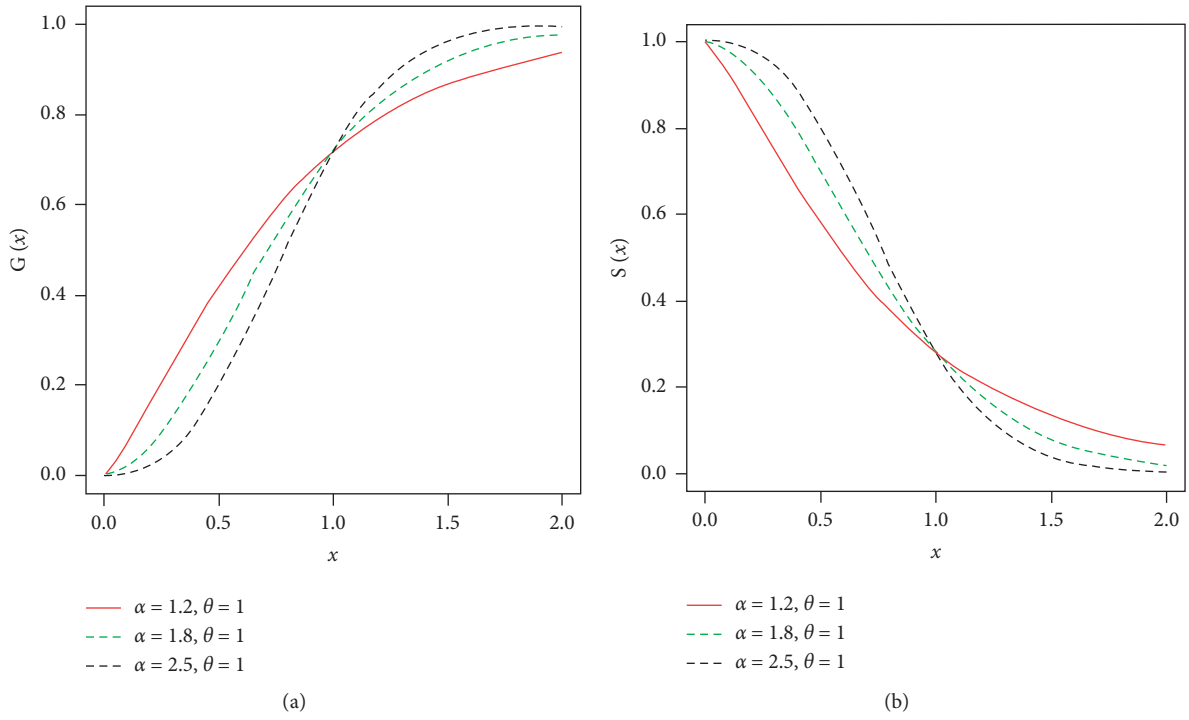
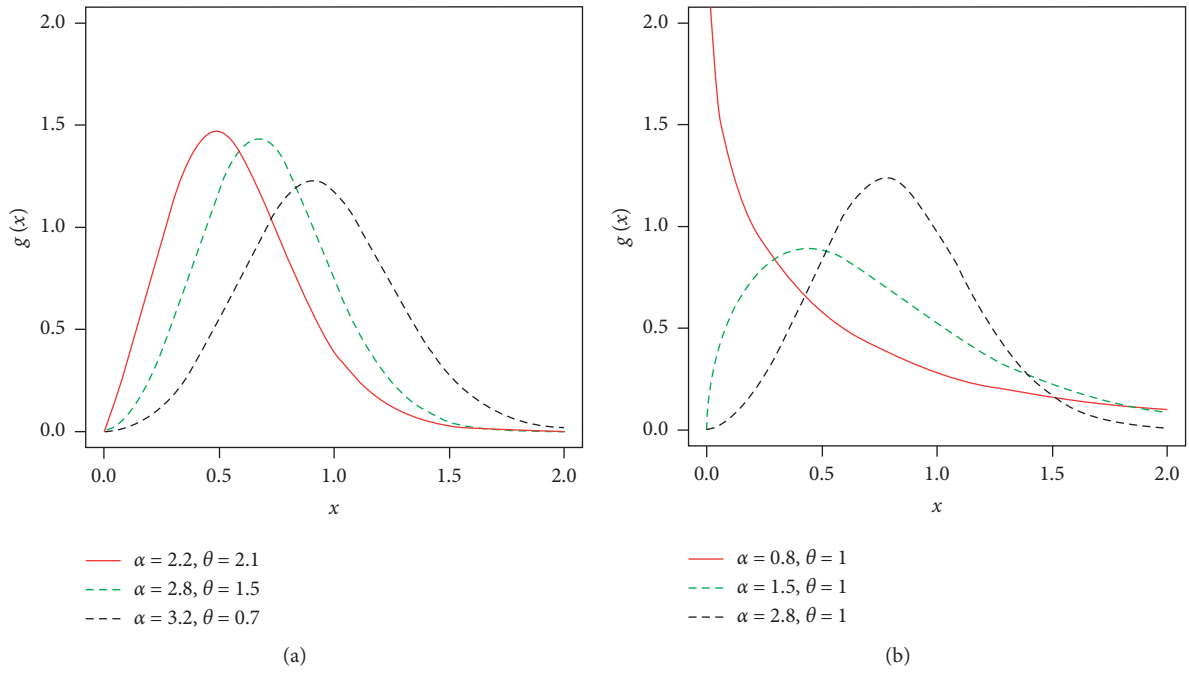FIGURE 5: Different plots for the cdf and sf of the NEP-Weibull distribution.

