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 Koucheryavy1

1Tampere University of Technology, Tampere, Finland
2RUDN University, Moscow, Russia
3Federal 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

Copyright © 2017 Aleksandr Ometov et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 carryable and wearable equipment on a daily basis, which is a dramatic
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.
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
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 \( v(t) \), 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 \( x(t) \) \( \epsilon \mathbb{R}_+ \), which is the overall duration spent at time \( t \) for the failed element recovery. We obtain a two-dimensional process \( (v(t), x(t)) \) with an extended steady-state space \( E = \{(0), (1, x), (2, x), \ldots, (k, x)\} \).

We further denote \( p_i(t) \) 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 \( i \) \((i = 1, k)\). The time taken to “repair” the failed element is in the range \((x, x + dx)\). The probabilities in question are calculated as

\[
\begin{align*}
p_0 (t) & = P \{v(t) = 0\}, \\
p_1 (t, x) dx & = P \{v(t) = 1, x < x(t) < x + dx\}, \\
p_2 (t, x) dx & = P \{v(t) = 2, x < x(t) < x + dx\}, \\
& \vdots \\
p_k (t, x) dx & = P \{v(t) = k, x < x(t) < x + dx\}.
\end{align*}
\]

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

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

\[
P \{x \leq A < x + \Delta \mid A \geq x\} = \frac{P \{x \leq A < x + \Delta\}}{P \{A \geq x\}}
\]

\[
= \frac{A(x + \Delta) - A(x)}{1 - P \{A < x\}} = \frac{a(x) \Delta}{1 - A(x)} = \frac{a(x) \Delta}{e^{-ax}}
\]

\[
= a \Delta + o(\Delta).
\]

Second Event. Starting from the considered time instant, an element will be restored during the time \( \Delta \), provided that it has already been in “repair” for \( x \) time units.

\[
P \{x \leq B < x + \Delta \mid B \geq x\} = \frac{P \{x \leq B < x + \Delta\}}{P \{B \geq x\}}
\]

\[
= \frac{B(x + \Delta) - B(x)}{1 - P \{B < x\}} = \frac{b(x) \Delta}{1 - B(x)} = \delta(x) \Delta.
\]

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 \( p_i(t + \Delta) \) 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 \( \Delta \) (with probability \( 1 - \alpha \Delta \)) or (ii) if there was \( 1 \) failed element and it has been restored during time \( \Delta \) (with probability \( \delta(x) \Delta \)), while no other active element has failed during this time (with probability \( 1 - \alpha \Delta \)).

\[
p_0 (t + \Delta) = P \{v(t + \Delta) = 0\}
\]

\[
= p_0 (t) (1 - \alpha \Delta)
\]
Figure 3: Steady-state transitions in the considered system.

\begin{align*}
\int_{0}^{\alpha} p_{1}(t, x) \cdot \delta(x) \cdot \Delta \cdot (1 - \alpha \Delta) dx \\
= p_{0}(t) (1 - \alpha \Delta) \\
+ \int_{0}^{\alpha} p_{1}(t, x) \cdot \delta(x) \cdot \Delta dx.
\end{align*}

Hence, we establish the state-related equations as

\begin{align*}
p_{0}(t + \Delta) &= p_{0}(t) \cdot (1 - \alpha \Delta) + \int_{0}^{\alpha} p_{1}(t, x) \cdot \delta(x) \cdot \Delta dx, \quad (5)
\end{align*}

and, similarly,

\begin{align*}
p_{1}(t + \Delta, x + \Delta) dx & \quad = P \{ \nu(t + \Delta) = 1, x + \Delta < x(t + \Delta) < x + \Delta + dx \} \\
& \quad = p_{1}(t, x) dx \cdot (1 - \alpha \Delta) \cdot (1 - \delta(x) \cdot \Delta). \quad (6)
\end{align*}

After the necessary transformations, we obtain

\begin{align*}
p_{1}(t + \Delta, x + \Delta) &= p_{1}(t, x) \cdot (1 - \alpha \Delta) \cdot (1 - \delta(x) \cdot \Delta). \quad (7)
\end{align*}

