AI-driven Programmable Protocols for 6G and Beyond

Pranav Jha*, Shwetha Kiran*, Rashmi Kamran*, Prasanna Chaporkar*, Abhay Karandikar*†
Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India*
Secretary to the Government of India, Department of Science & Technology, New Delhi, India†
Email: pranavjha@ee.iitb.ac.in*, shwethak@iitb.ac.in*, rashmi.kamran@iitb.ac.in*,
chaporkar@ee.iitb.ac.in*, karandi@ee.iitb.ac.in*

***Abstract*—Artificial Intelligence/Machine Learning (AI/ML) technologies have started making a rapid foray into mobile communication networks. It is anticipated that 'AI for networks', i.e., incorporation of AI/ML technology into network design, will play a crucial role in future mobile networks. One area that may benefit significantly from the usage of 'AI for networks' is the design and optimization of communication protocols in mobile networks. Communication protocols, the building blocks of all communication networks, have traditionally been organized into fixed layered architectures. In addition, they typically possess immutable behaviour with limited user/service specific variations. For example, existing networks, such as the 5th Generation (5G) network, uses the same set of protocols to support all users and services, with limited user/service specific configurations. Considering the requirement to support a vast array of services in future mobile networks, it is time to rethink this traditional ossified layered protocol architecture. In this paper, we propose programmable and adaptive protocols based future mobile network architecture. The proposed architecture can harness AI/ML technology to dynamically design, program and optimize the communication protocols. As discussed in this paper, a mobile network architecture with an AI-driven flexible and customizable communication protocols, is better suited to handle diverse user/service/system requirements, network conditions, and performance requirements of future.**

***Index Terms*—AI/ML based Mobile Networks, Service specific Communication Primitives selection, Future Networks, 6G.**

I. Introduction

Cellular mobile networks, originally designed to cater to a single type of user, mobile human users, and to provide a single service, the voice service, have emerged as the primary vehicle for connectivity supporting an increasing diversity of services like multimedia streaming, video, online gaming, industrial automation, sensing and augmented reality. It supports a proliferating user diversity as well, from human users to machine type users, and from stationary users to users moving at extremely high-speeds, e.g., upto 500 km/h [1]. The user and service diversity is expected to grow further in the 6th Generation (6G) or International Mobile Telecommunications (IMT)-2030 networks [2].

Mobile communication networks comprise a set of Network Entities/Functions (NEs/NFs) connected via communication interfaces. The individual communication interface uses a layered protocol architecture, with a fixed set of interface-specific protocols. For example, the radio (Uu) interface between a User Equipment (UE) and a base station (gNodeB (gNB)) in the 3rd Generation Partnership Project (3GPP) 5G network uses PHYsical (PHY) layer, Medium Access Control (MAC) layer, Radio Link Control (RLC) layer, Packet Data Convergence Protocol (PDCP) layer, and Service Data Adaptation Protocol (SDAP) layer in user (data) plane. User data traverses through a fixed and predetermined set of 5G NE/NFs, e.g., gNBs, User Plane Functions (UPF), and interfaces, e.g., Uu and N3, utilizing the same set of protocol layers. All user data flows (services) are treated in the same manner by the network via creation of data tunnels to carry the service flows. The behavior of the protocol layers is the same with limited user/service-specific variations. While this approach of using a fixed set of interfaces and protocols has been in vogue in mobile networks for long, it may not be the most suitable to efficiently cater to the enormous range of emerging user and service requirements.

It should be noted that there have been efforts to enhance the adaptability and flexibility of mobile networks over the years. One of the capabilities added in this direction is the heterogeneity of Radio Access Networks (RAN). Multiple access types, such as cellular access, Wireless Local Area Network (WLAN) access and satellite access have been included in a converged 5G System (5GS) to efficiently support varied services and users. The time is ripe now to extend this idea further towards design of flexible and adaptable communication protocols and interfaces.

Usage of Artificial Intelligence/Machine Learning (AI/ML) in existing mobile networks, e.g. the 5G network, is currently limited to service plane. The 5G network provides an end-to-end data path to support different AI/ML services with required Quality of Service (QoS), i.e., supports 'network for AI' paradigm. However, as explored in this article, we can utilize 'AI for networks' paradigm for run-time design, evolution and optimization of communication interfaces and protocols in future (6G) mobile networks enabling user/service-specific adaptive and efficient protocols in these networks.