FIGURE 6: Different plots for the pdf of the NEP-Weibull distribution.

$$\ell(\alpha, \theta) = -n \log(e - 1) + n \log \alpha + n \log \theta - \sum_{i=1}^{n} \theta x_i^{\alpha}$$

$$+ \sum_{i=1}^{n} \log\left[1 + e^{-\theta x_i^{\alpha}}\right] - \sum_{i=1}^{n}\left[1 - e^{-\theta x_i^{\alpha}}\right] \qquad (21)$$

$$+ \sum_{i=1}^{n}\left(1 - \frac{e^{-\theta x_i^{\alpha}}}{e^{\left(1 - e^{-\theta x_i^{\alpha}}\right)}}\right).$$

On behalf of the model parameters, the partial derivatives of $\ell(\alpha, \theta)$ is given by

$$\frac{\partial}{\partial \alpha}\ell(\alpha, \theta) = \frac{n}{\alpha} - \theta\sum_{i=1}^{n}(\log x_i)x_i^{\alpha} - \theta\sum_{i=1}^{n}\frac{(\log x_i)x_i^{\alpha}e^{-\theta x_i^{\alpha}}}{\left(1 + e^{-\theta x_i^{\alpha}}\right)}$$

$$+ \theta\sum_{i=1}^{n}\left(\frac{(\log x_i)x_i^{\alpha}e^{-\theta x_i^{\alpha}}e^{-\theta x_i^{\alpha}}\left(1 - e^{-\theta x_i^{\alpha}}\right)\left(1 + e^{-\theta x_i^{\alpha}}\right)}{e^{2\left(1 - e^{-\theta x_i^{\alpha}}\right)}}\right)$$

$$- \theta\sum_{i=1}^{n}\frac{(\log x_i)x_i^{\alpha}e^{-\theta x_i^{\alpha}}}{\left(1 - e^{-\theta x_i^{\alpha}}\right)},$$

$$\frac{\partial}{\partial \theta}\ell(\alpha, \theta) = \frac{n}{\theta} - \sum_{i=1}^{n}x_i^{\alpha} - \sum_{i=1}^{n}\frac{x_i^{\alpha}e^{-\theta x_i^{\alpha}}}{\left(1 + e^{-\theta x_i^{\alpha}}\right)} - \sum_{i=1}^{n}\frac{x_i^{\alpha}e^{-\theta x_i^{\alpha}}}{\left(1 - e^{-\theta x_i^{\alpha}}\right)}$$

$$+ \sum_{i=1}^{n}\left(\frac{x_i^{\alpha}e^{-\theta x_i^{\alpha}}e^{-\theta x_i^{\alpha}}\left(1 - e^{-\theta x_i^{\alpha}}\right)\left(1 + e^{-\theta x_i^{\alpha}}\right)}{e^{\left(1 - e^{-\theta x_i^{\alpha}}\right)}}\right). \qquad (22)$$

Solving the expressions $(\partial/\partial\alpha)\ell(\alpha, \theta) = 0$ and $(\partial/\partial\theta)\ell(\alpha, \theta) = 0$ yields the MLEs of α and θ, respectively.

