Towards a 5G Communication Architecture for the Internet of Musical Things

Marco Centenaro
Fondazione Bruno Kessler
Trento, Italy
mcentenaro@fbk.eu

Paolo Casari
University of Trento
Trento, Italy
paolo.casari@unitn.it

Luca Turchet
University of Trento
Trento, Italy
luca.turchet@unitn.it

***Abstract*—The stringent reliability and latency requirements of the Internet of Musical Things (IoMusT) paradigm are outside the reach of the current generation of cellular networks, thus significant expectations fall on the upcoming fifth-generation (5G) of cellular systems. In this paper, we outline for the first time a communication architecture for a future IoMusT based on the 5G framework, introducing two prominent use cases and designing a model of such a system. Moreover, considering the realistic needs of smart musical tools, we also define high-level service requirements and sketch some relevant evaluation scenarios for future investigation.**

I. Introduction

The Internet of Things (IoT) paradigm finds application in multiple fields, spanning environmental monitoring, industrial control, and many others [1]. Recently, there has been a growing interest towards the application of the IoT to the domain of musical interactions over networks, leading to the emergence of the Internet of Musical Things (IoMusT) [2] field. The IoMusT vision requires the design of a new class of IoT devices (the so-called *musical things*) and communication systems dedicated to musical purposes. For instance, it is envisioned that musical instruments will equip with a communication interface, making them smart musical instruments (SMIs) [3]. Moreover, the people attending a performance will utilize innovative multisensory interfaces both to enrich the music listening experience, and to participate actively in the music creation process [4], [5].

The IoMusT may revolutionize the traditional concept of musical interaction in many ways. A prominent example is the so-called networked music performance (NMP), whose widespread availability could allow musicians to interact and perform together without being co-located, thanks to SMIs. Both wired and wireless networks can connect SMIs and support IoMusT traffic. However, using cables can be problematic in certain cases, such as when assembling/disassembling musical devices on stage [6], [7]. Instead, wireless network interfaces provide better opportunities, especially a much smoother user experience in terms of instrument (self)-configuration, and freedom of movement for the musicians. On the other hand, an effective remote and distributed music performance entails extremely strict quality of service (QoS) requirements, such as very low communication latency, low and constant jitter (i.e., the variation of latency), and high audio quality (i.e., low packet losses that generate unperceivable signal dropouts) [8]. Satisfying these key performance indicators (KPIs) makes it possible for the performers to maintain a stable tempo, to remain synchronous and, more generally, to benefit from a high-quality interaction experience [9, Ch. 3].

Wireless cellular technologies are prominent candidates to connect SMIs. In particular, we expect the fifth-generation (5G) of cellular systems to be a fundamental enabler of the IoMusT paradigm, because it will overcome the packet latency and reliability limitations of the current fourth-generation (4G) systems. Nevertheless, the absence of a dedicated infrastructure for the IoMusT calls for a careful configuration of the 5G system. For this reason, in this paper, we aim at laying the foundation of a 5G-enabled IoMusT framework by introducing building blocks, technological enablers, and design principles for an effective network configuration, in order to support IoMusT users as much as possible. We also highlight the role of edge computing in conjunction with the 5G deployment.

The remainder of the paper is organized as follows: Section II surveys the related work in the scientific and standardization literature, providing the foundation of our proposed architecture; Section III presents an overview of such architecture, describing the various components, and characterizing its KPIs; a discussion on the identified gaps and future research directions follows in Section IV. Finally, we draw some concluding remarks in Section V.

II. Related Work

A. Technological Enablers of Networked Music Performances

Gabrielli and Squartini [9, Ch. 4], survey candidate wireless communication standards for NMP. These technologies encompass both short-range systems (such as proprietary audio-specific solutions and IEEE 802.11a/b/g/n/ac) and long-range technologies (such as IEEE 802.11af). Short-range technologies operate mainly in the 2.4 GHz industrial, scientific, and medical (ISM) frequency band. While proprietary solutions are not suitable for NMP because they usually provide unidirectional signal transmission only, the short-range standards of the IEEE 802.11 family suffer from the typically high interference level found in ISM frequency bands, from a medium access latency that tends to increase significantly with the size of the network, as well as from distributed channel access management issues such as the well-known hidden node problem. On the other hand, the IEEE 802.11af operates on sub-GHz frequency bands previously used for television broadcast, which show favorable propagation properties. Even in this case, the uneven availability of sub-GHz bands across different

ISSN 2305-7254

countries jeopardizes the exploitation of such technology for NMP.

As a consequence of the above limitations, the vast majority of the existing frameworks for NMP relies on *wired* network infrastructures [10, Tab. 3]. One of the most advanced frameworks is LOLA ("LOw LAtency audio visual streaming system"), a system specifically conceived for distributed performing arts interaction over advanced packet networks [11], [12]. LOLA leverages a dedicated wired network infrastructure and specialized hardware, limiting its scalability and cost-effectiveness.