Individual communication protocols in mobile networks comprise a set of granular communication messages (or communication primitives). These granular communication primitives, with precise semantics, are exchanged between peer network entities to enable information transfer between them. Examples of such communication primitives are:

- 3GPP 5G New Radio (NR) MAC layer - Random Access

Preamble Transmission (sent by a UE to a gNB) [3]
- 3GPP 5G NR - MAC layer - Timing Advance Command (gNB to UE) [3]
- 3GPP 5G NR - MAC layer - Scheduling Request (UE to gNB) [3]
- 3GPP 5G NR – RLC layer – Acknowledge Mode Data Transfer (UE/gNB to gNB/UE) [4]
- 3GPP 5G NR - PDCP layer - Ciphering [5]
- 3GPP 5G NR - PDCP layer - Integrity Protection [5]
- IEEE 802.11 – Beacon frame (broadcast by IEEE 802.11 Access Point (AP)) [6]
- IEEE 802.11 - Association Request frame (Mobile Station (MS) to AP) [6]
- IEEE 802.11 Data frame (MS/AP to AP/MS) [6]

As is obvious, the set of communication primitives used by a specific protocol layer (and the interface) of a mobile network does not change over time and across nodes; the structure and semantics of these primitives (i.e., the type of information and the number of bits they carry) are typically defined during the standardization process of the technology (and documented in a standard). These primitives continue to be used by NEs/NFs with extremely limited and slow evolution during the lifetime of the technology. For example, the MAC layer on a 3GPP 5G gNB and 5G UE will always use the primitives as defined in the 3GPP TS 38.321 [3] irrespective of the network condition or user requirements. Therefore, even if there is only one UE connected to a 5G gNB, the UE MAC layer will still need to send the 'Scheduling Request' primitive to gNB to request radio resources from the gNB. User data can be sent by the UE only when dedicated radio resources are granted by the gNB. Interestingly, it may not be desirable for the UE to follow the above sequence in this scenario from the efficiency perspective; instead it could directly send the data over the radio channel similar to an IEEE 802.11 data frame (following Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) scheme), as there is no other UE connected to the gNB and there will not be any conflict if it chooses to do so. However, this kind of flexibility is currently impossible as the 3GPP 5G MAC layer primitives are not allowed to change dynamically.

In this paper, we propose a programmable, adaptive protocols based mobile network architecture that can support such scenarios. It utilizes user/service/system requirements, network conditions, and the network performance to dynamically select and optimize the communication primitives used between network nodes thereby supporting programmable protocols instead of protocols with a predefined set of communication primitives and immutable behaviour. Such programmable protocols are apparently feasible in this era of softwarized networks, when most network functionality is written in software. Furthermore, AI/ML techniques can be used to aid in the selection and optimization of communication primitives. To summarize, the network nodes in the proposed architecture do not utilize a fixed and immutable set of communication protocols, e.g., there are no IEEE 802.11 APs or gNBs in the RAN as far as the communication protocols used by the radio nodes are concerned. Every pair of a UE and a BS (communicating with each other) use a dynamically curated communication protocol (from the selected primitives), which evolve over time in conjunction with user/service/system requirements and network conditions.

The rest of the paper is organised as follows: Section II covers literature review. Section III identifies the limitations of the existing network protocol design. Section IV identifies the requirements for an adaptive protocol design in the mobile network. Section V details the concept of automated protocol optimization and communication primitives selection using AI/ML. Section VI discusses the benefits and challenges of implementing the proposed concept in mobile networks. Further, a conclusion is provided in Section VII.