At the next step, we have

\begin{align*}
p_{2}(t + \Delta, x + \Delta) dx & \quad = P \{ \nu(t + \Delta) = 2, x + \Delta < x(t + \Delta) < x + \Delta + dx \} \\
& \quad = p_{2}(t, x) dx \cdot (1 - \alpha \Delta) \cdot (1 - \delta(x) \cdot \Delta) \\
& \quad + p_{1}(t, x) dx \cdot \alpha \Delta. \quad (8)
\end{align*}

Further, we arrive at the equation

\begin{align*}
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. \quad (9)
\end{align*}

Similarly, we produce the equations for \( p_{i}(t + \Delta, x + \Delta), \)

\( i = 3, k - 1 \) as

\begin{align*}
p_{i}(t + \Delta, x + \Delta) &= p_{i}(t, x) \cdot (1 - \alpha \Delta) \cdot (1 - \delta(x) \cdot \Delta) \\
& \quad + p_{i-1}(t, x) \alpha \Delta, \quad i = 3, k - 1. \quad (10)
\end{align*}

At the next step, we establish

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

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

Then, we obtain for the general case the equation

\begin{align*}
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. \quad (12)
\end{align*}

The boundary conditions in our system are

\begin{align*}
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, \\
& \quad \vdots \\
p_{k-1}(t + \Delta, 0) &= \int_{0}^{t} p_{k}(t, x) \delta(x) dx. \quad (13)
\end{align*}

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

\begin{align*}
p_{0}(t + \Delta) &= p_{0}(t) \cdot (1 - \alpha \Delta) + \int_{0}^{t} p_{1}(t, x) \cdot \delta(x) \cdot \Delta dx, \\
p_{0}(t + \Delta) &= p_{0}(t) - p_{0}(t) \cdot \alpha \Delta + \int_{0}^{t} p_{1}(t, x) \cdot \delta(x) \cdot \Delta dx, \quad (14)
\end{align*}

\begin{align*}
p_{0}(t + \Delta) - p_{0}(t) &= -p_{0}(t) \cdot \alpha \Delta + \int_{0}^{t} p_{1}(t, x) \cdot \delta(x) \cdot \Delta dx, \\
\frac{1}{\Delta} \cdot (p_{0}(t + \Delta) - p_{0}(t)) &= -p_{0}(t) \cdot \alpha + \int_{0}^{t} p_{1}(t, x) \cdot \delta(x) dx.
\end{align*}
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

$$p_1(t + \Delta, x + \Delta) = p_1(t, x) \cdot (1 - \alpha \Delta) \cdot (1 - \delta(x) \cdot \Delta)$$

$$p_1(t + \Delta, x + \Delta) - p_1(t, x) \cdot \alpha \Delta = -p_1(t, x) \cdot \delta(x) \cdot \Delta,$$

$$p_1(t + \Delta, x + \Delta) - p_1(t, x) \cdot \alpha \Delta = -p_1(t, x) \cdot \delta(x) \cdot \Delta,$$

$$p_1(t + \Delta, x + \Delta) - p_1(t, x + \Delta) = -p_1(t, x) \cdot \delta(x) \cdot \Delta,$$

$$\frac{p_1(t + \Delta, x + \Delta) - p_1(t, x)}{\Delta} = -p_1(t, x) \cdot \delta(x). \quad (17)$$

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

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

Similarly, based on (9), we have

$$\frac{p_2(t + \Delta, x + \Delta) - p_2(t, x + \Delta)}{\Delta} + \frac{p_2(t, x + \Delta) - p_2(t, x)}{\Delta} = -p_2(t, x) \cdot (\alpha + \delta(x) \cdot \Delta). \quad (19)$$