In this context, the potential of wireless cellular systems as connectivity providers for the IoMusT is yet unexplored. Cellular radio access technologies could provide end-to-end, IP-based connectivity between SMIs (which would constitute a brand-new class of user equipments (UEs)) without the limitations of IEEE 802.11af, and implying a minimum configuration effort for the user via the plug&play concept. Moreover, while current 4G cellular systems focus on supporting mobile broadband traffic, the 5G of such systems will be able to support ultra-reliable low-latency communication (URLLC), i.e., guarantee packet latency values down to 1 ms and a reliability of 99.999% [8], [13]. For these reasons, new industrial initiatives have started focusing on networked music performance over wireless cellular networks. An interesting case study discussed throughout this paper is that of Elk, a Swedish company that develops technologies enabling a new generation of connected musical instruments and audio processors [14].

B. Embedded Audio Systems

Recent music technology advances led to the emergence of embedded platforms dedicated to digital audio processing that are suitable for creating musical things as well as to build IoMusT applications on top of them. A notable example is represented by the Bela board [15], a cape for the Beaglebone Black that enables low-latency audio signal processing.

Nevertheless, most embedded systems designed for audio processing offer a limited range of connectivity options. A notable exception is Elk's Audio OS, an embedded Linux-based operating system, that guarantees processing latency below 1 ms. It is highly optimized not only for low-latency and high-performance audio processing, but also for handling wireless connectivity to local and remote networks using the most widespread communication protocols.

C. IoMusT Ecosystems

Recent endeavors in IoMusT research explored the creation of ecosystems around IoMusT technologies, proposing preliminary architectures based on Semantic Web technologies to foster interoperability across heterogeneous musical things. The semantically-enriched IoMusT architecture reported in [16] relies on a semantic audio server, embedded audio systems, and edge computing techniques. In particular, the SPARQL Event Processing Architecture described in [17] was used as an interoperability enabler allowing multiple prototypes of musical things to cooperate. However, Semantic Web technologies are not suitable for IoMusT applications relying on real-time aspects, as the Semantic Web stack is oriented towards static scenarios, where information evolves at a low rate [17]. To cope with this issue, Viola *et al.* improved the architecture reported in [16] by using the Constrained Application Protocol (CoAP) [18], a lightweight IoT protocol for machine-to-machine communication [19].

Nevertheless, neither of the above architectures has been tested yet within actual IoMusT ecosystems as yet. In addition, they have been developed around wireless local area network technologies (i.e., Wi-Fi), whereas their application in distributed musical performance contexts (such as those envisioned for 5G systems) represents today an unexplored opportunity.

D. Standardization Efforts

The standardization body of cellular networks – the Third Generation Partnership Project (3GPP) – investigated the potential of the 5G-enabled distribution of audio-visual content and services in [20]. The report describes a variety of use cases, several of which deal with high-quality audio acquisition, mixing, and dispatching. The most relevant use case in the scope of the IoMusT paradigm is *audio streaming in live performances* [20, §5.2], dealing with live stage events such as concerts, musicals, or theatre events where several artists perform in front of an audience. Here, UE-type microphones generate multiple audio streams; after audio mixing, these streams return to the musicians via UE-type in-ear monitoring devices thanks to a 5G infrastructure.

The typical system parameters for audio streaming in live performances use cases are provided in [20, Tab. 5.2.1-1] in terms of various KPIs. Moreover, the European Telecommunications Standards Institute (ETSI) has recently established a new industry specification group on augmented reality, whose purpose is to define a framework for the inter-operability of augmented reality components, systems and services [21]. We envision that such an initiative can foster the entire extended reality umbrella, including virtual reality, with the creation of new kinds of musical experiences based, e.g., on virtual avatars or virtual objects [2], [5], which are enabled by the interworking between smart musical things and augmented/virtual reality devices.

III. Proposed Framework for a 5G-Enabled Internet of Musical Things (IoMusT)

We illustrate our envisioned communication architecture for a 5G-enabled IoMusT in Fig. 1. Such architecture shall include the following elements:

- *UE-type musical things*, that include audio I/O hardware, an audio processing system (e.g., the Elk Audio OS), and a 5G communication module;
- a *5GS* providing low-latency packet delivery and extremely reliable packet transmission. The system combines a next-generation radio access network (NG-RAN) comprising 5G base stations – the so-called next-generation NodeBs (gNBs) – and a 5G Core (5GC) network, which transfer (digital) audio traffic among musical things, possibly with the mediation of audio application services providing, e.g., stream processing and content caching;