5.2. Modeling the Life Span of WSNs. The subsection illustrates the applicability of the NEP-Weibull distribution by analyzing real-life data. The data set is representing the life span of the WSNs. The scaled TTT transform plot is a graphical method of identifying the nature of the failure rate function of a dataset. As known in the literature, the rate of failure is increasing (decreasing) if the TTT plot is concave (convex). From the TTT transform plot sketched in Figure 7, we can see that the life span of the WSNs data has increasing failure rate behavior. Furthermore, the histogram and box plots of the data set are also provided in Figure 7.

The summary measures of the data related to the life span of the WSNs are provided in Table 1.

The NEP-Weibull is applied to the life span of the WSMs, and a comparison of the NEP-Weibull is made with the exponentiated Weibull (EW), Marshall-Olkin Weibull (MOW), and alpha power transformed Weibull (APTW) distributions. The cdfs of the competing distributions are as follows:

(i) EW distribution

$$F(x) = \left(1 - e^{-\theta x^{\alpha}}\right)^{\gamma}, \quad x \geq 0, \alpha, \gamma, \theta > 0. \qquad (23)$$

(ii) MOW distribution

$$F(x) = 1 - \frac{\sigma e^{-\theta x^{\alpha}}}{\left(1 - (1 - \sigma)e^{-\theta x^{\alpha}}\right)}, \quad x \geq 0, \alpha, \gamma, \sigma > 0. \qquad (24)$$

(iii) APTW distribution

$$F(x) = \frac{\alpha_1^{\left(1 - e^{-\theta x^{\alpha}}\right)} - 1}{\alpha_1 - 1}, \quad x \geq 0, \alpha, \gamma, \alpha_1 > 0, \alpha_1 \neq 1. \qquad (25)$$

To show the best fitting of the competing distributions, certain analytical measures are considered. These analytical measures include the Akaike information criterion (AIC), Bayesian information criterion (BIC), Hannan–Quinn information criterion (HQIC), consistent Akaike information criterion (CAIC), and Kalmogrove-Simrnov (KS) statistic with p-value. The formulas for these measures are as follows:

(iv) The AIC

$$\text{AIC} = 2k - 2\ell. \qquad (26)$$

(v) The BIC

$$\text{BIC} = k \log(n) - 2\ell. \qquad (27)$$

(vi) The HQIC

$$\text{HQIC} = 2k \log[\log(n)] - 2\ell. \qquad (28)$$

(vii) The CAIC

$$\text{HQIC} = \frac{2nk}{n - k - 1} - 2\ell, \qquad (29)$$

where ℓ denotes the log-likelihood function evaluated at the MLEs, k is the number of model parameters and n is the sample size.

(viii) The KS

$$\text{KS} = \sup_x\left[G_n(x) - G(x)\right], \qquad (30)$$

where $G_n(x)$ is the empirical cdf and $\sup_x$ is the supremum of the set of distances.