II. Literature Review

As modern networks become more intricate, conventional protocol optimization techniques are proving to be less efficient as they require extensive manual design and tuning. Recent advancements in AI/ML methods have captured the interest of the networking community. With their capability to operate in complex environments and support decision-making processes, AI/ML-based approaches offer the potential to enhance network performance. This work [7] proposes AI-powered mobile network architecture which uses deep learning algorithms to achieve efficient and intelligent admission control strategies for QoS provision. This article [8] delves into active queue management implementation in high-latency environments within disaggregated RAN deployments for 5G and beyond networks. In this work [9], AI/ML-driven protocol selection for individual flows is proposed using a multi-armed bandit-based learning algorithm for different use cases. However, flows are categorized under different protocols based on transmission rates only, which makes the scope of this work limited. Following few articles have focused on redesigning MAC layer protocol using AI/ML methods: [10] introduces a Deep Reinforcement Learning (DRL) framework to automate and optimize protocol design. By decomposing protocols into functional blocks, the proposed DeepMAC system learns and adapts MAC layer configurations in 802.11 WLANs based on network conditions. Article [11] proposes a DRL-based framework using Proximal Policy Optimization to design adaptive, application-specific MAC protocols that outperform IEEE 802.11ac in throughput and latency. The framework allows flexible protocol reconfiguration to meet diverse QoS needs. Work in [12] presents a novel Multi-Agent DRL framework for distributed MAC protocol design, enabling individual nodes to learn and adapt from local observations within the ns-3 environment. Authors in [13] explore the use of on-device machine learning to enhance MAC layer protocols, particularly focusing on Multi-User Multiple Input Multiple Output (MU-MIMO) grouping, demonstrating improvements in system capacity, latency, and adaptability across WLAN and Massive MIMO scenarios. These works focus either on achieving network optimization using existing protocols or enhancing MAC protocol generalization through observation abstraction in learning-based systems. Our proposal goes further by enabling programmable and adaptive protocol design across all layers, providing a holistic architecture that supports dynamic protocol customization.

III. Limitations of the existing network architecture and protocol design

As discussed above, existing mobile networks utilize a fixed suite of communication protocols comprising a set of predetermined and immutable granular communication primitives with precise semantics. The communication protocols and the corresponding primitives differ across Radio Access Technologies (RATs), such as, WLANs, 5G NR or 3GPP Long Term Evolution (LTE) but are unchangeable within a RAT. One of the limitations of the existing mobile network architecture and the protocol design is that there is no service/user specific handling in these networks; in essence, all users and services are treated more or less in the same manner by existing networks and protocols. Fixed routes/paths are defined for all user communication, with a predetermined set of network nodes in the path. These network nodes use a layered architecture with a fixed set of unchanging communication protocols. Communication primitives of these protocols are not dynamically selected, adapted and applied in existing mobile networks as per the traffic conditions, performance indicators, user/service and network requirements. Such a static network and protocol design may not be best equipped to handle diverse user/service/network requirements of the future. Examples of two RAN nodes (LTE and 5G-NR) are shown in Figure 1, the behaviour of the protocol layers for all UEs connected to these RAN nodes does not change over time.

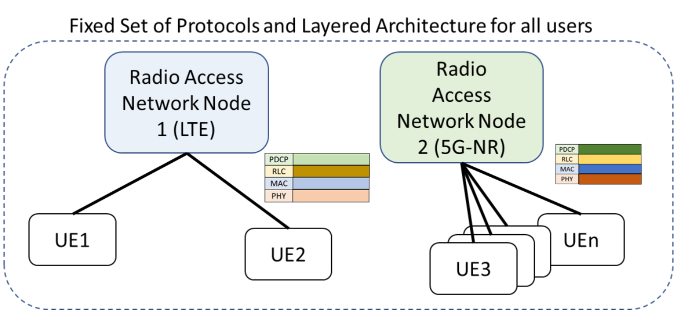

Fig. 1. Fixed Set of Protocols and Layered Architecture for all users in existing mobile networks.

In this context, the following questions on protocol design for future mobile networks has to be explored:

- Is our current approach of using fixed set of protocols and nodes appropriate for future networks?
- Can we use core network nodes in the data path only for a subset of users according to their needs? For example, place them in the data path for mobile users but not for stationary users.
- Can service specific protocols be flexibly chosen for a particular flow? For example, do not use GPRS Tunnelling Protocol (GTP) tunnels for stationary users.
- Can different protocols (or variants) be used for different applications or users? e.g., Different protocols for 'low latency' vs. 'latency tolerant' applications.
- Are different variants of control plane protocols (e.g., Radio Resource Control (RRC) or Non-Access Stratum (NAS)) and different variants of User Plane Protocols (e.g., MAC, RLC, PDCP, SDAP) possible?
- Can the set of communication primitives, used by a protocol, be changed and adapted dynamically based on real-time network conditions and user/service/network needs instead of using an immutable structure?