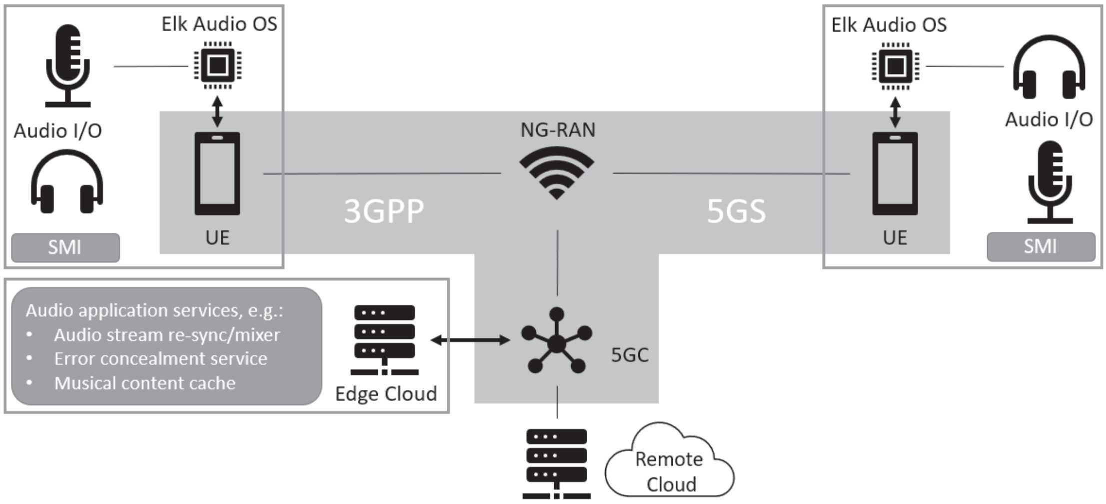

Fig. 1. 5G-based IoMusT communication architecture overview, entailing UE-type SMIs, 3GPP 5GS, and cloud computing platforms (remote and edge)

- *cloud computing platforms*, which host such application services. These platforms may be located either in the remote cloud (central datacenter) if they perform latency-tolerant tasks, or at the edge of the network if they perform latency-critical tasks. In this paper, we will mainly consider edge cloud computing platforms, due to the stringent service requirements of the IoMusT.

A. 5G Enablers

Within the scope of 3GPP, the 5GS [22] represents the successor of 4G Evolved Packet System (EPS). The 5GS introduces multiple innovations on both the radio access network (RAN) and core network (CN).

As far as the access network is concerned, a new air interface called New Radio (NR) [23] is key to overcoming the limitations of the legacy Long-Term Evolution (LTE) standard. Despite being still based on orthogonal frequency division multiple access (OFDMA) like LTE, the NR standard is characterized by a redesigned and flexible structure both in the time domain and in the frequency domain, thanks to the introduction of short transmission time interval (TTI) and of additional subcarrier spacing (SCS) options. In conjunction with reduced processing times at both the UE and the gNB, grant-free transmissions [24], antenna diversity, and multi-connectivity [25], the 5G NR can meet the latency-reliability constraints posed by the URLLC paradigm.

Regarding the CN, the 5GC introduced a disruptive service-based architecture, in which the network elements and operations are organized into network functions (both physical and virtual), that produce and consume one another's *services*. Such a design extensively softwarizes the CN infrastructure, allowing flexible network deployment and (re)configuration. One example is the seamless integration of multi-access edge computing (MEC) platforms into the 5GC administrative domain: being considered as an application function, the MEC host can interact with the CN to negotiate traffic routing policies and exploit value-added services provided by the 5GS [26], [27]. In such a time-critical application as the IoMusT, the role of the MEC is as important as that of the RAN and of the transport network. The MEC is a convenient candidate site to mix digital audio streams, as well as host other smart, machine learning-based processing functions that, e.g., automatically reconstruct flows by filling gaps originating from bursts of transmission errors and unrecoverable packet losses.

One of the strength of the 5GS lies in its support for heterogeneous traffic types, each having different QoS requirements. In particular, other than URLLC, a 5GS shall be able to support simultaneously enhanced mobile broadband (eMBB) traffic as well as massive machine-type communication (mMTC). The *network slicing* concept guarantees the management of the distinct traffic types: here, the (unique) physical network infrastructure can be divided into multiple virtual networks serving a given service type [23, §16.3]. We also recall that a smooth transition between the two generations of 3GPP technologies will happen, starting from the new gNB that will operate with a 4G Evolved Packet Core (EPC) in the so-called non-standalone (NSA) mode.

B. Envisioned Use Cases

We envision two 5G-enabled IoMusT use cases:

1) *ultra-reliable low-latency networked music performance (URLL-NMP)*, entailing IoMusT users rehearsing and playing together from remote places thanks to UE-type SMIs. This use case extends the 3GPP's *audio streaming in live performances* use case [20, §5.2] by relaxing the constraint of having co-located musicians and audience, thus yielding 5G-enabled networked music performances. We sketch this use case and a sample communication protocol stack in Figs. 2 and 3, respectively. In essence, the typical data flow among the IoMusT users involves sending audio streams through the 5GS, which conveys them to a MEC server running audio application services (such as audio re-synchronization/mixing or advanced functions for error concealing). The MEC server finally returns a mixed audio stream to each musician

within a maximum tolerable delay that makes it possible to maintain a smooth musical performance. Let us remark that the MEC may not be always present or needed: in those cases, the 5GS just routes the audio traffic between the involved IoMusT users. In any event, we envision that the MEC assistance becomes more and more important as the amount of involved streams and the distance between IoMusT users grows;

2) *fast server interaction*, entailing an enhanced one-to-one interaction between budding/expert musicians and a server. For example, by streaming the played music towards the server through a SMI, each musician may automatically trigger a software agent instantiated in the MEC, which can respond in real-time for improvisation, composition, or learning purposes [28]. This use case also includes new kinds of human-machine interaction entailing, e.g., the expressiveness of generated music [29].

