Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 9625687, 15 pages
https://doi.org/10.1155/2017/9625687

Research Article

Reliability-Centric Analysis of Offloaded Computation in Cooperative Wearable Applications

Aleksandr Ometov,[1,2] Dmitry Kozyrev,[2] Vladimir Rykov,[2] Sergey Andreev,[1] Yuliya Gaidamaka,[2,3] and Yevgeni Koucheryavy[1]

[1] *Tampere University of Technology, Tampere, Finland*
[2] *RUDN University, Moscow, Russia*
[3] *Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia*

Correspondence should be addressed to Aleksandr Ometov; aleksandr.ometov@tut.fi

Received 11 August 2017; Accepted 7 November 2017; Published 7 December 2017

Academic Editor: Mostafa Zaman Chowdhury

Motivated by the unprecedented penetration of mobile communications technology, this work carefully brings into perspective the challenges related to heterogeneous communications and offloaded computation operating in cases of fault-tolerant computation, computing, and caching. We specifically focus on the emerging augmented reality applications that require reliable delegation of the computing and caching functionality to proximate resource-rich devices. The corresponding mathematical model proposed in this work becomes of value to assess system-level reliability in cases where one or more nearby collaborating nodes become temporarily unavailable. Our produced analytical and simulation results corroborate the asymptotic insensitivity of the stationary reliability of the system in question (under the "fast" recovery of its elements) to the type of the "repair" time distribution, thus supporting the fault-tolerant system operation.

1. Introduction and Motivation

With the rapid proliferation of wireless communications technology, pushed by the avalanche of demanding services, mobile users are preparing to meet much higher levels of service experience than ever before. This poses an unprecedented burden on the existing as well as emerging wireless infrastructures [1]. Indeed, the numbers of connected handheld devices have increased dramatically over the past decade [2], and this growth is only expected to continue further with the introduction of wearables.

Today's transition toward next-generation (5G) mobile technology brings along an astounding variety of networked devices, like low-end wearables, augmented and virtual reality eye-wear, and self-driving cars, among many others [3, 4]. This paradigm shift unlocks a large diversity of completely new applications and scenarios, which not only require conventional resources (e.g., bandwidth) but also impose other important constraints, primarily along the lines of latency and reliability [5]. The latter is due to an ongoing evolution of pre-5G networking ecosystem, as it departs from serving primarily human-to-human (H2H) applications to embracing diverse machine-type communications (MTC) services [6, 7].

However, despite significant progress in circuit design, supported by cheaper production processes [8] and increasingly high pace of technological revolution [9], many application-level tasks are still infeasible to be executed in the resource-limited end-user devices [10]. This is because of different factors, such as reduced computation capabilities, high-speed mobility, and stringent security. The above is especially true for the more constrained Internet-of-Things (IoT) objects, including wearables, which are still designed so that their efficiency is maintained only when completing a certain set of predefined tasks, which may hamper their applicability.

To make matters worse, our present-day devices still considerably depend on energy efficiency of communication [11]. People are therefore forced to charge their carriable and wearable equipment on a daily basis, which is a dramatic departure from the time when mobile phone was able to last for a week on one charging cycle. The energy harvesting and transfer techniques, including wireless charging or generating energy from heat, are still in their infancy, and their mass adoption may not take off any time soon [12], thus leaving the devices with extremely limited battery lifetimes.

Fortunately, in many cases a viable option for "on-the-fly" execution of a processing-hungry task is by offloading (delegating) it to more resource-rich and power-independent node(s). Then, after the input information has been processed remotely (possibly, in a distributed manner), it would be returned to the original constrained device hence saving its resources, as opposed to executing the task locally on the device [13]. Allocating more computing resources in the network and supplying user devices with appropriate communications technology to improve the energy efficiency via offloaded computation pose new challenges. These include the need for reliable wireless connectivity and failure-free remote processing of information, which could be resolved with the advent of communication and computing cooperation (3C) techniques [14, 15].

The development of 3C solutions has a long history behind, as it comes from conventional cloud computing [16], through fog structures [17], all the way to the network edge [18, 19]. This transition, in turn, underpins a new concept of mobile edge computing (MEC) as an emerging computing and caching paradigm [20]. MEC was initially employed in 2013 to facilitate the execution of services at the edge of a network, when IBM and NSN introduced a platform that could run applications on a mobile base station. These early efforts only had local impact [21] and did not consider other important aspects, like application migration and interoperability, among others.

MEC grew to its current importance a year after, when ETSI launched their Industry Specification Group (ISG) for mobile edge computing in 2014. With this specification, MEC aimed to "provide an IT service environment and cloud computing capabilities at the edge of the mobile network." This group also pursues creation of an open ecosystem, where service providers can deploy their applications across multi-vendor MEC platforms, and the target services are expected to include augmented reality, intelligent video acceleration, connected cars, and IoT gateways, to name a few [22]. Once standardization is complete, telecommunication companies will become capable of deploying this service environment in their infrastructures.

In light of the above, the main contributions and the structure of this paper are as follows. We address an important question of whether MEC as part of the 3C system can operate reliably in advanced 5G-grade scenarios, such as "connected law enforcement officer" in urban settings. The subsequent Section 2 offers a brief review of computation offloading and communications techniques for smart mobile devices, mainly focusing on network-edge operation. Further, the use of specialized wearable equipment utilized in the proposed system design is discussed in Section 3. The paper continues with our developed analytical model in Section 4. In Section 5, we deliver the corresponding numerical results. Our study is concluded in the last section.

2. Communications and Computing Cooperation Strategies

2.1. Review of Emerging 3C Solutions. One of the most prominent solutions under the umbrella of 5G ecosystem from the 3C perspective is computation offloading [23]. This paradigm has been coined for resource-constrained and energy-hungry devices, which have difficulty in handling their heavy computation tasks and thus need to delegate those to more resourceful nodes. This allows to make a decisive step forward in the direction of content-oriented networking, in contrast to connection-oriented technology, thus targeting an increase in cooperative network capacity [24].

To this end, the stringent requirements of new applications and services are pushing the network operators to develop diverse communications capabilities in the form of heterogeneous networks (HetNets) [25–27]. For example, cellular macrocells can deliver coverage, while smaller cells may supply capacity, hence providing the users with better quality of experience (QoE) and lowering harmful interference [28]. Therefore, HetNets emerge as advanced architectures that are able to serve demanding user scenarios, by utilizing increasingly diverse wireless technologies across the same geographical area, but posing a complex challenge of backhaul efficiency [29].

Making backhaul more effective than in already deployed HetNet solutions may be costly for the network operators. On the other hand, radio access network (RAN) capacity can potentially be increased by deploying higher numbers of small cells. As one of the approaches to overcome this challenge, various caching techniques have been proposed and utilized [30, 31]. These provide the end users with easy access to content of high demand, which is stored locally instead of being transferred repeatedly over "expensive" communications links, especially during peak hours.

The range of network architectures to be utilized is very wide. Accounting for the fact that any of the proposed techniques can be executed not only on the infrastructure side but also by offloading tasks onto other user devices in proximity, device-to-device (D2D) technologies have recently seen a renewed research interest [32, 33]. Further facilitated by the latest progress in affordable memory capacity, transparent caching in key locations allows boosting the content delivery as well as improving the network resource utilization, even when the same content is not requested by the users simultaneously.

2.2. Emerging User Applications. As follows from the above, capable mobile devices (smartphones, high-end wearables, cars, etc.) are changing the lives of people. Already in 2015, the time spent by the average user online has been up to 5 hours daily [34] and it is commonly anticipated to grow further, since people turn more to their mobile devices as opposed to legacy stationary equipment. This, in turn, creates higher loads due to repetitive content downloads that could however be cached either on the network or on the user side.

A particularly challenging use case in the above context, enabled by the decisive advancements in wearable and computing technologies, is augmented reality (AR), which opens the door to genuinely interactive user experience [35]. Different from virtual reality (VR), AR aims to supplement the real world instead of creating an entirely artificial environment. To this end, various objects in the person's surroundings become the target items for computer-generated annotations, which require complex and often real-time computation. This, in turn, substantiates the need for content delivery transformation, thus producing a diversity of computation-heavy and delay-sensitive AR applications [36].