All the computations are carried out using the optim() R-function with the argument method "SANN"; see Appendix. A model with lowest values for these measures could be chosen as the best model to fit the data. The MLEs of the
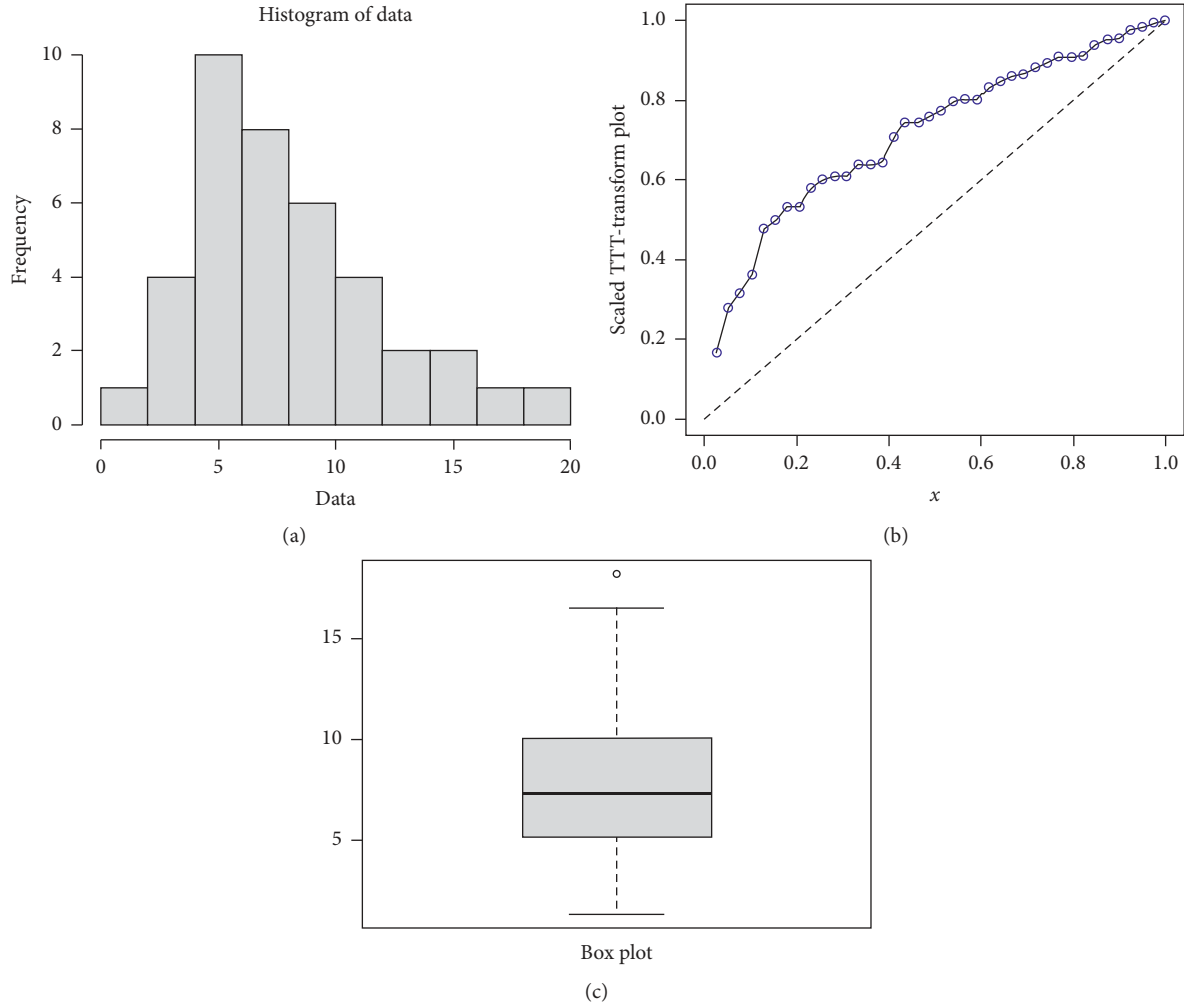
(a)

(b)

(c)

FIGURE 7: The histogram, TTT transform, and box plots of the life span of the WSNs data.

TABLE 1: The summary measures of the life span of the WSNs data.

Min.	1st qu.	Median	Mean	3rd qu.	Max.
1.300	5.150	7.300	7.836	10.050	18.200

model parameters are presented in Table 2, the AIC, CAIC, BIC, and HQIC are presented in Table 3, whereas the KS and the p-value are provided in Table 4.

From the results provided in Tables 3 and 4, we can see that the analytical measures for the NEP-Weibull distribution are AIC = 215.5336, CAIC = 215.8670, BIC = 218.8607, QIC = 216.7274, KS = 0.0740 with p-value = 0.9831. These measures of the second closet model are AIC = 217.5140, CAIC = 218.1997, BIC = 222.5047, HQIC = 219.3046, KS = 0.0790 with p-value = 0.9679.