Solution to some of the questions above, if not all, requires a dynamic approach to select, evolve and apply communication primitives on mobile network nodes. Adaptive selection, evolution, and application (or configuration) of communication primitives at network nodes are very much possible in this era when network nodes are typically software-defined, i.e., implemented in software. The selection/optimization of specific communication primitives between communicating nodes in a network can be based on AI/ML. The approach can help achieve a flexible and adaptive protocol design in mobile networks and will result in better network performance using the same set of resources.

IV. Considerations for an adaptive protocol design for Future Networks

Following are some considerations for designing an autonomous 6G network based on selection and optimization of communication primitives.

- The system should allow runtime selection and optimization of communication primitives used between different network nodes.
- A base set of communication primitives for the system may comprise of primitives associated with protocols from different wireless networks.
- The system can support a centralized mechanism for selecting and applying communication primitives on mobile network nodes.
- The selection/adaptation of communication primitives should take into account user/service requirements (QoS), other network requirements like network energy efficiency, performance and network conditions.
- The system should be able to apply optimized communication primitives on network nodes dynamically.

In summary, dynamic and adaptive protocol design should lead to an efficient mobile network. System requirements like Service Level Agreements (SLAs) and Key Performance Indicators (KPIs) (e.g., number of users served, total throughput) are important goals to meet in this context.

V. Adaptive protocol design and optimization in 6G network: a proposed approach

Taking into account the above considerations, we propose a dynamic and adaptive protocol optimization for 6G networks. In this regard, we explore the possibility of an AI/ML framework, learning from real-time conditions to dynamically select, adapt and apply appropriate communication primitives in 6G networks. The network nodes shall use dynamically changing communication protocols instead of a fixed and immutable protocol, as is the case today. One possible approach is to use Reinforcement Learning, starting with base Communication Primitives (CPs) to analyze real-time network conditions and performance and arrive at a more appropriate set of communication primitives to meet the requirements. Figure 2 provides an overview of the conceptual idea of communication primitives set for different protocols/layers and selecting service-specific CPs based on diversified needs. For example, Service-2 which is a messaging app kind of service may not need any CP for PDCP and RLC, whereas Service-1 (buffered video) and Service-3 (Voice over IP (VOIP)) may need CPs associated with all layers. These examples will become clearer in Section V-B where primitives selection for few use cases are explained in detail.

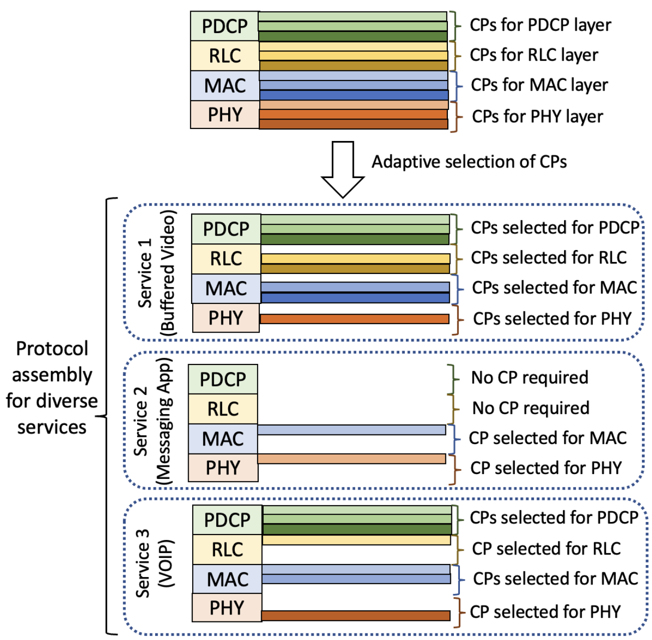

Fig. 2. Defining Communication Primitives (CPs) set for different protocols and selecting service specific CPs based on diversified needs.