C. Key Performance Indicators

1) Latency: We can characterize the end-to-end (E2E) latency of the audio communication from the time the (analog) audio signal leaves the generating SMI until its delivery to the musician as follows:

$$\begin{aligned}\mathcal{D} = \tau_{\text{audio,upstream}} &+ \tau_{\text{tx,uplink}} \\ &+ \tau_{\text{transport}} + \tau_{\text{proc}} \\ &+ \tau_{\text{tx,downlink}} + \tau_{\text{audio,downstream}} ,\end{aligned} \tag{1}$$

where

- $\tau_{\text{audio,upstream}}$ is the time that a musical thing takes to acquire and digitize an analog audio signal in a suitable manner before passing it to the transmission module. Note that the outgoing traffic pattern (the so-called *transfer interval* [20]) towards the 5G module is deterministic;
- $\tau_{\text{tx,uplink}}$ includes UE processing delay, transmission time, and processing time at the gNB side;
- $\tau_{\text{transport}}$ is the delay component caused by the transmission of the audio stream from the gNB that serves the transmitting UE towards the gNB serving the receiving UE. This interval includes the time to reach a server that may process the data before forwarding them to the destination;
- τ_{proc} is the time taken by the MEC server (if present/needed) to process the incoming audio stream, e.g., to re-synchronize/mix it with streams from other SMIs or to retrieve cached audio content;
- $\tau_{\text{tx,downlink}}$ is the counterpart of $\tau_{\text{tx,uplink}}$, and includes the gNB processing delay, the transmission time, and the UE processing delay. Note that, due to the different direction of the transmission (downlink vs. uplink), we likely have that $\tau_{\text{tx,downlink}} \neq \tau_{\text{tx,uplink}}$;
- $\tau_{\text{audio,downstream}}$ accounts for the operations that the receiving musical thing must perform in order to serve an analog audio signal to the musician. This delay depends on whether a MEC server is used to re-sync and mix audio streams or not. In the absence of the MEC server, each receiving SMI would manage packet reordering, audio syncing and error concealment independently, therefore typically $\tau_{\text{audio,downstream}} \neq \tau_{\text{audio,upstream}}$.

2) Reliability: This KPI measures how long a given system performs its intended function under well-defined conditions. This concept is coupled with that of *availability*, which instead is a measure of the percentage of time the system is in an operable state. We stress that a reliable system has also a high availability, while a highly available system may not be reliable. Thus, in the following, we refer to the concept of reliability as a KPI.

As far as network-layer packet transmissions are concerned, the reliability is typically defined as the percentage of the amount of sent network layer packets which reach another system entity within the time constraint required by the targeted service, divided by the total number of sent network layer packets [30, §3.1]. Similarly to what we did with $\mathcal{D}$, we propose to decompose the E2E reliability into multiple reliability components as follows:

$$\mathcal{R} = p_{\text{succ,uplink}} \cdot p_{\text{succ,transport}} \cdot p_{\text{succ,comp}} \cdot p_{\text{succ,downlink}} , \tag{2}$$

where

- $p_{\text{succ,uplink}}$ is the success probability of the uplink transmission;
- $p_{\text{succ,transport}}$ is the success probability of the packet forwarding across the back-haul transport network;
- $p_{\text{succ,comp}}$ represents the probability of error-free data processing at the computing platform side;
- $p_{\text{succ,downlink}}$ is the success probability of the downlink transmission.

Note that we neglect the non-idealities of the audio processing unit. Moreover, it is reasonable to assume that $p_{\text{succ,transport}} \to 1$ due to an almost error-free backbone transport network. On the other hand, we may assume $p_{\text{succ,comp}} \to 1$ only for a remote/datacenter cloud deployment, but it may not be the case for edge deployments. Thus, the reliability of a 5G-enabled IoMusT can be approximated as follows:

$$\mathcal{R} \simeq p_{\text{succ,uplink}} \cdot p_{\text{succ,comp}} \cdot p_{\text{succ,downlink}} . \tag{3}$$

In other words, we can safely assume that the overall reliability is due to the reliability given by the radio links in uplink and downlink and the dependability of the edge computing platform.

3) Service Coverage: In 3GPP's technical documents, the term "service area" commonly refers to a geographic region where a 3GPP communication service is accessible [30]. For 5G-enabled IoMusT, the concept of service area more closely concerns the topology and transport delay of the network that connects the performers, rather than their geographic distance. For this reason, in the following we will denote the accessibility of IoMusT services as "service coverage," and we will characterize this KPI by identifying three relevant scenarios.