From the above discussion, we can see the NEP-Weibull distribution has the smallest values of the analytical measures and the highest p-value amongst the competing models. Therefore, we can conclude that the NEP-Weibull distribution may be a good candidate distribution for modeling data in the fields of computer science technology, particularly modeling the life span of

TABLE 2: The MLEs of the fitted models to the life span of the WSNs data.

Distributions	α	γ	θ	σ	α_1
NEP-Weibull	2.1795	—	0.0070	—	—
APTW	2.4746	—	0.0031	—	0.3007
EW	1.709	1.64990	0.0340	—	—
MOW	2.2191	—	0.0081	1.1637	—

TABLE 3: The AIC, CAIC, BIC, and HQIC of the fitted models to the life span of the WSNs data.

Distributions	AIC	CAIC	BIC	HQIC
NEP-Weibull	215.5336	215.8670	218.8607	216.7274
APTW	217.5140	218.1997	222.5047	219.3046
EW	217.6039	218.2896	222.5946	219.3945
MOW	218.3380	219.0237	223.3287	220.1286

the WSNs. Furthermore, in support of the numerical results, the empirical cdf and Kaplan–Meier survival, probability-probability (PP), and quantile-quantile (QQ) plots of the NEP-Weibull distribution are provided in Figure 8. From Figure 8, we can see that the NEP-Weibull

TABLE 4: The KS and p-value of the fitted models to the life span of the WSNs data.

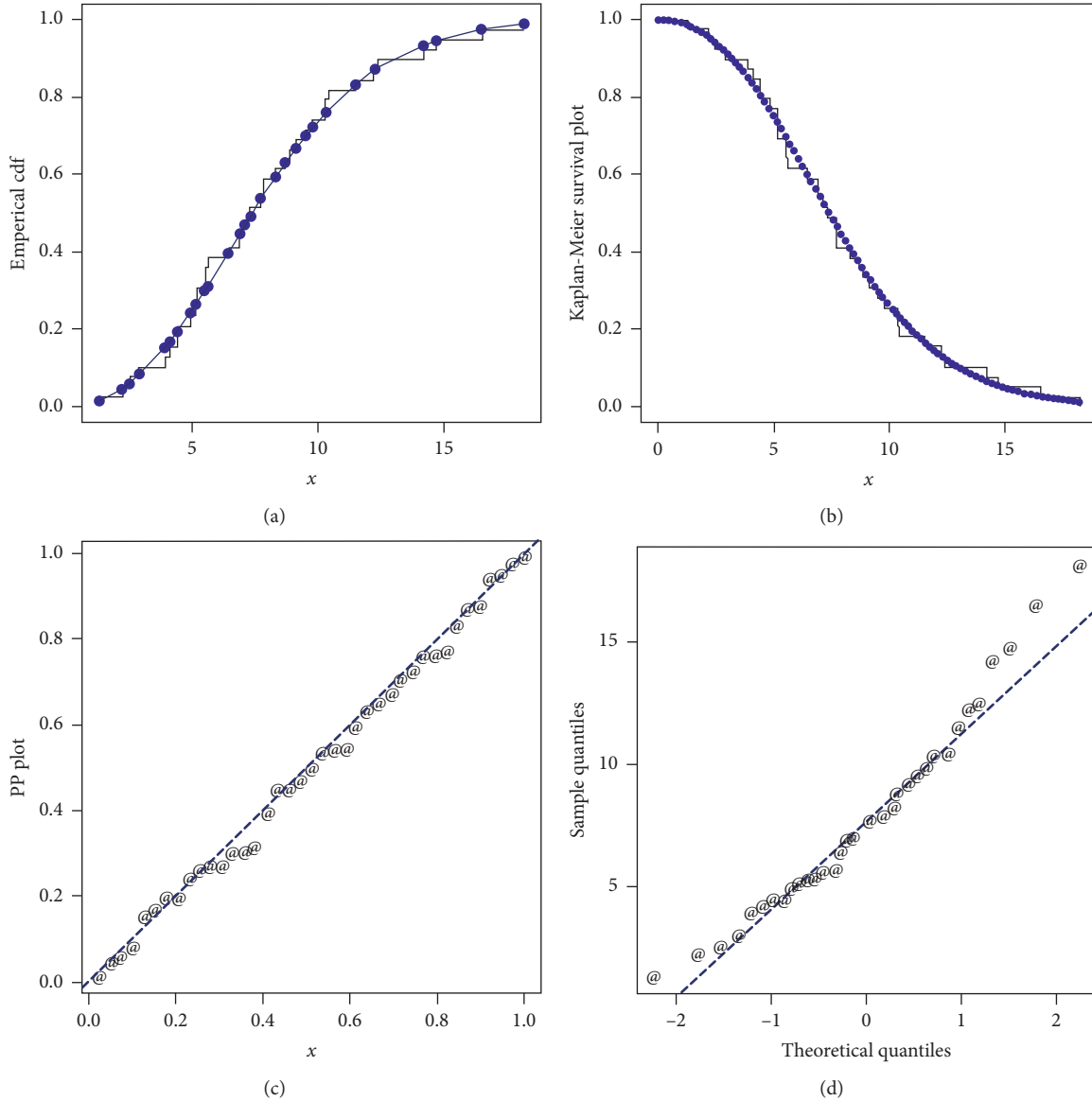| Distributions	SK	p-value
NEP-Weibull	0.0740	0.9831
APTW	0.0790	0.9679
EW	0.0905	0.906
MOW	0.1040	0.7920