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

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

Finally, it follows from (10) that

$$p_k(t + \Delta, x + \Delta) = p_k(t, x) \cdot (1 - \delta(x) \cdot \Delta)$$

$$+ p_{k-1}(t, x) \cdot \alpha \Delta,$$

$$p_k(t + \Delta, x + \Delta) = p_k(t, x) \cdot \delta(x) \cdot \Delta$$

$$+ p_{k-1}(t, x) \cdot \alpha \Delta,$$

$$p_k(t + \Delta, x + \Delta) - p_k(t, x + \Delta) + p_k(t, x + \Delta)$$

$$- p_k(t, x) = -p_k(t, x) \cdot \delta(x) + p_{k-1}(t, x) \cdot \alpha,$$

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

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

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

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

$$p_1(t, 0) = \alpha \cdot p_0(t) + \int_0^t p_2(t, x) \delta(x) \, dx,$$

$$p_2(t, 0) = \int_0^t p_3(t, x) \delta(x) \, dx,$$

$$\vdots$$

$$p_{k-1}(t, 0) = \int_0^t p_k(t, x) \delta(x) \, dx.$$  

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

$$p_0 = \lim_{t \to \infty} p_0(t),$$

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

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

$$\alpha p_0 = \int_0^\infty p_1(x) \cdot \delta(x) \, dx,$$

$$\frac{dp_1(x)}{dx} = -\alpha \cdot p_1(x),$$

$$\frac{dp_2(x)}{dx} = -\alpha \cdot p_1(x) + \alpha \cdot p_1(x),$$

$$\vdots$$

$$\frac{dp_{k-1}(x)}{dx} = -\alpha \cdot p_{k-1}(x) + \alpha \cdot p_{k-1}(x),$$

$$\frac{dp_k(x)}{dx} = -\alpha \cdot p_{k-1}(x) + \alpha \cdot p_{k-1}(x),$$

$$p_1(0) = \alpha \cdot p_0 + \int_0^\infty p_2(x) \cdot \delta(x) \, dx,$$

$$p_2(0) = \int_0^\infty p_3(x) \cdot \delta(x) \, dx,$$

$$\vdots$$

$$p_{k-1}(0) = \int_0^\infty p_k(x) \cdot \delta(x) \, dx.$$
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

\[
\alpha p_0 = \int_0^\infty p_1 (x) \cdot \delta (x) \, dx,
\]

\[
\frac{dp_1 (x)}{dx} = -\left( \alpha + \delta (x) \right) \cdot p_1 (x),
\]

\[
\frac{dp_2 (x)}{dx} = -\left( \alpha + \delta (x) \right) \cdot p_2 (x) + \alpha \cdot p_1 (x),
\]

\[
\frac{dp_3 (x)}{dx} = -\delta (x) \cdot p_3 (x) + \alpha \cdot p_2 (x),
\]

\[
p_1 (0) = \alpha \cdot p_0 + \int_0^\infty p_2 (x) \cdot \delta (x) \, dx,
\]

\[
p_2 (0) = \int_0^\infty p_3 (x) \cdot \delta (x) \, dx.
\]

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

\[
\frac{dp_1 (x)}{dx} = -\left( \alpha + \delta (x) \right) \cdot p_1 (x),
\]

\[
\frac{dp_3 (x)}{p_1 (x)} = -\alpha \cdot \delta (x) \, dx,
\]

\[
d \ln p_1 (x) = d \ln (1 - B (x)) + d \ln e^{-\alpha x},
\]

\[
d \ln p_2 (x) = d \ln (1 - B (x)) \cdot e^{-\alpha x},
\]

\[
\ln p_1 (x) = \ln (1 - B (x) \cdot e^{-\alpha x}) + \ln C_1,
\]

\[
\ln p_2 (x) = \ln (1 - B (x) \cdot e^{-\alpha x} \cdot C_1),
\]

\[
p_1 (x) = C_1 \cdot e^{-\alpha x} \cdot (1 - B (x)).
\]

(27)

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

\[
p_2 (x) = (\alpha C_1 x + C_2) \cdot e^{-\alpha x} \cdot (1 - B (x)) \cdot \frac{1 - b (\alpha)}{b (\alpha)},
\]

\[
p_3 (x) = C_3 \cdot (1 - B (x)) - (C_2 + C_1 (\alpha x + 1)) \cdot e^{-\alpha x} \cdot (1 - B (x)).
\]

(28)