While modern AR devices are already capable of advanced hardware acceleration [37], the opportunities for software-based data analysis remain rather limited. Today, the data is typically sent to the dedicated processing device by utilizing a cable connection. Hence, leading vendors are actively pursuing the implementation of mobile and truly wireless AR/VR technology [38]. However, the task of pattern matching, especially in the video processing applications, is still very challenging for the standalone smart glasses, and hence distributed/offloaded computation methods rapidly become one of the key areas of interest [39].

2.3. Computation Offloading Techniques. The notion of "cloud" first appeared in the late 70s, thus shaping the concept of distributed storage and computation. Since then, the evolution of cloud-based techniques counted a long history and in this paper we sketch the state of the art in this area only briefly.

(1) Cloud and Fog Computing. Generally, the concept of migrating the computation and data from portable and desktop devices to larger data centers is known as cloud computing [40]. The applications of interest are therefore delivered as services to the end users over the Internet, while the hardware for actually running those is "rented" from the service providers. Cloud-service clients help receive more capacity during the peak demand hours, reduce the device costs, and increase the network utilization dramatically, primarily at the network edge.

Another development in cloud computing is named fog computing [41]. The original definition was, "extension of the cloud computing paradigm (that) provides computation, storage, and networking services between end devices and traditional cloud servers" [42]. Technically, fog computing is a supplement for the conventional cloud computing that allows analyzing localized information "on-the-fly" by taking advantage of the locally deployed servers instead of forwarding all of the data to the remote server.

(2) Mobile Edge Computing. The core idea behind the concept of MEC is to execute the demanding applications even closer to the end user, that is, on the most proximate base stations [43]. The main benefits delivered to both the network operators and the end clients are as follows: (i) lower latency; (ii) higher bandwidth; and (iii) improved location awareness.

(3) Mobile Cloud Computing. Another noteworthy direction in the development of computation offloading from MEC techniques is mobile cloud computing (MCC) [44, 45]. Based on the concept of mobile delegation, MCC leads to decentralized storage and facilitates distribution of execution between various mobile nodes that may operate at the network edge [46].

Solutions for MCC are many, but in cases where wireless connection to the base station (or, more generally, to the cloud) is not reliable, the devices can group together and solve the task collaboratively [47]. Existing multiplatform frameworks already allow migrating a part of the device's code to the neighboring "helper" nodes [48]. When such delegation is complete, the memory, CPU state, and other application-related resources are utilized independently. After the task at hand is finalized, the result is delivered back to the requester device.

In the following subsection, we briefly elaborate on the caching techniques that correspond to the reviewed offloading mechanisms.

2.4. Caching Strategies. Similarly to the case of classifying the computing strategies, two main groups of caching techniques could be identified. One involves using the network infrastructure, while the other is based on utilizing distributed user devices in proximity.

(1) Conventional Caching Strategies. First, infrastructure-based caching strategies are to be considered. In the Long Term Evolution (LTE) cellular networks by 3GPP, two types of locations appear attractive for deploying 5G-grade caches [49]: (i) in the Evolved Packet Core (reflecting the cloud computing paradigm) or (ii) in the RAN on the base station side (representing the MEC concept). For example, caching the content library at the femtocell stations (so-called femtocaching [50]) holds a promise to mitigate the backhaul requirements in next-generation 5G-grade HetNets [51].

(2) Distributed Caching Strategies. On the other hand, mobile user equipment already has a significant amount of data storage at its disposal. Since most of the users tend to consume similar content in their daily life [52] (e.g., news, social network data, blogs, and popular location-based files), the possibilities of utilizing direct D2D links for delivering popular content arise as an attractive solution [53]. Moreover, the costs of such technology integration appear to be more affordable as compared to the infrastructure-based caching.

The discussed development of 3C toward MEC and MCC solutions opens numerous opportunities for emerging 5G applications. The following section outlines our representative scenario of interest, where the goal is to select the most appropriate approach for supporting an advanced wearable application.

3. Utilization of High-End Wearables

In the rest of this paper, we focus on the *distributed* caching and computing strategies within the scope of MEC and MCC paradigms. Assuming that modern user devices are already capable of storing large amounts of information, this data needs to be delivered to more computationally powerful nodes for further processing [54]; see our considered example in Figure 1.

Figure 1: Connected law enforcement officer scenario.

One of the feasible approaches here is to enable dynamic data prefetching, where mobile but more computationally powerful devices (or infrastructure nodes) acquire and deliver the relevant information to less computationally powerful equipment [55]. The latter could become the basis for the following three main applications:

(i) *Caching of popular content*: news, videos, maps, and more could be stored in the devices based on their proximity and then delivered directly if requested by other nodes in the network [56].

(ii) *Incentive mechanisms*: users in proximity may assist the target device based on a certain incentive mechanism, be it operator bonuses, seller discounts, or access to another service.

(iii) *Special cases*: users may be engaged in assisting the dedicated devices of special forces, that is, police, firefighters, and ambulance, among others, thus providing them with the required computing and caching resources upon request.

In all of the above listed cases, there may be situations when the device/network has the knowledge that it will be approached by another device in need later on. Then, the content can be precached in the "helper" device proactively by taking advantage of a reliable connection to the infrastructure, while refraining from utilizing this expensive connectivity for the target requester device [57]. The security, privacy, and anonymity remain as important challenges to facilitate such operation but are left out of scope of this work. The authors assume that all data is encrypted and, for example, pairwise key cryptography is utilized.

3.1. Smart Devices of Interest. Today, various types of special forces are equipped with advanced devices [58], including smart cameras, health monitors, and communications systems. While health monitoring may not require real-time and heavy data analytics [59, 60], the assessment of the surrounding environment can have tremendous demands in terms of communications and computation.

Present-day technology revolution is primarily pushed by the industry giants. For example, Connected Law Enforcement Officer program by Motorola facilitates real-time collaboration experience between different specialized devices to improve the response times within the ecosystem of mission-critical services. This solution makes it possible for teams of people, in the field or in the office, to effectively communicate and cooperate by exchanging voice and data, securely, reliably, and in real time, regardless of network, carrier, or device, whether deployed inside the client's network infrastructure or in the cloud as a hosted/managed solution.

In order to improve the operating efficiency and accelerate decision making, the devices need to collaborate

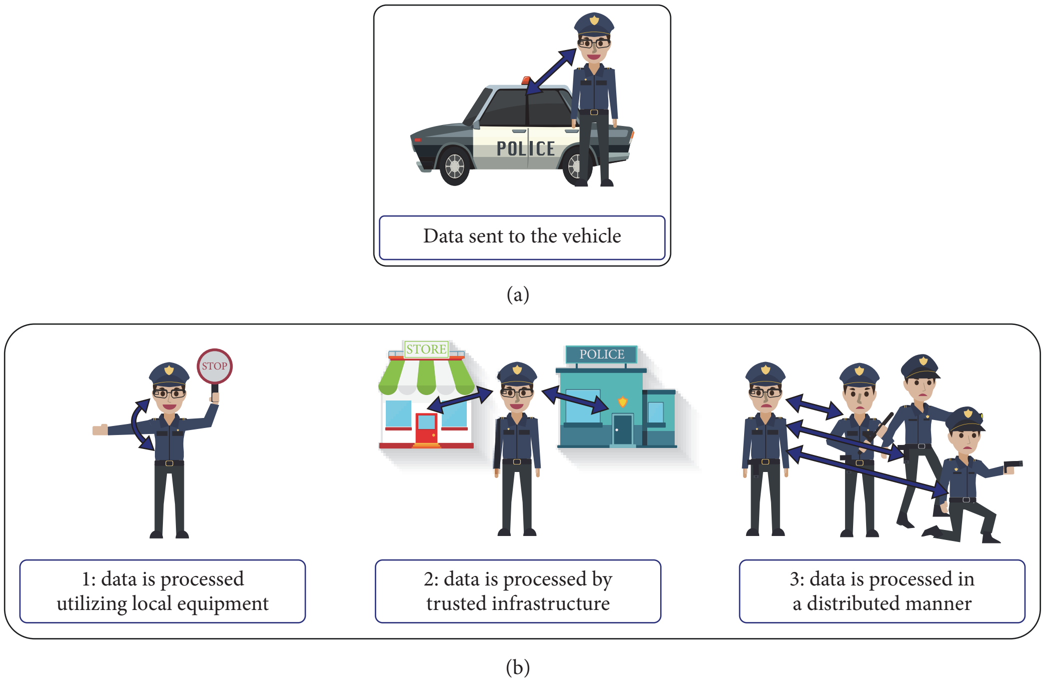

Figure 2: Our considered scenarios for real-time data offloading.

seamlessly and share information "on-the-fly." The considered set of mobile devices typically includes the following:

(i) *Video speaker microphone*: it combines a body-mounted camera with a high-quality radio speaker microphone and a touchscreen interface, for intuitive capture, storage, and management of video, audio, and image evidence.