(a)

(b)

(c)

(d)

FIGURE 8: The empirical cdf and Kaplan–Meier survival, PP and QQ plots of the NEP-Weibull distribution for the life span of the WSNs data.

distribution provides a close fit to the life span of the WSNs data.

6. Future Work

Instead of a combination of G-IoT and BD, many limitations are faced by corporations and governments in developing smart cities to move people from rural to urban areas. These limitations include the sustainability of technologies, development of smart sensors, reliable and strong Internet connectivity, and efficient use of resources. New modifications to smart sensors and new technologies must be introduced in Big Data for data mining and new decision-making approaches for all G-IoT and BD domains and new G-IoT-BD architectures.

7. Conclusion

Humans move to cities all over the world for the sake of a secure, better, and more sustainable life. Due to this, the

vision of a smart green city is becoming a reality. It is made possible by the implementation of modern technologies in the environment that can interact with the IoT. The IoT allows the wireless connectivity of anything from anywhere, anytime. This connectivity plays a very important role in the development of numerous fields, such as health, transport, industry, smart education, and waste management. In this study, we conducted a detailed review of the G-IoT-BD combination for smart city development. We found that the G-IoT alone is not able to manage the smart green city concept; it needs BD to support it and enhance its performance. Big data provides different techniques that permit valuable data mining from random data. A new statistical model called NEP-Weibull distribution is introduced and applied to the life span of the WSNs data. The comparison of the proposed model was made with EW, MW, and APTW distributions. Certain analytical measures are considered to decide about the goodness of fit of the competing distributions. Based on the results of the analytical measures, we found that the proposed model provides the best fit for data.

Appendix

Note: In the following R code, a is α, t is used θ, and pm is used for proposed model.

```
data < -read.csv(file.choose(), header = TRUE)
data = data[, 1]
data = data[!is.na(data)]
data
#
# Probability density function
#
pdf_pm < - function(par, x)
{
a = par[1]
t = par[2]
a * t * (x̂(a − 1)) * exp(−t * x̂a) * (1 + exp(−t * x̂a)) * (1/
(exp(1 − exp( − t * x̂a)))) *
(1/(exp(1) − 1)) * exp(1 − ((exp(−t * x̂a))/(exp(1  −  exp
(−t * x̂a)))))
}
#
# Cumulative distribution function
#
cdf_pm < - function(par,x)
{
a = par[1]
t = par[2]
(exp(1 − ((exp(−t * x̂a))/(exp(1 − exp(−t * x̂a))))) − 1)/
(exp(1) − 1)
}
```

```
set.seed(0)
goodness.fit(pdf = pdf_pm, cdf = cdf_pm, starts = c(0.5,
0.5), data = data,
method = "SANN", domain = c(0, Inf), mle = NULL).
```

Data Availability

This work is mainly a methodological development and has been applied to secondary data related to the life span of the wireless sensor networks, but if required, data will be provided.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

References

[1] S. H. Alsamhi, O. Ma, M. S. Ansari, and Q. Meng, "Greening internet of things for greener and smarter cities: a survey and future prospects," *Telecommunication Systems*, vol. 72, no. 4, pp. 609–632, 2019.