A. Programmable Protocols based Network Architecture

Figure 3 provides an overview of the architecture that allows communications primitives selection and optimization. Network Orchestrator under Management Functions provides the base primitives and system requirements for bootstraping the mobile network to begin with. Service Orchestrator collects service level requirements from the users and supplies it to the AI/ML framework. The framework learns from the base communication primitives, as well as data collected from the mobile network (UE, RAN and Core network functions). In addition, the framework is also aware of existing network conditions like network traffic load, mobility and radio conditions of the UE and QoS requirements for the service. As the framework selects and optimizes the minimum set of primitives to meet service requirements, it is applied to the different network nodes of the mobile network. Other system requirements like data throughput, number of users supported, would also be considered to arrive at a set of optimized communication primitives to support a particular user/service.

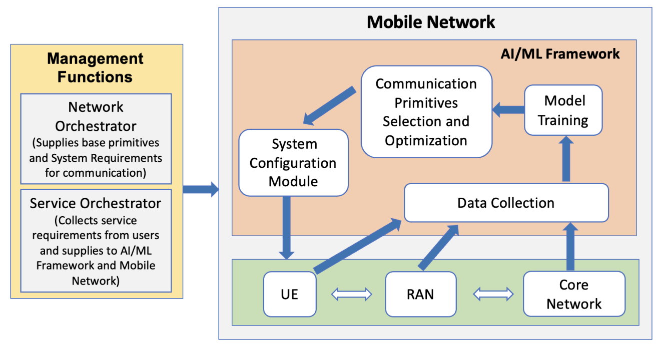

Fig. 3. Architectural framework for AI/ML based Communication Primitives optimization

B. Communication primitives selection use cases

Figure 4 provides few examples of CP selection based on the traffic conditions and service requirements. The labeling and description of the CPs for the examples use cases are listed in Figure 5. CPs for functionalities like scheduling and medium access, error control, and encryption/integrity protection are considered to explain the concept of AI/ML driven CP selection. Please note that a base CP set (preconfigured at the network entities) can be used for any incoming UE for initial communication. Based on dynamic selection considering real time scenarios and user/service specific requirements, this CP set can be reconfigured dynamically. Few use cases for dynamic selection of CPs for UEs based on traffic types are as follows:

- Let's say UE1 is consuming multimedia content like news. For optimized delivery of data in this case, following primitives can be selected: CPs1, CPr2 and CPi3. CPs1 is required for data transfer over a common channel. CPr2 is required for 'in sequence' delivery of media content. Since it is news, there is no need to encrypt data, hence only header compression i.e., CPi3 is chosen. CPs1 here refers to MAC layer primitive of IEEE 802.11, where UEs share a common channel for communication with the access network node. However, in sequence (CPr2) delivery of data and header compression (CPi3) are communication primitives from the 3GPP 5G/LTE access. To summarize, we can use relevant primitives from different access technologies to ensure optimal handling of a particular service.
- Assuming UE2 is downloading a file and following primitives are chosen to achieve efficient data delivery: CPs1, CPr1, CPr2, CPi1 and CPi2. CPs1 is needed for data transfer over a common channel, CPr1 is chosen to provide Automatic Repeat reQuest (ARQ). CPr2 is needed to ensure 'in sequence' delivery of data. CPi1

provides ciphering and CPi2 provides integrity protection. Since UE2 is engaged in file download and dealing with large packet sizes, usage of header compression primitive (CPi3) is not necessary. Similar to the UE1 example, here also we are using CPs1, the IEEE 802.11 MAC layer primitive for data transfer, whereas CPr1, CPr2, CPi1 and CPi2 are primitives from the 3GPP 5G/LTE access. In essence, relevant primitives are selected, irrespective of the technology they come from, for optimized handling of a particular service. This fundamental change in the way we look at the mobile network is the novelty of this solution.

- Since there are only 2 UEs under RAN Node 1, a simpler CSMA/CA (used in IEEE 802.11) kind of access scheme would suffice. Later when more UEs join RAN Node 1, CSMA/CA access could be switched to centralized control (based on 3GPP LTE/5G).
  This kind of dynamic switching of access methods brings additional flexibility to the system, especially in the era of software defined networks and radios. This is another fundamental change in the way we think about access technologies, it doesn't have to be fixed for the entirety of its life cycle, rather customizable based on the network conditions and requirements.