(ii) *Portable radio*: it provides reliable "push-to-talk" communications as well as superior audio quality for essential day-to-day and emergency two-way radio communications.

(iii) *Handheld smartphone*: it offers reliable broadband connectivity and enables users to enjoy intelligent support of the entire team. With its rugged form factor, powerful data applications, and a quick-response user interface, it delivers "always-on" mission-critical user experience to improve situational awareness for police, firefighter, rescue, and emergency medical services as well as for organizations that support critical infrastructure.

(iv) *Connected car*: it is equipped with, for example, TETRA radio, mobile computing features, and mission-critical 3GPP LTE broadband connectivity; it supports intuitive data applications as well as automatic number plate recognition (ANPR), video surveillance, lights, and sirens functionality, which can be accessed and managed effortlessly. This vehicle is particularly of interest for our case study due to its superior computation and caching capabilities, as well as more stable connectivity to the network infrastructure.

(v) *AR glasses* are expected to join this set soon: they assist the user in a range of tasks and provide real-time assessment of the surrounding environment for faster response [61].

The considered application ecosystem includes real-time video streaming, computer-aided dispatch, unified "push-to-talk," electronic citation, report writing, and data capture. In what follows, we primarily focus on the representative use case where a group of high-tech equipped officers on duty collaborate with each other. However, our findings generally hold valid for a range of similar mission-critical scenarios as discussed below.

3.2. Scenario of Interest. In this work, we concentrate on a study case that involves, for example, daily police officer operation, as demonstrated in Figure 2(a). The considered conventional duties include travel of the officer using the law enforcement vehicle, that is, a car, a motorcycle, or any other means of transport [62]. In this case, the resource-limited device is the AR glasses, which for the most of the time has the possibility of connecting to a resource-rich processing unit. The car can feature the required datasets and hardware to simultaneously process a high number of pattern-matching tasks from the AR glasses thus acting as the MEC/MCC unit. By this means, the police officer's gear can effectively enjoy higher processing power.

Another study case reflects the scenarios, where the AR-user leaves the proximity with the powerful processing unit, as shown in Figure 2(b). This may happen for a long list of reasons, like an emergency call, due to operational causes, and so on. This scenario comprises the following modes:

(1) *Patrolling (b1)*: if the processing unit is far away, the AR glasses may keep analyzing the most crucial

patterns in their vicinity according to the user preferences. Here, the battery resource becomes the main limiting factor, since running complex computation on this single device may significantly degrade energy efficiency.

(2) *Visiting connected buildings (b2)*: if after losing direct connectivity to the processing node (discontinuing pure MCC operation), the AR-equipped user is moving toward, for example, a police station, shopping mall, or hospital, the respective pattern-matching sets can be securely transferred to the trusted infrastructure as in the MEC case. Hence, the AR glasses may be assisted by relying on the distributed pattern-matching functionality and therefore save their limited resources.

(3) *Emergency call (b3)*: if an emergency situation occurs, the officer may need to arrive at the scene as part of the dispatched unit. In this case, the control center has the knowledge on who participates in the mission and can identify the vehicles and, potentially, other available (mobile) stations in proximity. Therefore, the required datasets can be migrated accordingly onto these nearby devices for the AR glasses to take advantage of the distributed collaborative processing.

On the one hand, utilizing "helper" devices with more abundant computation and energy resources becomes the desired option. However, in case a group of police officers turns out to be on a mission without direct access to such MEC/MCC unit (or the connection is not sufficiently reliable), the data in question may be distributed among the group of officers so that the latter could run collaborative computation in a distributed manner; that is, the number of simultaneously processing/caching "servers" becomes more than one.

Since the above cases a and $b1$ do not pose much of a research challenge, we concentrate below on modeling a scenario built around *several* content-caching and processing devices that serve the tagged unit. We are specifically interested in assessing the reliability perspective of such operation, which is discussed in detail in the following section.

4. Proposed Reliability-Centric Modeling

This section introduces a mathematical framework suitable for analyzing the fault-tolerant collaborative operation of tagged and helper devices. In this work, we assume that the "active" (available) and the "repair" (unavailable) operation times for cooperating system elements are independent. The notation employed further on is summarized in Main System Model Notations.

4.1. General Considerations. Consider a random process $\nu(t)$, which corresponds to the number of failed collaborating elements at time t, where the set of states in the system is $E = 0, 1, \ldots, k$. To characterize the behavior of the system in question by relying on a Markov process, we introduce an additional variable $x(t) \in R_+$, which is the overall duration spent at time t for the failed element recovery. We obtain a two-dimensional process $(\nu(t), x(t))$ with an extended steady-state space $\mathscr{E} = \{(0), (1, x), (2, x), \ldots, (k, x)\}$.

We further denote $p_0(t)$ as the probability that at time t the system is in the state $i = 0$, and $p_i(t; x)$ is the probability density function (in continuous component) that captures the probability that at time t the system is in state i ($i = \overline{1, k}$). The time taken to "repair" the failed element is in the range $(x, x + dx)$. The probabilities in question are calculated as

$$
\begin{aligned}
p_0(t) &= P\{\nu(t) = 0\}, \\
p_1(t, x)\,dx &= P\{\nu(t) = 1, x < x(t) < x + dx\}, \\
p_2(t, x)\,dx &= P\{\nu(t) = 2, x < x(t) < x + dx\}, \\
&\vdots \\
p_k(t, x)\,dx &= P\{\nu(t) = k, x < x(t) < x + dx\}.
\end{aligned} \tag{1}
$$

Consider the probabilities of two events that are important for our further discussion.

First Event. A system element has failed during time Δ under the condition that it has already worked for x units of time before its failure.

$$
\begin{aligned}
P\{x \le A < x + \Delta \mid A \ge x\} &= \frac{P\{x \le A < x + \Delta\}}{P\{A \ge x\}} \\
&= \frac{A(x + \Delta) - A(x)}{1 - P\{A < x\}} = \frac{a(x)\Delta}{1 - A(x)} = \frac{\alpha e^{-\alpha x}\Delta}{e^{-\alpha x}} \\
&= \alpha\Delta + o(\Delta).
\end{aligned} \tag{2}
$$

Second Event. Starting from the considered time instant, an element will be restored during the time Δ, provided that it has already been in "repair" for x time units.

$$
\begin{aligned}
P\{x \le B < x + \Delta \mid B \ge x\} &= \frac{P\{x \le B < x + \Delta\}}{P\{B \ge x\}} \\
&= \frac{B(x + \Delta) - B(x)}{1 - P\{B < x\}} = \frac{b(x)\Delta}{1 - B(x)} = \delta(x)\Delta.
\end{aligned} \tag{3}
$$

Steady-state transition diagram corresponding to the operation of the system in question is illustrated in Figure 3.

By applying the law of total probability, we transition to Kolmogorov's system of differential equations, which allows obtaining the steady-state probabilities of the system at hand. Consider the probability $p_0(t + \Delta)$ that corresponds to the case where there are 0 failed elements at time instant $t + \Delta$. This situation may appear (i) if at time instant t either there were 0 failed elements or the active one has not failed during time instant Δ (with probability $1 - \alpha\Delta$) or (ii) if there was 1 failed element and it has been restored during time Δ (with probability $\delta(x)\Delta$), while no other active element has failed during this time (with probability $(1 - \alpha\Delta)$).

$$
\begin{aligned}
p_0(t + \Delta) &= P\{\nu(t + \Delta) = 0\} \\
&= p_0(t)(1 - \alpha\Delta)
\end{aligned}
$$

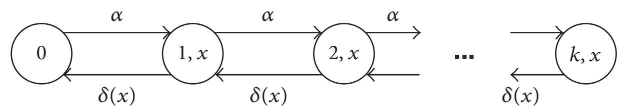

Figure 3: Steady-state transitions in the considered system.

$$
\begin{aligned}
&+ \int_0^t p_1(t,x) \cdot \delta(x) \cdot \Delta \cdot (1-\alpha\Delta)\,dx \\
&= p_0(t)(1-\alpha\Delta) \\
&+ \int_0^t p_1(t,x) \cdot \delta(x) \cdot \Delta\,dx.
\end{aligned} \tag{4}
$$