[2] O. Golubchikov and M. Thornbush, "Artificial intelligence and robotics in smart city strategies and planned smart development," *Smart Cities*, vol. 3, no. 4, pp. 1133–1144, 2020.

[3] M. F. Tuysuz and R. Trestian, "From serendipity to sustainable green internet of things: technical, industrial and political perspective," *Computer Networks*, vol. 182, pp. 10–27, 2020.

[4] A. Booth, "Searching for qualitative research for inclusion in systematic reviews: a structured methodological review," *Systematic Reviews*, vol. 5, no. 1, pp. 1–74, 2016.

[5] A. E. Varjovi and S. Babaie, "Green internet of things: vision, applications and research challenges," *Sustainable Computing: Informatics and Systems*, vol. 28, pp. 1–18, 2020.

[6] W. Shafik, S. M. Matinkhah, P. Etemadinejad, and M. N. Sanda, "Reinforcement learning rebirth, techniques, challenges, and resolutions," *JOIV: International Journal on Informatics Visualization*, vol. 4, no. 3, pp. 127–135, 2020.

[7] X. Chen, S. Liu, J. Lu, P. Fan, and K. B. Letaief, "Smart channel sounder for 5g iot: from wireless big data to active communication," *IEEE Access*, vol. 4, pp. 8888–8899, 2016.

[8] W. Shafik, S. M. Matinkhah, M. N. Sanda, and S. S. Afolabi, "A 3-dimensional fast machine learning algorithm for mobile unmanned aerial vehicle base stations," *International Journal of Advances in Applied Sciences*, vol. 10, no. 1, pp. 28–38, 2020.

[9] X. Liu, A. Liu, T. Wang et al., "Adaptive data and verified message disjoint security routing for gathering big data in energy harvesting networks," *Journal of Parallel and Distributed Computing*, vol. 135, pp. 140–155, 2020.

[10] W. Shafik, S. M. Matinkhah, and M. N. Sanda, "Network resource management drives machine learning: a survey and future research direction," *Journal of Communications*

Technology, Electronics and Computer Science, vol. 30, pp. 1–15, 2020.

[11] J. Ruan, Y. Wang, F. T. S. Chan et al., "A life cycle framework of green iot-based agriculture and its finance, operation, and management issues," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 90–96, 2019.

[12] M. Ghasemaghaei and G. Calic, "Assessing the impact of big data on firm innovation performance: big data is not always better data," *Journal of Business Research*, vol. 108, pp. 147–162, 2020.

[13] Y. Wu, H. Huang, N. Wu, Y. Wang, and M. Z. Wang, "An incentive-based protection and recovery strategy for secure big data in social networks," *Information Sciences*, vol. 508, pp. 79–91, 2020.

[14] C.-c. Qi, "Big data management in the mining industry," *International Journal of Minerals, Metallurgy and Materials*, vol. 27, no. 2, pp. 131–139, 2020.

[15] M. Maksimovic, "The role of green internet of things (G-IoT) and big data in making cities smarter, safer and more sustainable," *International Journal of Computing and Digital Systemss*, vol. 6, no. 4, pp. 175–184, 2017.

[16] C. S. Nandyala and H.-K. Kim, "Green IoT agriculture and healthcare application (GAHA)," *International Journal of Smart Home*, vol. 10, no. 4, pp. 289–300, 2016.

[17] D. X. Li, He. Wu, and L. Shancang, "Internet of things in industries: a survey," *IEEE: Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2233–2243, 2014.

[18] T. Jamal and S. A. Butt, "Malicious node analysis in manets," *International Journal of Information Technology*, vol. 11, no. 4, pp. 859–867, 2019.

[19] S. A. John, "Research directions for the internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 3–9, 2014.

[20] Z. Haider, T. Jamal, M. Asam, S. Butt, and A. Ajaz, "Mitigation of wireless body area networks challenges using cooperation," *International Journal of Security and Its Applications*, vol. 14, no. 1, pp. 15–30, 2020.