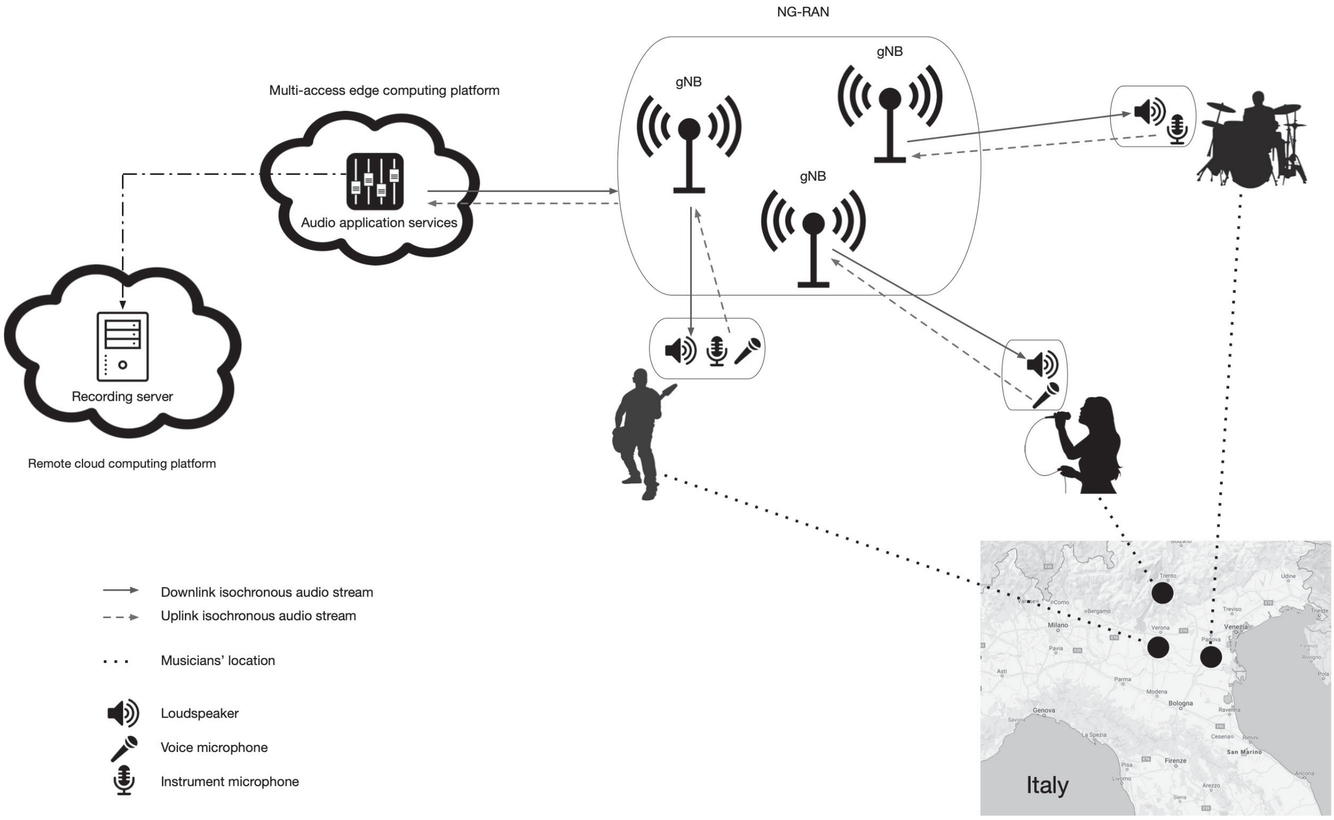

Fig. 2. Illustration of the envisioned URLL-NMP use case. In this sample scenario, three players are distributed in a regional area of Northern Italy, and exploit UE-type microphones and loudspeakers to perform together. The 5GS infrastructure is in charge of managing the generated audio flows (singing voices and musical instruments signals). A MEC platform in proximity of the NG-RAN may provide audio stream processing, e.g., providing each artist with a personalized audio mix through the loudspeakers. A remote cloud platform may also carry out latency-tolerant tasks, e.g., recording and storing the musical session.

First, let us consider by a rough estimation of the audio communication time. Considering the Elk audio systems [31], the board produces a protocol data unit (PDU) comprising 32 audio samples (each of 32 bits) for each audio channel, plus some overhead (about 256 bits including UDP headers). Since two audio channels are considered, the total PDU size is ≈290 bytes. Redundancy mechanisms help compensate for packet losses: after the transmitter sends a packet, the following packet includes 20% redundant information. Being the sampling frequency equal to 48 kHz and considering the redundancy, the packet transmission rate is approximately one packet every $32/(48 \cdot 10^3) \cdot (4/5) \approx 0.533$ ms. This is the value of $\tau_{\text{audio,upstream}}$. Hence, with a fast audio board in each SMI [31], we may assume that $\tau_{\text{audio,upstream}} \ll \mathcal{D}$. We may also assume that $\tau_{\text{audio,downstream}}$ has the same order of magnitude as $\tau_{\text{audio,upstream}}$. Therefore, the main delay components are due to the over-the-air transmissions, back-haul routing, and processing. We recall that we assume that the reliability of the audio processor is 100%.

While the operations of an SMI on an audio streams are inherently local tasks, the transmission of audio samples over the air requires the multiplexing of traffic from multiple users, which places the burden on the RAN and transport networks. In particular, multiple IoMusT users co-located in the same geographical area could potentially lead the 5GS to worse latency ($\tau_{\text{tx,uplink}}$ and $\tau_{\text{tx,downlink}}$) and reliability performance ($p_{\text{succ,uplink}}$ and $p_{\text{succ,downlink}}$). An increasing load would have consequences also on the computing platform, as well. In both the envisioned use cases, an edge computing platform may be instrumental to achieve low-latency IoMusT user interaction. With reference to Fig. 2, it is reasonable to assume that each MEC host will serve a certain geographical area, thus if the amount of IoMusT users in that area increases, the computing load will increase as well. An increasing offered load leads to a higher latency component $\tau_{\text{transport}}$, but may also impact the reliability term $p_{\text{succ,comp}}$.

