Leveraging Synergies Between AI and Networking to Build Next Generation Edge Networks

Sen Lin[1*], Ming Shi[1*], Anish Arora[1], Raef Bassily[1], Elisa Bertino[2], Constantine Caramanis[3], Kaushik Chowdhury[4], Eylem Ekici[1], Atilla Eryilmaz[1], Stratis Ioannidis[4], Nan Jiang[5], Gauri Joshi[6], Jim Kurose[7], Yingbin Liang[1], Zhiqiang Lin[1], Jia Liu[1], Mingyan Liu[8], Tommaso Melodia[4], Aryan Mokhtari[3], Rob Nowak[9], Sewoong Oh[10], Srini Parthasarathy[1], Chunyi Peng[2], Hulya Seferoglu[11], Ness Shroff[1], Sanjay Shakkottai[3], Kannan Srinivasan[1], Ameet Talwalkar[6], Aylin Yener[1], Lei Ying[8]

[1]The Ohio State University, [2]Purdue University, [3]University of Texas, Austin, [4]Northeastern University,
[5]University of Illinois, Urbana Champaign, [6]Carnegie Mellon University, [7]University of Massachusetts, Amherst,
[8]University of Michigan, [9]University of Wisconsin, Madison, [10]University of Washington, [11]University of Illinois, Chicago

***Abstract*—Networking and Artificial Intelligence (AI) are two of the most transformative information technologies over the last few decades. Building upon the synergies of these two powerful technologies, we envision designing next generation of edge networks to be highly efficient, reliable, robust and secure. To this end, in this paper, we delve into interesting and fundamental research challenges and opportunities that span two major broad and symbiotic areas: AI for Networks and Networks for AI. The former deals with the development of new AI tools and techniques that can enable the next generation AI-assisted networks; while the latter focuses on developing networking techniques and tools that will facilitate the vision of distributed intelligence, resulting in a virtuous research cycle where advances in one will help accelerate advances in the other. A wide range of applications will be further discussed to illustrate the importance of the foundational advances developed in these two areas.**

I. Introduction

The astonishing successes of Artificial Intelligence (AI), especially in Machine Learning (ML), provide an opportunity to design next generation (XG) networks that are "intelligent" in many different ways. In particular, the focus will be on the wireless edge, since most of the growth is expected to happen with wireless devices in the network edge, and not the network core. Imagine a network edge that supports diverse devices, such as robots, mobiles and self-driving cars, and a wide range of distributed AI applications, such as smart aerospace, remote healthcare, smart manufacturing, AI-assisted education and the operation of 6G and beyond (6G+) cellular networks. We envision such a future network edge, managed by AI, supporting distributed intelligence. The physical infrastructure (towers, radios, routers, edge cloud, etc.) of this wireless edge will be controlled, managed and maintained by AI driven service robots (connected through the same network). Further, the virtual infrastructure (link optimization, load balancing, anomaly/outage detection and self-healing, etc.) of the network will be maintained by intelligent network agents that continually learn and evolve, without the need for human intervention. In the meanwhile, these AI agents and physical devices need to be robust to failures, extreme loads, weather events, and adversarial attacks.

However, it is extremely challenging to develop a distributed intelligent plane in order to control these inherently complex edge networks, consisting of large numbers of dispersed network elements, comprising heterogeneous software and hardware components and modules. Traditional approaches to managing networks have been based on heuristic designs derived from domain knowledge or using (simplified) theoretical models. Neither of these approaches will be suffice to control these networks because of their scale and complexity; their mobility and dynamics; the stringent constraints posed by their applications; and the expectations of for security, privacy, and adaptability. Moreover, an AI-centric black box approach will not work either, because today's successful AI algorithms require massive amounts of data co-located with computation. In stark contrast, the fast time scales and the decentralized nature of networks and available data force instead a distributed paradigm for AI in the edge networks that requires the proper blending of domain knowledge and AI.

In order to address these challenges, we envision promising research directions from two different perspectives, i.e., AI for Networks and Networks for AI, where advances in one will accelerate advances in the other through a virtuous cycle. To design next generation hyperscalable heterogeneous and dynamic networks that are highly efficient, reliable, robust, and secure, new AI tools and techniques will need to be developed to ensure that these networks are self-healing and self-optimized. On the other hand, these networks will in turn

*Equal contribution; Corresponding authors. Email: lin.4282@osu.edu, shi.1796@osu.edu.

be designed to unleash the power of collaboration to help solve long-standing AI challenges, making AI more efficient, interactive, and privacy preserving.