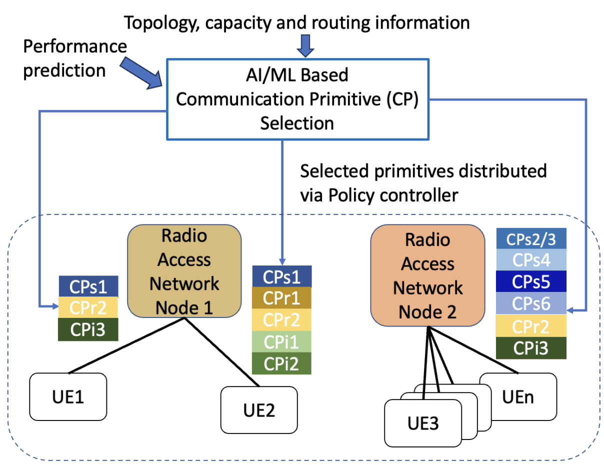

Fig. 4. Examples of Communication Primitives selection in the network.

- UE3 connected to RAN Node 2 is located slightly farther away from the base station, the associated user is engaged in a financial transaction (internet traffic). Following primitives are selected to achieve reliable unicast communication: CPs2, CPs3, CPs4, CPs5, CPs6, CPr2, CPi3. CPs2/3 provides scheduling request and response, CPs4 provides data transfer over dedicated resource, CPs5 provides timing advance for the radio channel considering the UE is slightly farther away from the base station. CPs6 provides Hybrid ARQ (HARQ) to ensure error correction handling and CPr2 provides 'in sequence' delivery of data. CPi3 provides header compression and since security is already handled at the application (https) layer additional encryption/integrity primitives (CPi1/2) are not required here. All the primitives used here are based on 3GPP LTE/5G.

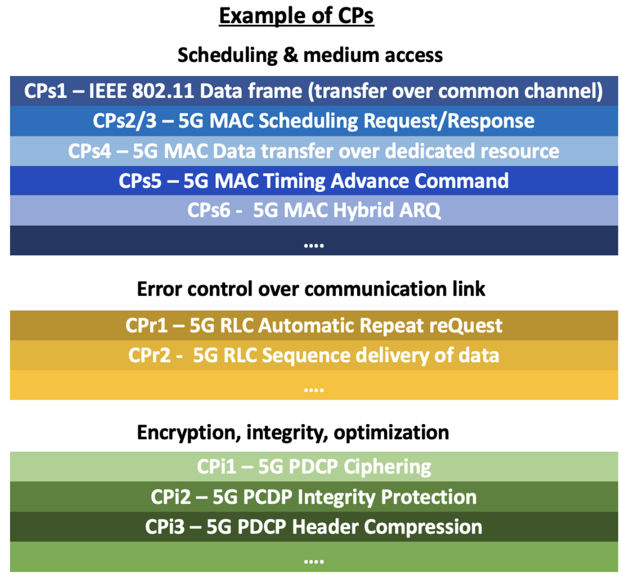

Fig. 5. Communication primitives set considered for the examples shown in Figure 4.

- As shown in figure 4 many UEs are connected to RAN Node 2, and it uses 5G/LTE based centralized access control as that may be ideal.
- If some of the UEs under RAN Node 2 are assumed to be stationary, then timing advance CP wouldn't be needed to serve them.

Considering these examples, it is evident that the network can forego redundant processing and achieve efficiency. Furthermore, modifiable network nodes and automation of such modifications as explained in above examples is closer to reality with technological advancements like AI/ML and Software Defined Networking (SDN).

VI. Benefits and Challenges

The proposed framework is a fundamental design-level transition from the conventional network architecture and protocol design principles, taking into account the diversity of services expected from future networks. It brings many benefits, a few are described below:

- Improved performance: The scheme will likely have a significant impact on the system performance. As shown earlier, existing mobile network protocols have significant redundancies, which brings down the network and the service performance. Selection of appropriate communication primitives can improve network and service performance considerably by saving redundant procedures and resources. It can also reduce the overall cost of providing services in the network.
- Flexibility and Adaptability: The proposal allows for dynamic selection, and optimization of the communication protocols in the network, rendering immense flexibility and adaptability to the network in heterogeneous

environments. There are no RAT-specific nodes with fixed protocol behaviour. Every node is fungible and can transform its protocol behaviour with changes in state/requirements/performance.