Based on the above observations, a limited number of IoMusT users would result in $\tau_{\text{tx,uplink}}$, τ_{proc}, $\tau_{\text{tx,downlink}}$, $\tau_{\text{audio,downstream}} \to 0$, and $p_{\text{succ,proc}} \to 1$. As a result, $\mathcal{D} \simeq \tau_{\text{transport}}$, thus almost the entire latency budget may be employed to cover the round-trip time between each user and the serving cloud computing platform, which performs audio processing (e.g., syncing, mixing, error concealing, caching). In this respect, we remark that the geographical distance between each IoMusT user and the server depends on the back-haul network topology, thus the definition of *proximity* among performers relates to *traffic routing* delay, rather to geographical distance [32]. Thus, in the absence of *closely or fully integrated* administrative network domains, a given use case may be feasible or not.

According to the above reasoning, we can define the IoMusT service coverage based on the following two parameters:

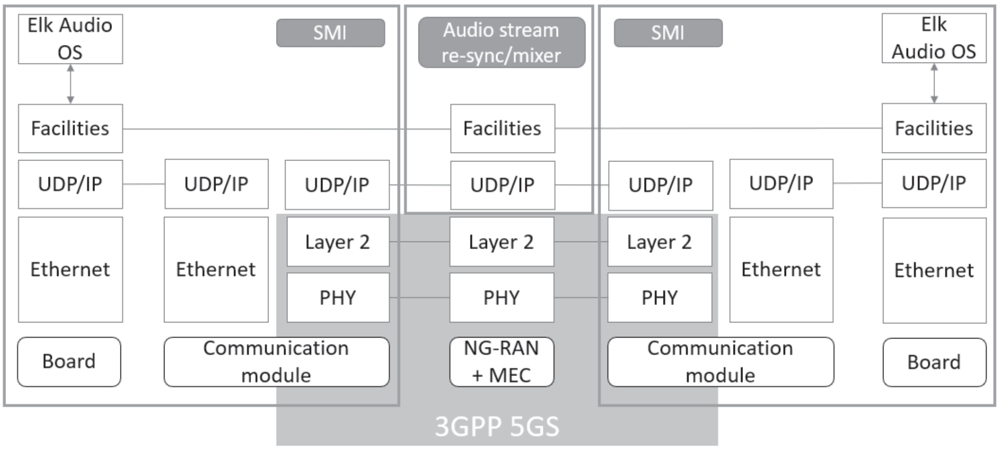

Fig. 3. Characterization of the IoMusT communication protocol stack for the URLL-NMP use case. Two SMIs communicate with the MEC server, which synchronizes and mixes the respective audio streams. Note that each SMI integrates an audio processing board as well as a communication module that interfaces the board with the 5G network.

- number of IoMusT users served by the same MEC platform, i.e., the MEC load;
- number of IoMusT users served by the same RAN infrastructures, i.e., the RAN load.

Specifically, we identify three different scenarios according to the interplay of these parameters.

Scenario 1: high-density performer distribution — Multiple IoMusT users connect locally to the same RAN, i.e., to the same gNB or to a set of gNBs that relies on the same MEC server. This scenario tests the stress on the RAN in moderate transport delay conditions, which offers some leeway to absorb part of the delay budget in radio links, e.g., to correct transmission errors and ensure a smooth audio flow.

Scenario 2: low-density, wide-area performer distribution — Multiple IoMusT users are distributed over a wide area. Hence, they connect to different local gNBs, which forward the audio flows to a MEC server chosen to optimize the transport performance. This scenario tests the stress on the transport network, assuming that it consumes most of the delay budget to connect performers located farther than in Scenario 1.

Scenario 3: intermediate-density, wide-area performer distribution — IoMusT users appear in clusters: each cluster is mostly distributed locally, but the clusters are distributed over an intermediate/wide area. This scenario makes it possible to explore the interplay among the RAN, the transport, and the MEC server delay components.

D. High-Level Service Requirements

According to several studies, rhythmic synchronization among multiple performers is optimal as long as the acoustic delay remains below 20-30 ms. Skilled musicians may even tolerate an absolute maximum of 50 ms of delay, without affecting the music performance [10]. For the URLL-NMP use case, we may therefore set the total delay budget to an intermediate value of $\mathcal{D} = 30$ ms, Instead, the fast server interaction use cases comprises several sub-use cases, each of which entails a different latency requirement. For example, in case the edge server underpins an interactive IoMusT performance, the latency requirement should be similar to that of URLL-NMP, whereas if the server's role is to proactively cache musical content, higher latency budgets are typically tolerable.

As far as the packet reliability $\mathcal{R}$ is concerned, air interface reliability lower bounds between $1 - 10^{-3}$ and $1 - 10^{-9}$ typically satisfy the so-called *tactile Internet* paradigm [33], of which the IoMusT is part. However, such service guarantees typically hold only under well-defined conditions and depend on many factors, e.g., the varying offered traffic load. In any case, as also seen for the latency, the upper bounds on the reliability value should be a function of the audio application. In particular, we may consider as a reference the audio streaming in live performances use case [20, Tab. 5.2.1-1], which mandates a packet error ratio $< 10^{-6}$ for a packet size corresponding to 1 ms of audio data, in order to guarantee no audio dropouts or audible interference occurrences, and assuming no/basic error concealment algorithms. Such a constraint should apply to both the envisioned use cases.