The rest of the paper is organized as follows. In order to make progress on solving the grand challenges on AI for networks and Networks for AI, we take a divide and conquer approach. We first introduce AI for Networks in Section II, where new AI techniques will be designed to achieve four goals: (1) re-engineering the physical fabric for 6G+ wireless communications; (2) designing and controlling next generation networks by taking into account practical resource constraints; (3) generalizing for multi-agent, possibly non-cooperative, network entities; (4) guaranteeing that the network edge is secure and intrusion-free. In Section III, we focus on Networks for AI, where the networks will be re-engineered such that (1) distributed AI adapts its operation seamlessly by taking into account physical constraints; (2) communication and computation resources are adaptively allocated to serve the needs of distributed AI applications; (3) collaborative analysis of data will be enabled for effective AI; (4) protection from information leakage and attacks will be achieved. Applications for these research directions are presented in Section IV, followed by the conclusions in Section V.

II. AI for Networks

While networks have evolved over time to incorporate the latest technological advances and improve fabric capacity, it is human professionals that determine policies and technical advances. Instead, we aim to leverage the power of AI towards designing networks with minimal human interventions for most network operations. We propose to design a distributed intelligence plane that has two goals to impact current and future networks. First, current networks could be renovated by replacing the antiquated management plane with autonomously interacting AI agents that adaptively optimize the data and control planes for efficient and robust performance. Second, we could re-imagine secure future networks through the lens of an AI-driven distributed intelligence to co-design the data, control and management planes.

A. AI for Physical Layer Design

Due to the complex, non-stationary and distributed nature of current and future networks, it becomes significantly important to leverage the power of AI to not only better understand the dynamic physical environment, but also make it controllable. This points to the need of using AI to leverage the domain knowledge to develop better networking solutions, and also designing new AI algorithms to engineer the physical environment itself for further improved performance.

1) Leveraging Physical Knowledge: There have been increasing interests in incorporating physics-based models with neural network architectures to solve complex physics problems [1]–[4]. Two of the places where AI could play a critical role are (i) to expedite the estimate or prediction for the physical environment in the network [5], and (ii) to encode relevant physical models of interest as a layer within a multi-layer neural network for faster convergence [6]. For example, in the mmWave beam alignment problem, non-stationary kernelized bandits under inference constraints have been studied in the time-varying multi-path environment [7]. By leveraging the beam correlation information, a constrained UCB-type of algorithm was proposed with theoretically-guaranteed low regret and constraint violations. It is also of great interests to design physical-based learning techniques so as to improve the network performance in various scenarios, e.g., beam-scan time and beam-selection performance.

2) Engineering the Physical Environment: In the edge networks, each sub-system could contain heterogeneous resource-constrained edge devices in order to satisfy different service requirements. As a result, it becomes more important to adaptively engineer the network topology based on the mobility of the edge nodes and actively change the physical environment for communication [8], [9]. For example, an over-the-air federated learning algorithm was proposed in [10] with joint adaptive computation and power adjustment. This algorithm adaptively chooses device local iterations and scales power at both the device and server levels. Moreover, [11] developed a channel-quality-aware over-the-air learning algorithm with theoretical convergence rate performance guarantees. This algorithm leverages channel state information estimation and adaptively scales channel inversion to mitigate the impacts of wireless channel fading. An important future direction is to adapt these algorithms to higher level network dynamics, e.g., network mobility, node density, number and type of applications, flows, and highly-dynamic edge demand.

3) Discovering Communication Algorithms via Deep Learning: Traditional efforts mainly focus on improving linear codes, e.g., turbo [12], low-density parity check [13], and polar codes [14], for reliable communication by individual human ingenuity. However, as ML has been playing a more and more important role in networking, utilizing ML to design more efficient codes, e.g., search for non-linear codes, becomes an essential idea. Although using ML could expedite discovery and expand the searching space of codes, it is required to design neural architectures that can handle many unique challenges, e.g., large number of code words, varying channel conditions and large block length. Promising gains have been shown in existing work [15], [16]. Recently, Kronecker Operation (KO) codes that outperform the state-of-the-art reliability performance on the standardized additive white Gaussian noise channel was proposed in [17]. KO codes are a family of computationally efficient deep-learning driven pairs of encoders and decoders. The design of KO codes verifies the ideas of utilizing ML in better exploring codes, and motivates great interests in further discovering a much richer class of new nonlinear algebraic structures to increase the capacity region of the network.

B. AI-based Network Resource Allocation and Control

Control and allocation of network resources lie at the core of every network. Compared with classical convex-optimization-based network utility optimization, leveraging online learning and reinforcement learning techniques to develop efficient, fair, and safe data-driven AI-empowered mechanisms has received considerable attention. However, due to the various new challenges, including non-stationary dynamics at multiple timescales, non-convex objectives and combinatorial constraints on resource availability, new methods and algorithms need to be developed.