Hence, we establish the state-related equations as

$$
p_0(t+\Delta) = p_0(t) \cdot (1-\alpha\Delta) + \int_0^t p_1(t,x)\,\delta(x)\,\Delta\,dx, \tag{5}
$$

and, similarly,

$$
\begin{aligned}
&p_1(t+\Delta, x+\Delta)\,dx \\
&= P\{\nu(t+\Delta) = 1, x+\Delta < x(t+\Delta) < x+\Delta+dx\} \\
&= p_1(t,x)\,dx \cdot (1-\alpha\Delta) \cdot (1-\delta(x) \cdot \Delta).
\end{aligned} \tag{6}
$$

After the necessary transformations, we obtain

$$
p_1(t+\Delta, x+\Delta) = p_1(t,x) \cdot (1-\alpha\Delta) \cdot (1-\delta(x) \cdot \Delta). \tag{7}
$$

At the next step, we have

$$
\begin{aligned}
&p_2(t+\Delta, x+\Delta)\,dx \\
&= P\{\nu(t+\Delta) = 2, x+\Delta < x(t+\Delta) < x+\Delta+dx\} \\
&= p_2(t,x)\,dx \cdot (1-\alpha\Delta) \cdot (1-\delta(x) \cdot \Delta) \\
&\quad + p_1(t,x)\,dx \cdot \alpha\Delta.
\end{aligned} \tag{8}
$$

Further, we arrive at the equation

$$
\begin{aligned}
p_2(t+\Delta, x+\Delta) &= p_2(t,x) \cdot (1-\alpha\Delta) \cdot (1-\delta(x) \cdot \Delta) \\
&\quad + p_1(t,x) \cdot \alpha\Delta.
\end{aligned} \tag{9}
$$

Similarly, we produce the equations for $p_i(t+\Delta, x+\Delta)$, $i = \overline{3, k-1}$ as

$$
\begin{aligned}
p_i(t+\Delta, x+\Delta) &= p_i(t,x) \cdot (1-\alpha\Delta) \cdot (1-\delta(x)\,\Delta) \\
&\quad + p_{i-1}(t,x)\,\alpha\Delta, \quad i = \overline{3, k-1}.
\end{aligned} \tag{10}
$$

At the next step, we establish

$$
\begin{aligned}
&p_k(t+\Delta, x+\Delta)\,dx \\
&= P\{\nu(t+\Delta) = k, x+\Delta < x(t+\Delta) < x+\Delta+dx\} \\
&= p_k(t,x)\,dx \cdot (1-\delta(x) \cdot \Delta) + p_{k-1}(t,x)\,dx \cdot \alpha\Delta \\
&\quad \cdot (1-\delta(x) \cdot \Delta) \\
&= p_k(t,x)\,dx \cdot (1-\delta(x) \cdot \Delta) + p_{k-1}(t,x)\,dx \cdot \alpha\Delta \\
&\quad - p_{k-1}(t,x)\,dx \cdot \alpha \cdot \delta(x) \cdot \Delta^2 \\
&= p_k(t,x)\,dx \cdot (1-\delta(x) \cdot \Delta) + p_{k-1}(t,x)\,dx \cdot \alpha\Delta,
\end{aligned} \tag{11}
$$

where $\Delta^2 \ll \Delta \Rightarrow o(\Delta^2) \rightarrow 0$.

Then, we obtain for the general case the equation

$$
\begin{aligned}
p_k(t+\Delta, x+\Delta) &= p_k(t,x) \cdot (1-\delta(x) \cdot \Delta) \\
&\quad + p_{k-1}(t,x) \cdot \alpha\Delta.
\end{aligned} \tag{12}
$$

The boundary conditions in our system are

$$
\begin{aligned}
p_1(t+\Delta, 0) &= p_0(t) \cdot \alpha + \int_0^t p_2(t,x)\,\delta(x)\,dx, \\
p_2(t+\Delta, 0) &= \int_0^t p_3(t,x)\,\delta(x)\,dx, \\
&\vdots \\
p_{k-1}(t+\Delta, 0) &= \int_0^t p_k(t,x)\,\delta(x)\,dx.
\end{aligned} \tag{13}
$$

After the transformations, and based on (5), we establish

$$
\begin{aligned}
&p_0(t+\Delta) \\
&\quad = p_0(t) \cdot (1-\alpha\Delta) + \int_0^t p_1(t,x) \cdot \delta(x) \cdot \Delta\,dx, \\
&p_0(t+\Delta) \\
&\quad = p_0(t) - p_0(t) \cdot \alpha\Delta + \int_0^t p_1(t,x) \cdot \delta(x) \cdot \Delta\,dx, \\
&p_0(t+\Delta) - p_0(t) \\
&\quad = -p_0(t) \cdot \alpha\Delta + \int_0^t p_1(t,x) \cdot \delta(x) \cdot \Delta\,dx, \\
&\frac{1}{\Delta} \cdot \left(p_0(t+\Delta) - p_0(t)\right) \\
&\quad = -p_0(t) \cdot \alpha + \int_0^t p_1(t,x) \cdot \delta(x)\,dx.
\end{aligned} \tag{14}
$$

Passing to the limit $\Delta \to 0$, we obtain

$$\frac{dp_0(t)}{dt} = -\alpha \cdot p_0(t) + \int_0^t p_1(t,x) \cdot \delta(x)\, dx. \quad (15)$$

It follows from (7) that

$$\begin{aligned} p_1(t+\Delta, x+\Delta) &= p_1(t,x) \cdot (1-\alpha\Delta) \cdot (1-\delta(x)\cdot\Delta) \\ &= p_1(t,x) - p_1(t,x)\cdot\alpha\Delta - p_1(t,x)\cdot\delta(x)\cdot\Delta, \\ p_1(t+\Delta, x+\Delta) &- p_1(t,x+\Delta) + p_1(t,x+\Delta) \\ &- p_1(t,x) = -p_1(t,x)\,\alpha\Delta - p_1(t,x)\,\delta(x)\,\Delta, \end{aligned} \quad (16)$$

$$\begin{aligned} &\frac{p_1(t+\Delta, x+\Delta) - p_1(t,x+\Delta)}{\Delta} \\ &+ \frac{p_1(t,x+\Delta) - p_1(t,x)}{\Delta} = -p_1(t,x)\,\alpha \\ &- p_1(t,x)\,\delta(x). \end{aligned} \quad (17)$$

Passing to the limit $\Delta \to 0$, we calculate

$$\frac{\partial p_1(t,x+\Delta)}{\partial t} + \frac{\partial p_1(t,x)}{\partial x} = -(\alpha+\delta(x)) \cdot p_1(t,x). \quad (18)$$

Similarly, based on (9), we have

$$\begin{aligned} &\frac{p_2(t+\Delta, x+\Delta) - p_2(t,x+\Delta)}{\Delta} \\ &+ \frac{p_2(t,x+\Delta) - p_2(t,x)}{\Delta} \\ &= -p_2(t,x)\,(\alpha+\delta(x)) + p_1(t,x)\,\alpha. \end{aligned} \quad (19)$$

Further, passing to the limit $\Delta \to 0$, we obtain

$$\begin{aligned} &\frac{\partial p_2(t,x+\Delta)}{\partial t} + \frac{\partial p_2(t,x)}{\partial x} \\ &= -(\alpha+\delta(x)) \cdot p_2(t,x) + \alpha p_1(t,x). \end{aligned} \quad (20)$$

Finally, it follows from (10) that

$$\begin{aligned} p_k(t+\Delta, x+\Delta) &= p_k(t,x) \cdot (1-\delta(x)\cdot\Delta) \\ &+ p_{k-1}(t,x)\cdot\alpha\Delta, \\ p_k(t+\Delta, x+\Delta) &= p_k(t,x) - p_k(t,x)\cdot\delta(x)\cdot\Delta \\ &+ p_{k-1}(t,x)\cdot\alpha\Delta, \\ p_k(t+\Delta, x+\Delta) &- p_k(t,x+\Delta) + p_k(t,x+\Delta) \\ &- p_k(t,x) = -p_k(t,x)\cdot\delta(x) + p_{k-1}(t,x)\cdot\alpha, \\ &\frac{p_k(t+\Delta, x+\Delta) - p_k(t,x+\Delta)}{\Delta} \\ &+ \frac{p_k(t,x+\Delta) - p_k(t,x)}{\Delta} = -\delta(x)\cdot p_k(t,x) + \alpha \\ &\cdot p_{k-1}(t,x). \end{aligned} \quad (21)$$