[21] S. Gupta, S. Modgil, and A. Gunasekaran, "Big data in lean six sigma: a review and further research directions," *International Journal of Production Research*, vol. 58, no. 3, pp. 947–969, 2020.

[22] S. Mostafavi and W. Shafik, "Fog computing architectures, privacy and security solutions," *Journal of Communications Technology, Electronics and Computer Science*, vol. 24, pp. 1–14, 2019.

[23] G. Koppe, A. L. Meyer, and D. Durstewitz, "Deep learning for small and big data in psychiatry," *Neuropsychopharmacology*, vol. 15, pp. 1–7, 2020.

[24] A. S. H. Abdul-Qawy and T. Srinivasulu, "SEES: a scalable and energy-efficient scheme for green IoT-based heterogeneous wireless nodes," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 4, pp. 1571–1596, 2019.

[25] H. Hassani, X. Huang, and E. Silva, "Big data and climate change," *Big Data and Cognitive Computing*, vol. 3, no. 1, pp. 1–12, 2019.

[26] G. Knieps, "Internet of Things, future networks, and the economics of virtual networks," *Competition and Regulation in Network Industries*, vol. 18, no. 4, pp. 240–255, 2017.

[27] W. Shafik, S. M. Matinkhah, and M. Ghasemzadeh, "Fog-mobile edge performance evaluation and analysis on internet of things," *Journal of Advance Research in Mobile Computing*, vol. 1, no. 3, pp. 1–17, 2019.

[28] D. Feldman, M. Schmidt, and C. Sohler, "Turning big data into tiny data: constant-size coresets for k-Means, PCA, and projective clustering," *SIAM Journal on Computing*, vol. 49, no. 3, pp. 601–657, 2020.

[29] W. Shafik, S. M. Matinkhah, and M. Ghasemzadeh, "A fast machine learning for 5g beam selection for unmanned aerial vehicle applications," *Information Systems & Telecommunication*, vol. 7, no. 28, pp. 262–278, 2019.

[30] S. Suoniemi, L. Meyer-Waarden, A. Munzel, A. R. Zablah, and D. Straub, "Big data and firm performance: the roles of market-directed capabilities and business strategy," *Information & Management*, vol. 57, no. 7, Article ID 103365, 2020.

[31] H. Zhu, "Big data and artificial intelligence modeling for drug discovery," *Annual Review of Pharmacology and Toxicology*, vol. 60, no. 1, pp. 573–589, 2020.

[32] W. Shafik and S. A. Mostafavi, "Knowledge engineering on internet of things through reinforcement learning," *International Journal of Computer Applications*, vol. 177, no. 44, pp. 975–8887, 2019.

[33] S. Shilo, H. Rossman, and E. Segal, "Axes of a revolution: challenges and promises of big data in healthcare," *Nature Medicine*, vol. 26, no. 1, pp. 29–38, 2020.

[34] W. Shafik, S. M. Matinkhah, and M. Ghasemzadeh, "Theoretical understanding of deep learning in UAV biomedical engineering technologies analysis," *SN Computer Science*, vol. 1, no. 6, pp. 1–3, 2020.

[35] E. Hargittai, "Potential biases in big data: omitted voices on social media," *Social Science Computer Review*, vol. 38, no. 1, pp. 10–24, 2020.

[36] Z. Ahmad, G. G. Hamedani, and N. S. Butt, "Recent developments in distribution theory: a brief survey and some new generalized classes of distributions," *Pakistan Journal of Statistics and Operation Research*, vol. 15, pp. 87–110, 2019.

[37] Z. Ahmad, E. Mahmoudi, S. Dey, and S. K. Khosa, "Modeling vehicle insurance loss data using a new member of T-X family of distributions," *Journal of Statistical Theory and Applications*, vol. 19, no. 2, pp. 133–147, 2020.

[38] Z. Ahmad, E. Mahmoudi, G. G. Hamedani, and O. Kharazmi, "New methods to define heavy-tailed distributions with applications to insurance data." *Journal of Taibah University for Science*, vol. 14, no. 1, pp. 359–382, 2020.