- Higher energy efficiency: AI/ML-based optimization can consider energy-related constraints while selecting communication primitives that can result in improving the end-to-end energy efficiency of the network.

The key challenge in the implementation of this proposal may be the requirement of extensive learning of network conditions in real-time, which may not be possible in every scenario.

VII. Conclusion and Future Work

This paper introduces the idea of a novel mobile network architecture supporting programmable protocols through dynamic communication primitive selection and optimization. User/service requirements, network conditions, network performance, and other system requirements, such as data throughput are considered for selection and optimization of communication primitives. The system may start with a set of base communication primitives (base protocols) and evolve a new set of optimized communication primitives (protocols) using an ML-based scheme. The evolved primitives are applied to the peer communicating nodes of the mobile network. Since communication protocols are the heart of a communication network, the approach may allow for an immensely flexible and adaptable mobile network, which may dynamically evolve according to the requirements and the current state. The automated approach towards programmable protocol design as proposed here may go a long way towards an automated network design in future. Such an approach has the potential to arrive at an autonomous mobile network, which may allow for evolution of network protocols and emergence of a suitable network architecture from the current network state and requirements. It is also quite possible that such protocols and networks are more efficient viz-a-viz human-designed protocols and networks. As part of our future work we will investigate the usage of AI/ML based schemes towards selection and optimization of communication primitives in mobile networks.

Acknowledgments

We acknowledge the Department of Telecommunications (DOT), Ministry of Communications, Government of India for supporting the project.

References

[1] ITU-R M.2083-0, "IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond," *Recommendation*, 2015.
[2] ITU-R M.2160-0, "Framework and overall objectives of the future development of IMT for 2030 and beyond," *Recommendation*, 2023.
[3] 3GPP TS 38.321 V18.4.0 (2023-09), "Technical Specification Group Radio Access Network; NR; Medium Access Control (MAC) protocol specification(Release 18)," *Technical Specification*, 2024.
[4] 3GPP TS 38.322 V18.2.0 (2023-09), "Technical Specification Group Radio Access Network; NR; Radio Link Control (RLC) protocol specification (Release 18)," *Technical Specification*, 2024.
[5] 3GPP TS 38.323 V18.4.0 (2023-09), "Technical Specification Group Radio Access Network; NR; Packet Data Convergence Protocol (PDCP) specification(Release 18)," *Technical Specification*, 2024.
[6] IEEE, "IEEE Standard for Information technology-Telecommunications and information exchange between systems Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *Standard*, 2016.
[7] G. Luo, Q. Yuan, J. Li, S. Wang, and F. Yang, "Artificial intelligence powered mobile networks: From cognition to decision," *IEEE Network*, vol. 36, no. 3, pp. 136–144, 2022.
[8] A. Stoltidis, K. Choumas, and T. Korakis, "Active Queue Management in Disaggregated 5G and Beyond Cellular Networks Using Machine Learning," in *2024 19th Wireless On-Demand Network Systems and Services Conference (WONS)*, 2024, pp. 113–120.
[9] X. Zhang, S. Chen, Y. Im, M. Gorlatova, S. Ha, and C. Joe-Wong, "Towards Automated Network Management: Learning the Optimal Protocol Selection," in *2019 IEEE 27th International Conference on Network Protocols (ICNP)*, 2019, pp. 1–4.
[10] Pasandi, Hannaneh Barahouei and Nadeem, Tamer, "Towards a Learning-Based Framework for Self-Driving Design of Networking Protocols," *IEEE Access*, vol. 9, pp. 34 829–34 844, 2021.
[11] Keshtiarast, Navid and Petrova, Marina, "ML Framework for Wireless MAC Protocol Design," in *2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, 2024, pp. 560–565.
[12] Keshtiarast, Navid and Renaldi, Oliver and Petrova, Marina, "Wireless MAC Protocol Synthesis and Optimization With Multi-Agent Distributed Reinforcement Learning," *IEEE Networking Letters*, vol. 6, no. 4, pp. 242–246, 2024.
[13] Pasandi, Hannaneh B. and Nadeem, Tamer, "Autonomous On-Device Protocols: Empowering Wireless with Self-Driven Capabilities," in *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, 2024, pp. 1–6.