Passing to the limit $\Delta \to 0$, we find

$$\begin{aligned} &\frac{\partial p_k(t,x+\Delta)}{\partial t} + \frac{\partial p_k(t,x)}{\partial x} \\ &= -\delta(x)\cdot p_k(t,x) + \alpha \cdot p_{k-1}(t,x). \end{aligned} \quad (22)$$

The boundary conditions with respect to the passage to the limit $\Delta \to 0$ are

$$\begin{aligned} p_1(t,0) &= \alpha \cdot p_0(t) + \int_0^t p_2(t,x)\,\delta(x)\,dx, \\ p_2(t,0) &= \int_0^t p_3(t,x)\,\delta(x)\,dx, \\ &\vdots \\ p_{k-1}(t,0) &= \int_0^t p_k(t,x)\,\delta(x)\,dx. \end{aligned} \quad (23)$$

Since the process at hand is positively recurrent (due to finite state space as well as positive and bounded transition rates), there exist limit state probabilities under $t \to \infty$, which coincide with the stationary probabilities as

$$\begin{aligned} p_0 &= \lim_{t\to\infty} p_0(t), \\ p_i(x) &= \lim_{t\to\infty} p_i(t;x) \quad (i \in \{1,\ldots,k\}). \end{aligned} \quad (24)$$

Hence, we arrive at the following system of equations for the steady-state mode (i.e., the system of equilibrium equations) in the form

$$\begin{aligned} \alpha p_0 &= \int_0^\infty p_1(x)\cdot\delta(x)\,dx, \\ \frac{dp_1(x)}{dx} &= -(\alpha+\delta(x))\cdot p_1(x), \\ \frac{dp_2(x)}{dx} &= -(\alpha+\delta(x))\cdot p_2(x) + \alpha\cdot p_1(x), \\ &\vdots \\ \frac{dp_{k-1}(x)}{dx} &= -(\alpha+\delta(x))\cdot p_{k-1}(x) + \alpha\cdot p_{k-2}(x), \\ \frac{dp_k(x)}{dx} &= -\delta(x)\cdot p_k(x) + \alpha\cdot p_{k-1}(x), \\ p_1(0) &= \alpha\cdot p_0 + \int_0^\infty p_2(x)\cdot\delta(x)\,dx, \\ p_2(0) &= \int_0^\infty p_3(x)\cdot\delta(x)\,dx, \\ &\vdots \\ p_{k-1}(0) &= \int_0^\infty p_k(x)\cdot\delta(x)\,dx. \end{aligned} \quad (25)$$

As discussed previously, the established set of equations has an analytical closed-form solution, which we present in the following subsection for the special case of $k = 3$ redundant elements. This formulation, after the integration over the variable x, produces a result for the macrostate probabilities in the system. A macrostate probability p_i in the steady-state mode represents the limiting time-average fraction of time spent by the system in state i. Note that the sum $\sum_{i=0}^{k-1} p_i$ corresponds to the steady-state probability that at least one of k reserve (collaborating) elements is operational; that is, the system is functioning properly. Hence, in the general case we define the stationary reliability of the system as the steady-state probability of its failure-free operation, that is, $1 - p_k$.

Note that the established set of equations has an analytical closed-form solution for an arbitrary number of system elements. By increasing k, calculation of the undetermined coefficients becomes rather "bulky" while solving the corresponding inhomogeneous linear ordinary differential equations. In the following subsection, we consider an algorithm for solving the target system of differential equations in the special case.

4.2. Considerations for Target Scenario. To demonstrate the applicability of our general system modeling in the considered scenario, we further focus on the characteristic use case where a group of four police officers (one of them equipped with the AR glasses) are entering the building, while the squad leader wears a resource-hungry device that requires assistance from the devices of fellow officers. The squad moves along as a group but the connection to the assisting equipment may become unreliable (unavailable) due to unconstrained group mobility [63].

We begin with

$$\begin{aligned} \alpha p_0 &= \int_0^\infty p_1(x)\cdot\delta(x)\,dx, \\ \frac{dp_1(x)}{dx} &= -(\alpha+\delta(x))\cdot p_1(x), \\ \frac{dp_2(x)}{dx} &= -(\alpha+\delta(x))\cdot p_2(x) + \alpha\cdot p_1(x), \\ \frac{dp_3(x)}{dx} &= -\delta(x)\cdot p_3(x) + \alpha\cdot p_2(x), \\ p_1(0) &= \alpha\cdot p_0 + \int_0^\infty p_2(x)\cdot\delta(x)\,dx, \\ p_2(0) &= \int_0^\infty p_3(x)\cdot\delta(x)\,dx. \end{aligned} \tag{26}$$

Further, we focus on the solution obtained with Kolmogorov's system of differential equations. Let us initially consider the second equation

$$\begin{aligned} \frac{dp_1(x)}{dx} &= -(\alpha+\delta(x))\cdot p_1(x), \\ \frac{dp_1(x)}{p_1(x)} &= -(\alpha+\delta(x))\,dx, \\ \frac{dp_1(x)}{p_1(x)} &= -\delta(x)\,dx - \alpha dx, \\ \frac{dp_1(x)}{p_1(x)} &= -\frac{b(x)}{1-B(x)}dx - \alpha\,dx, \\ d\ln p_1(x) &= d\ln(1-B(x)) + d\ln e^{-\alpha x}, \\ d\ln p_1(x) &= d\ln(1-B(x))\cdot e^{-\alpha x}, \\ \ln p_1(x) &= \ln\left(1-B(x)\cdot e^{-\alpha x}\right) + \ln C_1, \\ \ln p_1(x) &= \ln\left(1-B(x)\cdot e^{-\alpha x}\cdot C_1\right), \\ p_1(x) &= C_1\cdot e^{-\alpha x}\cdot(1-B(x)). \end{aligned} \tag{27}$$

We solve the differential equations for $p_2(x)$ and $p_3(x)$ by applying the parameters variation method.

$$\begin{aligned} p_2(x) &= \left(\alpha C_1 x + C_2\right)\cdot e^{-\alpha x}\cdot(1-B(x)), \\ p_3(x) &= C_3(1-B(x)) - \left(C_2 + C_1(\alpha x+1)\right)\cdot e^{-\alpha x} \\ &\quad \cdot(1-B(x)). \end{aligned} \tag{28}$$

To express the coefficients C_2 and C_3 in terms of C_1, we utilize the boundary conditions

$$\begin{aligned} p_0 &= C_1\cdot\alpha^{-1}\cdot\tilde{b}(\alpha), \\ p_1(x) &= C_1\cdot e^{-\alpha x}\cdot(1-B(x)), \\ p_2(x) &= C_1\cdot e^{-\alpha x}\cdot(1-B(x))\cdot\frac{\left(1-\tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)}, \\ p_3(x) &= C_1\cdot\left(1-e^{-\alpha x}\right)\cdot(1-B(x))\cdot\frac{\left(1-\tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)}. \end{aligned} \tag{29}$$

Further, we evaluate the stationary probabilities for the macrostate p_1.

$$\begin{aligned} p_1 &= \int_0^\infty p_1(x)\,dx = \int_0^\infty C_1\cdot e^{-\alpha x}\cdot(1-B(x))\,dx \\ &= C_1\cdot\left(\int_0^\infty e^{-\alpha x}dx - \int_0^\infty e^{-\alpha x}\cdot B(x)\,dx\right) \\ &= C_1\cdot\left(\frac{1}{\alpha} - \frac{1}{\alpha}\tilde{b}(\alpha)\right) = C_1\cdot\left(\frac{1-\tilde{b}(\alpha)}{\alpha}\right), \end{aligned} \tag{30}$$

where $\tilde{b}(\alpha) = \alpha\tilde{B}(\alpha)$ is an LT for the PDF $b(x)$ at point $s = \alpha$.