Regarding the service coverage, the envisioned fast server interaction scenario features a one-to-one interaction between IoMusT and the cloud computing platform, thus entailing *traditional* challenges that fall under Scenario 3. Examples of these challenges are, e.g., the design of load balancing between remote cloud and edge cloud computing platforms, as well as the MEC host selection. On the other hand, URLL-NMP requires a more careful analysis, as it requires a one-to-many audio distribution mediated by the MEC server. As a matter of fact, according to the geographical position of the various members of the musical band as well as the density of other IoMusT users in the same area, multiple of our identified scenarios may occur simultaneously. In particular, for Scenario 1, we expect that the perceived user experience largely depends on the density of the performers and of the gNBs, as well as on the load of the MEC server that mixes the incoming audio streams or provides the cached musical content. For Scenario 2, the perceived user experience mostly depends on the average distance of the users (measured in

TABLE I. 5G-ENABLED IOMUST USE CASES AND RELATED E2E SERVICE REQUIREMENTS.

USE CASE	LATENCY	RELIABILITY	SERVICE COVERAGE
URLL-NMP	≤ 30 ms	$\leq 10^{-6}$	Scenario 1, 2, or 3
Fast server interaction	Sub-use-case-dependent		Scenario 3

terms of transport delay rather than geographical distance). Scenario 3, as an intermediate case between the previous two, involves a careful management of the various components of a 5G-enabled NMP, ideally enabling users located in another region or even another country to play together.

The mapping of the requirements against the identified use cases are provided in Table I.

IV. DISCUSSION: IDENTIFIED GAPS AND OPEN RESEARCH QUESTIONS

Various aspects of the proposed 5G communication architecture for IoMusT deserve more investigation.

- *Availability of 5G-enabled musical things* — At the time of writing, as also reflected in Fig. 3, state-of-the-art audio processing boards (e.g., Elk Audio OS) are not yet tightly integrated with the 5G communication module.
- *Availability of a 5GS* — Full-fledged deployments of standalone 5GS will take some time, thus the QoS provisioning for the envisioned IoMusT use cases may not be available soon, even though integrated 5G-audio systems may progressively become available on the market.
- *Characterization of the service coverage* — The IoMusT service coverage strongly depends on the considered scenario and network deployment conditions, including the coverage of the 5GS and the location and instantaneous load of MEC servers. Therefore, the evaluation of service coverage is an interesting open research question.
- *Optimal load balancing of edge tasks* — Each use case leads to a different optimization of the location of the MEC server that provides the audio services: for a distributed music performance, for example, the choice of the MEC server should minimize the data transport latency and offload the tasks of the on-instrument audio processing board; for the fast server interaction use case, instead, only one performer interacts with the server, and the best placement is likely closest to the performer. Anyway, all of these aspects deserve attention and research regarding, e.g., the optimization of the placement and the load of the involved MEC hosts.
- *Enhancement of network awareness and intelligence* — Innovative networking strategies to successfully manage parallel IoMusT sessions as well as individual sessions are needed to ensure the scalability of the 5G-enabled IoMusT. Such strategies may make use of artificial intelligence, e.g., to conceal errors introduced by the connect-compute chain.

V. CONCLUSION

In this paper, we characterized for the first time a 5G-based IoMusT framework, that can contribute to spread the availability of innovative musical services. Thanks to the technologies belonging to the 5G ecosystem, new use cases such as URLL-NMP and fast server interaction may be made available to musicians. We identified the KPIs to be considered for the evaluation of solutions that implement these use cases, and proposed a general communication architecture. Finally, we mentioned the most recent advancements in terms of hardware components, that could make the vision of 5G-enabled IoMusT real.

As future work, we will evaluate the interplay between the identified KPIs by means of analytical tools as well as computer simulations, in order to define and characterize the domain of application of the vision described in this paper.

ACKNOWLEDGMENT

The authors want to thank Stefano Zambon of Elk for his kind feedback on the Elk Audio OS.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: a survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010. [Online]. Available: https://doi.org/10.1016/j.comnet.2010.05.010

[2] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of Musical Things: Vision and Challenges," *IEEE Access*, vol. 6, pp. 61 994–62 017, 2018. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2872625

[3] L. Turchet, "Smart Musical Instruments: vision, design principles, and future directions," *IEEE Access*, vol. 7, pp. 8944–8963, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2876891

[4] L. Turchet, T. West, and W. M. M., "Touching the audience: Musical Haptic Wearables for augmented and participatory live music performances," *Journal of Personal and Ubiquitous Computing*, pp. 1–21, 2020. [Online]. Available: https://doi.org/10.1007/s00779-020-01395-2

[5] L. Turchet, M. Benincaso, and C. Fischione, "Examples of use cases with smart instruments," in *Proceedings of Audio Mostly Conference*, 2017, pp. 47:1–47:5. [Online]. Available: https://doi.org/10.1145/3123514.3123553

[6] S. J. Barro, T. M. Fernández-Caramés, and C. J. Escudero, "Enabling collaborative musical activities through wireless sensor networks," *Int. J. of Distributed Sensor Networks*, vol. 8, no. 3, p. 314078, 2012.

[7] J. Martinez-Avila, C. Greenhalgh, A. Hazzard, S. Benford, and A. Chamberlain, "Encumbered interaction: a study of musicians preparing to perform," in *Proceedings of the Conference on Human Factors in Computing Systems*. ACM, 2019.