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

\[
p_0 = C_1 \cdot \alpha^{-1} \cdot \bar{b} (\alpha),
\]

\[
p_1 (x) = C_1 \cdot e^{-\alpha x} \cdot (1 - B (x)),
\]

\[
p_2 (x) = C_1 \cdot e^{-\alpha x} \cdot (1 - B (x)) \cdot \frac{1 - \bar{b} (\alpha)}{b (\alpha)},
\]

\[
p_3 (x) = C_1 \cdot (1 - e^{-\alpha x}) \cdot (1 - B (x)) \cdot \frac{1 - \bar{b} (\alpha)}{b (\alpha)}.
\]

(29)

Further, we evaluate the stationary probabilities for the macrostate \( p_1 \).

\[
p_1 = \int_0^\infty p_1 (x) \, dx = \int_0^\infty C_1 \cdot e^{-\alpha x} \cdot (1 - B (x)) \, dx
\]

\[
= C_1 \cdot \left( \int_0^\infty e^{-\alpha x} \, dx - \int_0^\infty e^{-\alpha x} \cdot B (x) \, dx \right)
\]

\[
= C_1 \cdot \left( \frac{1 - \bar{b} (\alpha)}{\alpha} \right),
\]

(30)

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

Next, for the probability \( p_2 \) we have

\[
p_2 = \int_0^\infty p_2 (x) \, dx
\]

\[
= \int_0^\infty C_1 \cdot e^{-\alpha x} \cdot (1 - B (x)) \cdot \frac{1 - \bar{b} (\alpha)}{b (\alpha)} \, dx
\]

\[
= C_1 \cdot \left( \frac{1 - \bar{b} (\alpha)}{\alpha} \right).
\]
Further, for $p_3$ we obtain

$$p_3 = \int_0^\infty p_3(x) \, dx = \int_0^\infty C_1 \cdot (1 - e^{-\alpha x}) \cdot (1 - B(x)) \, dx$$

$$= C_1 \cdot \frac{(1 - \tilde{b}(\alpha))}{\tilde{b}(\alpha)} \cdot \int_0^\infty (1 - \tilde{b}(\alpha)) \cdot \left(1 - \frac{1}{\alpha} \cdot \frac{x - b(x)}{1 - \tilde{b}(\alpha)} \right) \, dx$$

$$= C_1 \cdot \frac{(1 - \tilde{b}(\alpha))^2}{\alpha \cdot \tilde{b}(\alpha)}.$$

(31)

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

$$C_1 = \frac{\alpha \cdot \tilde{b}(\alpha)}{\rho^{-1} \left(1 - \frac{1}{\alpha} \cdot \frac{1}{\tilde{b}(\alpha)} \right)} = \frac{\alpha \cdot \tilde{b}(\alpha)}{\rho^{-1} \left(1 - \frac{1}{\alpha} \cdot \frac{1}{\tilde{b}(\alpha)} \right)}.$$

(34)

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

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 $\rho$ for four different special cases of the $\langle M_3 | GI | 1 \rangle$ model, where GW stands for Gnedenko-Weibull distribution [64].

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.

\[ \begin{align*}
p_0 &= C_1 \cdot \frac{-\tilde{b}(\alpha)}{\alpha}, \\
p_1 &= C_1 \cdot \frac{(1 - \tilde{b}(\alpha))}{\alpha}, \\
p_2 &= C_1 \cdot \frac{(1 - \tilde{b}(\alpha))^2}{\alpha \cdot \tilde{b}(\alpha)},
\end{align*} \]

\[ \begin{align*}
p_3 &= C_1 \cdot \frac{(1 - \tilde{b}(\alpha))}{\tilde{b}(\alpha)} \cdot \left(1 - \frac{\hat{b}(\alpha)}{\alpha} \right).
\end{align*} \]

(33)

where $\hat{b}(\alpha) = b \cdot \alpha$.

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.
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 $\rho$ (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 $\rho$, 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
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 \( t \)
- \( \Delta \): Short time interval
- \( A \): Time to failure (random variable)
- \( \alpha \): 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