Next, for the probability p_2 we have

$$\begin{aligned} p_2 &= \int_0^\infty p_2(x)\,dx \\ &= \int_0^\infty C_1\cdot e^{-\alpha x}\cdot(1-B(x))\cdot\frac{\left(1-\tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)}dx \end{aligned}$$

$$= C_1 \cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \int_0^\infty e^{-\alpha x} \cdot (1 - B(x))\, dx$$
$$= C_1 \cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \cdot \left(\frac{1 - \tilde{b}(\alpha)}{\alpha}\right)$$
$$= C_1 \cdot \frac{\left(1 - \tilde{b}(\alpha)\right)^2}{\alpha \cdot \tilde{b}(\alpha)}. \tag{31}$$

Further, for p_3 we obtain

$$\begin{aligned}
p_3 &= \int_0^\infty p_3(x)\, dx = \int_0^\infty C_1 \cdot \left(1 - e^{-\alpha x}\right) \cdot (1 - B(x)) \\
&\cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} dx = C_1 \cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \cdot \int_0^\infty \left(1 - e^{-\alpha x}\right) \\
&\cdot (1 - B(x))\, dx = C_1 \cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \\
&\times \left(\int_0^\infty (1 - B(x))\, dx - \int_0^\infty e^{-\alpha x} dx \right. \\
&\left. + \int_0^\infty e^{-\alpha x} \cdot B(x)\, dx\right) = C_1 \cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \\
&\cdot \left(-\int_0^\infty x \cdot (-b(x))\, dx - \frac{1}{\alpha} + \frac{1}{\alpha}\tilde{b}(\alpha)\right) = C_1 \\
&\cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \cdot \left(\int_0^\infty x \cdot b(x)\, dx - \frac{1}{\alpha} + \frac{1}{\alpha}\tilde{b}(\alpha)\right) \\
&= C_1 \cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \cdot \left(EB - \frac{1}{\alpha} + \frac{1}{\alpha}\tilde{b}(\alpha)\right) = C_1 \\
&\cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \cdot \left(b - \frac{1}{\alpha} + \frac{1}{\alpha}\tilde{b}(\alpha)\right) = C_1 \\
&\cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \cdot \left(b - \frac{1 - \tilde{b}(\alpha)}{\alpha}\right),
\end{aligned} \tag{32}$$

where b is the parameter of random "repair" time (mathematical expectation for the failed element).

Therefore, the probabilities of the system macrostates are given as

$$p_0 = C_1 \cdot \frac{\tilde{b}(\alpha)}{\alpha},$$
$$p_1 = C_1 \cdot \left(\frac{1 - \tilde{b}(\alpha)}{\alpha}\right),$$
$$p_2 = C_1 \cdot \frac{\left(1 - \tilde{b}(\alpha)\right)^2}{\alpha \cdot \tilde{b}(\alpha)},$$
$$p_3 = C_1 \cdot \frac{\left(1 - \tilde{b}(\alpha)\right)}{\tilde{b}(\alpha)} \cdot \left(b - \frac{1 - \tilde{b}(\alpha)}{\alpha}\right). \tag{33}$$

By applying the normalization condition $p_0 + p_1 + p_2 + p_3 = 1$, we establish the constant C_1 as

$$C_1 = \frac{\alpha \cdot \tilde{b}(\alpha)}{\rho^{-1}\left(1 - \tilde{b}(\alpha)\right) + \tilde{b}(\alpha)}, \tag{34}$$

where $\rho^{-1} = EB/EA = b \cdot \alpha$.

Finally, we calculate the following steady-state probabilities for the considered "cold-standby repairable system" $\langle M_3/GI/1\rangle$ as

$$\begin{aligned}
p_0 &= \frac{\tilde{b}^2(\alpha)}{\rho^{-1}\left(1 - \tilde{b}(\alpha)\right) + \tilde{b}(\alpha)}, \\
p_1 &= \frac{\tilde{b}(\alpha)\left(1 - \tilde{b}(\alpha)\right)}{\rho^{-1}\left(1 - \tilde{b}(\alpha)\right) + \tilde{b}(\alpha)}, \\
p_2 &= \frac{\left(1 - \tilde{b}(\alpha)\right)^2}{\rho^{-1}\left(1 - \tilde{b}(\alpha)\right) + \tilde{b}(\alpha)}, \\
p_3 &= \frac{\left(1 - \tilde{b}(\alpha)\right)\left(\rho^{-1} - 1 + \tilde{b}(\alpha)\right)}{\rho^{-1}\left(1 - \tilde{b}(\alpha)\right) + \tilde{b}(\alpha)},
\end{aligned} \tag{35}$$

where $\rho^{-1} = b \cdot \alpha$.

As can be concluded from the produced expressions, there is a clear dependency of the system steady-state probabilities on the type of "repair" time distribution. However, this dependency becomes vanishingly small under the "fast repair" of failed elements. The numerical results are reported in the following section and they support this hypothesis.

5. Representative Numerical Results

This section summarizes selected numerical results obtained after applying our proposed analytical model to the scenario of interest described above. To this end, Figure 4 demonstrates the stationary system reliability $1 - p_3$ versus the relative recovery rate ρ for four different special cases of the $\langle M_3 \mid GI \mid 1\rangle$ model, where GW stands for Gnedenko-Weibull distribution [64].

As it can be observed in the figure, the stationary reliability $1 - p_3$ becomes close to 1 as ρ tends to infinity irrespective of the shapes of the activity and repair time distributions. However, it reaches the high value of $1 - p_3 = 0.99$ at different values of ρ. This characterizes the sensitivity of the system stationary reliability with respect to the shape of the activity and repair time distributions (note different rates here). We can assess the rate of convergence of $1 - p_3$ with the means of quantiles for a given reliability level. The corresponding simulations have been conducted with ρ from 1 to 25 for 10,000 time slots over 200 executions, thus making it possible to reach the accuracy of 99.999.

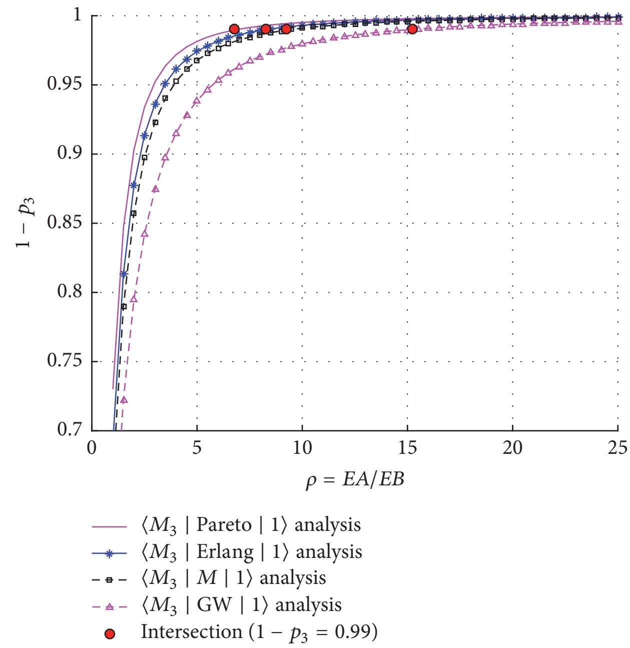

Figure 4: Analytical results for stationary system reliability $1 - p_3$ versus ρ for various "repair" time distributions.

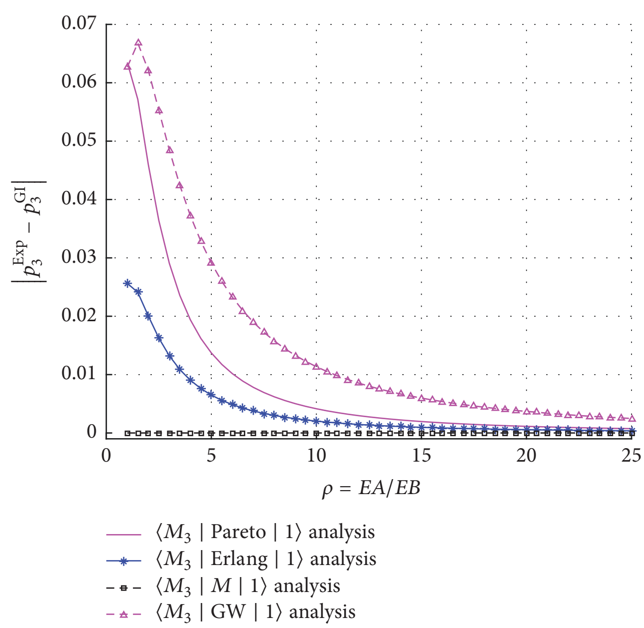

Figure 5: Absolute difference of values for system stationary reliability $1 - p_3$ versus ρ for various "repair" time distributions.