[8] X. Jiang, H. Shokri-Ghadikolaei, G. Fodor, E. Modiano, Z. Pang, M. Zorzi, and C. Fischione, "Low-latency networking: Where latency lurks and how to tame it," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 280–306, 2018.

[9] L. Gabrielli and S. Squartini, *Wireless Networked Music Performance*. Springer, 2016. [Online]. Available: https://doi.org/10.1007/978-981-10-0335-6_5

[10] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016. [Online]. Available: https://doi.org/10.1109/ACCESS.2016.2628440

[11] "The LOLA website," last visited: June 2020. [Online]. Available: https://lola.conts.it/

[12] C. Drioli, C. Allocchio, and N. Buso, “Networked performances and natural interaction via lola: Low latency high quality a/v streaming system,” in *International Conference on Information Technologies for Performing Arts, Media Access, and Entertainment*. Springer, 2013, pp. 240–250.

[13] M. Bennis, M. Debbah, and H. V. Poor, “Ultrareliable and low-latency wireless communication: Tail, risk, and scale,” *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.

[14] “The Elk website,” last visited: June 2020. [Online]. Available: https://elk.audio/

[15] A. McPherson and V. Zappi, “An environment for Submillisecond-Latency audio and sensor processing on BeagleBone black,” in *Audio Engineering Society Convention 138*. Audio Engineering Society, 2015. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17755

[16] L. Turchet, F. Viola, G. Fazekas, and M. Barthet, “Towards a Semantic Architecture for Internet of Musical Things applications,” in *IEEE Conference of Open Innovations Association (FRUCT)*. IEEE, 2018, pp. 382–390.

[17] L. Roffia, P. Azzoni, C. Aguzzi, F. Viola, F. Antoniazzi, and T. Salmon Cinotti, “Dynamic Linked Data: A SPARQL Event Processing Architecture,” *Future Internet*, vol. 10, no. 4, p. 36, 2018.

[18] Z. Shelby, K. Hartke, and C. Bormann, “The constrained application protocol (coap),” Tech. Rep., 2014.

[19] F. Viola, L. Turchet, and G. Antoniazzi, F. Fazekas, “C Minor: a Semantic Publish/Subscribe Broker for the Internet of Musical Things,” in *IEEE Conference of Open Innovations Association (FRUCT)*. IEEE, 2018, pp. 405–415. [Online]. Available: https://ieeexplore.ieee.org/document/8588087

[20] *Study on audio-visual service production; Stage 1 (Release 17)*, 3GPP TSG-SA Tech. Rep. 22.827, Rev. 17.1.0, Dec. 2019.

[21] *AR framework architecture*, ETSI ISG ARF Group Spec. 003, Rev. 1.1.1, Mar. 2020.

[22] *System architecture for the 5G system; Stage 2 (Release 16)*, 3GPP TSG-SA Tech. Spec. 23.501, Rev. 16.1.0, Jun. 2019.

[23] *NR; NR and NG-RAN overall description; Stage 2 (Release 15)*, 3GPP TSG-RAN Tech. Spec. 38.300, Rev. 15.5.0, Mar. 2019.

[24] N. H. Mahmood, R. Abreu, R. Böhnke, M. Schubert, G. Berardinelli, and T. H. Jacobsen, “Uplink grant-free access solutions for URLLC services in 5G New Radio,” in *Int. Symp. on Wireless Commun. Syst. (ISWCS)*, Oulu, Finland, Aug. 2019, pp. 607–612.

[25] M. Centenaro, D. Laselva, J. Steiner, K. Pedersen, and P. Mogensen, “System-level study of data duplication enhancements for 5G Downlink URLLC,” *IEEE Access*, vol. 8, pp. 565–578, Dec. 2019.

[26] S. Kekki *et al.*, “MEC in 5G networks,” ETSI ISG MEC, White Paper 28, Jun. 2018.

[27] M. Centenaro, S. Tomasin, N. Benvenuto, and S. Yang, “Predictive voice-over-Internet Protocol fallback over vehicular channels: Employing artificial intelligence at the edge of 5G networks,” *IEEE Vehicular Technology Magazine*, vol. 15, no. 2, pp. 72–78, Jun. 2020.

[28] L. Turchet, J. Pauwels, C. Fischione, and G. Fazekas, “Cloud-smart musical instrument interactions: Querying a large music collection with a smart guitar,” *ACM Transactions on the Internet of Things*, 2020 (In press).

[29] F. Pachet, “The continuator: Musical interaction with style,” *Journal of New Music Research*, vol. 32, no. 3, pp. 333–341, 2003.

[30] *Service requirements for the 5G system; Stage 1 (Release 17)*, 3GPP TSG-SA Tech. Spec. 22.261, Rev. 17.1.0, Dec. 2019.

[31] “Elk Audio OS,” last visited: June 2020. [Online]. Available: https://elk.audio/audio-os/

[32] R. Fanou, F. Valera, P. Francois, and A. Dhamdhere, “Reshaping the african internet: From scattered islands to a connected continent,” *Computer Communications*, vol. 113, pp. 25 – 42, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366417303535

[33] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, “5G-enabled tactile Internet,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, 2016.