Further, the similarity measure (absolute difference) between the plots in Figure 4 is presented in Figure 5. The results support our above statement about high asymptotic insensitivity of the system stationary reliability, which is clearly observed due to the proximity of the corresponding curves. For instance, starting from $\rho = 20$, all of the curves are almost indistinguishable.

We also note that explicit analytical expressions for the steady-state distribution of the system under consideration are not always easy to obtain. For this case, we construct a simulation model that approximates the analytical results for the system in question. We conduct our simulations by following the standard discrete event modeling approach. Previously discussed analytical values of the stationary system reliability $1-p_3$ versus ρ (see Figure 4) are supported with simulated data, and the results are displayed in Figure 6. The differences between the produced curves become marginal very quickly. Even for the relatively small values of ρ, the stationary reliability of the system $1 - p_3$ is already very close to 1 for all the four cases.

Comparing the results obtained with both analytical and simulation approaches, we confirm that they are in close

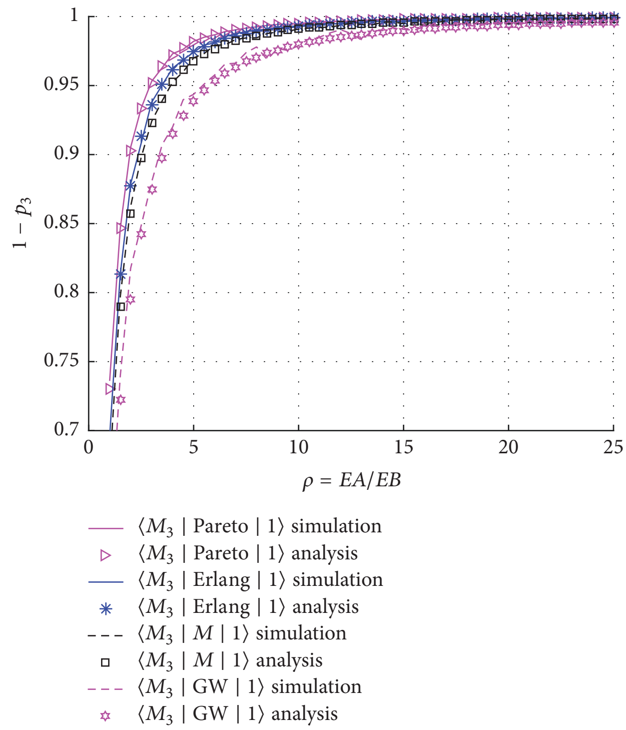

Figure 6: Analytical and simulation results of stationary system reliability $1 - p_3$ versus ρ (averaged values based on 200 replications) for various "repair" time distributions.

agreement. Hence, simulations can be employed in the cases where explicit analytical solution cannot be achieved due to its complexity or as part of a more complex simulation-based methodology that targets in-depth assessment of the fault-tolerant operation of MEC/MCC.

6. Main Conclusions

Today, mobile communications technology and resource-hungry applications are all over the market. As users expect to receive increasingly high levels of service experience, this introduces a burden on the existing wireless communications mechanisms. This work explores the challenge of delegating computing and caching functionality as the users move away from the network "center" to its edge, therefore supporting the emerging MEC paradigm.

With regard to the main contribution of this work, we propose a powerful mathematical methodology that allows assessing system-level reliability in cases where the "helper" nodes fail to provide reliable performance, thus allowing analysis of fault-tolerant MEC/MCC operation. The demonstrated modeling results reveal asymptotic insensitivity of the stationary reliability of the system under the "fast" recovery of its elements to the type of the "repair" time distribution.

In addition, the obtained steady-state stationary system distribution enables us to assess other operational variables, primarily with respect to the "helper" node performance, since the corresponding weighted coefficients may be added to the analysis readily. This permits characterizing the average stationary system performance where the assisting devices may belong to different classes.

Main System Model Notations

k:	Number of cooperating elements
$\nu(t)$:	Number of failed elements at time t
Δ:	Short time interval
A:	Time to failure (random variable)
α:	Failure rate of elements
$A(x) = 1 - e^{-\alpha x}, x \geq 0$:	Cumulative distribution function (CDF) of A
$a(x) = \alpha e^{-\alpha x}$:	Probability density function (PDF) of the random variable A
B:	Time to repair (random variable)
$B(x)$:	CDF of the random variable B
$b(x)$:	PDF of the random variable B
$\tilde{b}(s) = \int_0^\infty e^{-sx} b(x)dx$:	Laplace transform (LT) of the PDF $b(x)$
$EA = 1/\alpha$:	Mean time to failure of the main element
$b = EB$:	Mean time to repair the failed element
$\rho = EA/EB$:	Repair rate, relative recovery rate
$\delta(x) = b(x)/(1 - B(x))$:	Conditional PDF of the residual repair time of the element, which has already spent time t in the repair state.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The publication was supported by the Ministry of Education and Science of the Russian Federation (Project no. 2.882.2017/4.6).

References

[1] Cisco, *Global mobile data traffic forecast 2016–2021*, White Paper, 2017.

[2] Ericsson, "Ericsson mobility report: On the pulse of the Networked Society," http://www.ericsson.com/mobility-report, November 2017.

[3] J. G. Andrews, S. Buzzi, and W. Choi, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[4] J. Orlosky, K. Kiyokawa, and H. Takemura, "Virtual and augmented reality on the 5G highway," *Journal of Information Processing*, vol. 25, pp. 133–141, 2017.

[5] A. Orsino, A. Ometov, G. Fodor et al., "Effects of Heterogeneous Mobility on D2D- and Drone-Assisted Mission-Critical MTC in 5G," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 79–87, 2017.

[6] S. Hamdoun, A. Rachedi, and Y. Ghamri-Doudane, "A flexible M2M radio resource sharing scheme in LTE networks within an H2H/M2M coexistence scenario," in *Proceedings of the IEEE International Conference on Communications, ICC 2016*, Malaysia, May 2016.

[7] K. Zheng, F. Hu, W. Wang, W. Xiang, and M. Dohler, "Radio resource allocation in LTE-advanced cellular networks with M2M communications," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 184–192, 2012.

[8] "Smaller, faster, cheaper, over: the future of computer chips," The New York Times, 2015: https://www.nytimes.com/2015/09/27/technology/smaller-faster-cheaper-over-the-future-of-computer-chips.html?20r=0.

[9] "Why the wearable tech revolution has only just begun," The Telegraph, 2016: http://www.telegraph.co.uk/sponsored/technology/technology-trends/12044628/wearable-tech.html.

[10] M. Paterson and M. R. Glass, "The world through Glass: developing novel methods with wearable computing for urban videographic research," *Journal of Geography in Higher Education*, vol. 39, no. 2, pp. 275–287, 2015.

[11] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. U. Hassan, and C. Yuen, "Energy efficiency tradeoff mechanism towards wireless green communication: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 686–705, 2016.

[12] O. Galinina, H. Tabassum, K. Mikhaylov, S. Andreev, E. Hossain, and Y. Koucheryavy, "On feasibility of 5G-grade dedicated RF charging technology for wireless-powered wearables," *IEEE Wireless Communications Magazine*, vol. 23, no. 2, pp. 28–37, 2016.

[13] O. Galinina, A. Pyattaev, K. Johnsson, A. Turlikov, S. Andreev, and Y. Koucheryavy, "Assessing System-Level Energy Efficiency of mmWave-Based Wearable Networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 923–937, 2016.

[14] K. S. Deepak and A. V. Babu, "Improving energy efficiency of incremental relay based cooperative communications in wireless body area networks," *International Journal of Communication Systems*, vol. 28, no. 1, pp. 91–111, 2013.

[15] S. Ko, K. Huang, S. Kim, and H. Chae, "Live prefetching for mobile computation offloading," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3057–3071, 2017.

[16] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Network*, vol. 29, no. 2, pp. 6–14, 2015.

[17] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-computing-based radio access networks: Issues and challenges," *IEEE Network*, vol. 30, no. 4, pp. 46–53, 2016.

[18] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: the communication perspective," https://arxiv.org/abs/1701.01090.

[19] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-Efficient Resource Allocation for Mobile-Edge Computation Offloading," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1397–1411, 2017.

[20] J. O. Fajardo, I. Taboada, and F. Liberal, "Radio-aware service-level scheduling to minimize downlink traffic delay through mobile edge computing," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 158, pp. 121–134, 2015.

[21] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, A. Neal et al., "Mobile-edge computing introductory technical white paper," White Paper, Mobile-edge Computing (MEC) industry initiative, 2014.

[22] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing: A key technology towards 5G," *ETSI White Paper*, vol. 11, 2015.

[23] L. Person and P. Gotsurve, "World LTE Market – Opportunities and Forecasts, 2012-2020," Allied Market Research IC 14144, April 2014, p. 107.

[24] M. Zhang, H. Luo, and H. Zhang, "A survey of caching mechanisms in information-centric networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1473–1499, 2015.

[25] D. López-Pérez, I. Güvenç, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang, "Enhanced inter-cell interference coordination challenges in heterogeneous networks," *IEEE Wireless Communications Magazine*, vol. 18, no. 3, pp. 22–30, 2011.

[26] B. A. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: past, present, and future of programmable networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617–1634, 2014.

[27] M. Kountouris and N. Pappas, "HetNets and massive MIMO: Modeling, potential gains, and performance analysis," in *Proceedings of the 3rd IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications, IEEE APWC 2013*, pp. 1319–1322, Italy, September 2013.

[28] H. Zhang, X. Chu, W. Guo, and S. Wang, "Coexistence of Wi-Fi and heterogeneous small cell networks sharing unlicensed spectrum," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 158–164, 2015.

[29] S. Tombaz, P. Monti, F. Farias, M. Fiorani, L. Wosinska, and J. Zander, "Is backhaul becoming a bottleneck for green wireless access networks?" in *Proceedings of the 1st IEEE International Conference on Communications, ICC 2014*, pp. 4029–4035, Australia, June 2014.

[30] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, "Cache in the air: exploiting content caching and delivery techniques for 5G systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 131–139, 2014.

[31] X. Wang, X. Li, V. C. M. Leung, and P. Nasiopoulos, "A framework of cooperative cell caching for the future mobile networks," in *Proceedings of the 48th Annual Hawaii International Conference on System Sciences, HICSS 2015*, pp. 5404–5413, USA, January 2015.

[32] Z. Chen and M. Kountouris, "D2D caching vs. small cell caching: Where to cache content in a wireless network?" in *Proceedings of the 17th IEEE International Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2016*, UK, July 2016.

[33] R. Wang, J. Zhang, S. H. Song, and K. B. Letaief, "Mobility-Aware Caching in D2D Networks," *IEEE Transactions on Wireless Communications*, no. 8, pp. 5001–5015, 2017.

[34] "Americans spend an alarming amount of time checking social media on their phones," Digital Trends, 2015: http://www.digitaltrends.com/mobile/informate-report-social-media-smartphone-use/.

[35] R. Fedorov, D. Frajberg, and P. Fraternali, "A framework for outdoor mobile augmented reality and its application to mountain peak detection," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 9768, pp. 281–301, 2016.

[36] M. Mirahsan, R. Schoenen, and H. Yanikomeroglu, "HetHetNets: Heterogeneous Traffic Distribution in Heterogeneous Wireless Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2252–2265, 2015.

[37] "Augmented reality hardware is still way ahead of its software," Engadget, 2016: https://www.engadget.com/2016/06/02/augmented-reality-hardware-and-software-still-needs-work/.

[38] "Rivvr brings wireless VR to the oculus rift and HTC vive," TechCrunch, 2016: https://techcrunch.com/2016/12/15/rivvr-brings-wireless-vr-to-the-oculus-rift-and-htc-vive/.

[39] A. Y. C. Nee, S. K. Ong, G. Chryssolouris, and D. Mourtzis, "Augmented reality applications in design and manufacturing," *CIRP Annals - Manufacturing Technology*, vol. 61, no. 2, pp. 657–679, 2012.

[40] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis, and A. Vakali, "Cloud computing: distributed Internet computing for IT and scientific research," *IEEE Internet Computing*, vol. 13, no. 5, pp. 10-11, 2009.

[41] R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges," *Future Generation Computer Systems*, 2016.

[42] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proceedings of the 1st ACM Mobile Cloud Computing Workshop, MCC 2012*, pp. 13–15, Finland, August 2012.

[43] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: Architecture, applications, and approaches," *Wireless Communications and Mobile Computing*, vol. 13, no. 18, pp. 1587–1611, 2013.

[44] K. Wang, K. Yang, and C. Magurawalage, "Joint Energy minimization and resource allocation in C-RAN with mobile cloud," *IEEE Transactions on Cloud Computing*, 2016.

[45] K. Wang, K. Yang, X. Wang, and C. S. Magurawalage, "Cost-effective resource allocation in C-RAN with mobile cloud," in *Proceedings of the IEEE International Conference on Communications, ICC 2016*, Malaysia, May 2016.

[46] P. Bahl, R. Y. Han, L. E. Li, and M. Satyanarayanan, "Advancing the state of mobile cloud computing," in *Proceedings of the 3rd ACM Workshop on Mobile Cloud Computing and Services (MCS)*, pp. 21–28, Lake District, UK, June 2012.

[47] S. Abolfazli, Z. Sanaei, E. Ahmed, A. Gani, and R. Buyya, "Cloud-based augmentation for mobile devices: motivation, taxonomies, and open challenges," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 337–368, 2014.

[48] K. Gai, M. Qiu, and H. Zhao, "Cost-aware multimedia data allocation for heterogeneous memory using genetic algorithm in cloud computing," *IEEE Transactions on Cloud Computing*, 2016.

[49] S. Andreev, O. Galinina, A. Pyattaev et al., "Exploring synergy between communications, caching, and computing in 5G-grade deployments," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 60–69, 2016.

[50] N. Golrezaei, A. Molisch, A. G. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 142–149, 2013.

[51] B. Zhou, Y. Cui, and M. Tao, "Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6284–6297, 2016.

[52] Y. Wang, A. V. Vasilakos, Q. Jin, and J. Ma, "Survey on mobile social networking in proximity (MSNP): Approaches, challenges and architecture," *Wireless Networks*, vol. 20, no. 6, pp. 1295–1311, 2014.

[53] J. Hosek, M. Ries, P. Vajsar et al., "User's happiness in numbers: Understanding mobile YouTube quality expectations," in *Proceedings of the 38th International Conference on Telecommunications and Signal Processing, TSP 2015*, pp. 607–611, Czech Republic, July 2015.

[54] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: the role of proactive caching in 5G wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, 2014.

[55] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: how location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 102–112, 2014.

[56] Y. Wang, L. Wei, A. V. Vasilakos, and Q. Jin, "Device-to-Device based mobile social networking in proximity (MSNP) on smartphones: Framework, challenges and prototype," *Future Generation Computer Systems*, 2015.

[57] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3665–3676, 2014.

[58] J. Kuula, P. Kettunen, V. Auvinen, S. Viitanen, O. Kauppinen, and T. Korhonen, "Smartphones as an alerting, command and control system for the preparedness groups and civilians: Results of preliminary tests with the Finnish police," in *Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2013*, pp. 42–51, May 2013.

[59] H. Wang, D. Peng, W. Wang, H. Sharif, H.-H. Chen, and A. Khoynezhad, "Resource-aware secure ECG healthcare monitoring through body sensor networks," *IEEE Wireless Communications Magazine*, vol. 17, no. 1, pp. 12–19, 2010.

[60] T. Ma, P. L. Shrestha, M. Hempel, D. Peng, H. Sharif, and H.-H. Chen, "Assurance of energy efficiency and data security for ECG transmission in BASNs," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 1041–1048, 2012.

[61] "Tampere traffic wardens work quicker thanks to Google Glass app," YLE UUTISET, 2016: http://yle.fi/uutiset/osasto/news/tampere_traffic_wardens_work_quicker_thanks_to_google_glass_app/8766025.

[62] "2017 police vehicles article," Government Fleet-Feature, 2016: http://www.government-fleet.com/article/story/2016/11/2017-police-vehicles.aspx.

[63] A. Orsino, D. Moltchanov, M. Gapeyenko et al., "Direct Connection on the Move: Characterization of User Mobility in Cellular-Assisted D2D Systems," *IEEE Vehicular Technology Magazine*, vol. 11, no. 3, pp. 38–48, 2016.

[64] B. V. Gnedenko, Y. K. Belyayev, and A. D. Solovyev, *Mathematical Methods of Reliability Theory*, Academic Press, 2014.