1) Low-complexity and Sample-efficient AI-network Algorithms: AI has been an important tool to address the uncertain and complex network conditions in the operation of network services. Although new challenges arise due to the non-stationary, distributed and heterogeneous nature of the network systems, and uncertain network constraints, some recent results have already shown significant gains that can be reaped from employing advanced AI strategies for network optimization [6], [18]–[23]. For example, to improve the performance of vehicle detection in autonomous driving, [24] proposed several deep-learning-based frameworks for multi-modality fusing, which considerably outperform uni-modal detection. Moreover, safety constraints emerge as a critical obstacle when applying existing AI algorithms to practical network operation. For such a constrained Markov Decision process (MDP), model-free reinforcement learning (RL) approaches have been developed with theoretically-guaranteed low regret and constraint violation in [25]–[27]. Nonetheless, there still remain open problems on how to develop performance-guaranteed scalable AI algorithms for the dynamic and distributed network systems (under realistic hard and soft constraints [28]–[34]).

2) Algorithms with Mis-specified Models: Networks are highly complex systems, and classical control heuristics based on simplified theoretical models do not scale to hyper-scale networks due to model simplicity. Fortunately, [21] has shown that when the true environment model is realizable by a feature-based linear combination of base models, learning a near-optimal policy has polynomial sample complexity. However, significant challenges still exist when considering the real-worlds settings, due to the fact that a linear combination of base models may not accurately represent the dynamics of the complex networks. Thus, it is important to leverage non-linear function approximators (such as neural networks), in order to handle model mis-specifications and provide agnostic learning guarantees. To this end, a model-free RL algorithm was proposed in [35] for a mis-specified model where not only the average regret scales only polynomially with the mis-specification parameter, but also the space and per-episode time complexities are bounded when the number of episodes increases to infinitely large. Notably, to address the challenges from the real-world settings, there is still an urgent need to generalize the existing AI algorithms to more general MDP settings and fundamentally understand the power of deep learning in this direction.

3) Learning from Historical Data and Incomplete Network State: A practical challenge in network control is the lack of complete or sufficient data at runtime (e.g., in a multi-sensor scenario, LiDAR point clouds are not available to provide global information, even though camera images are available in a timely manner), so the network has to operate with limited knowledge of the network state. Because of this and the highly non-stationary network datasets, the promised performance guarantees using the traditional approaches may not be achievable. One promising remedy is to use local data for real-time control and also historical global data to determine policy. To this end, a new off-policy temporal difference learning method has been developed in [36], which improves upon the emphatic temporal difference learning [37] to conduct the off-policy value function evaluation with function approximation.

C. Multi-Agent Network Resource Allocation and Control

Sharing resources is central in networks that comprise a large and dynamic population of often competing, self-interested users. In addition to this, different parts of the global network may be subject to different local conditions and different organizational control and regulation. It is therefore crucial to develop AI-aided non-cooperative and distributed networks where resources are shared efficiently and fairly amongst self-interested users with dynamic demands.

1) Network as a Multi-Agent System: Sharing of resources and collaboration among agents have been prominent in current and future networks. In practice, each user can choose its own policy search algorithms, rather than a same type of algorithms that is studied in existing multi-agent learning method. This type of algorithmic heterogeneity requires us to start to seek new network stability and scalability [38], and the fundamental understanding of the underlying connections among different policy search algorithms [39]–[42]. It is also interesting to consider constraints in the multi-agent settings and to extend existing RL algorithms by taking the stability and scalability into account.

2) Fair Network Operations Among (Non-Cooperating) Users: A key challenge in network operation is the fair sharing of stochastic and heterogeneous resources under dynamic service requirements and network conditions. Fortunately, compared with the classical network utility optimization method, e.g., online convex optimization, AI is much more powerful in handling the non-convexity, unknown statistics and high dimensions. Thus, unique opportunities arise for designing new AI strategies to optimize the trade-off between efficiency and fairness in shared networks. A mini-batch Markovian sampled fully decentralized actor-critic algorithm was developed in [43] for fully decentralized multi-agent reinforcement learning problems. Even though there is no knowledge of joint actions, the sample complexity of this algorithm only scales polynomially with the number of agents. There are many other interesting future directions, including achieving theoretical performance guarantees under shared constraints, RL-based fair algorithm among non-cooperating users, and reducing communication overhead among agents.

3) Data Sharing and Augmented Learning for Distributed Network Operation and Resource Utilization: It is well-known that data sharing can lead to better system-level outcomes, e.g., the spectral efficiency gain in a distributed multi-user spectrum sensing when users are allowed to communicate and exchange information [44], [45]. Thus, an important open question when designing AI for network is to understand how much we can potentially gain by leveraging the shared data in the learning algorithms, while taking the adversarial inputs, delayed feedback and privacy into consideration. To this end, a distributed online algorithm has been developed in [46] for content allocation in networks of caches that attains a low regret under adversarial topology. Moreover, [47] investigated the HSIC (Hilbert-Schmidt independence criterion) bottleneck as a regularizer for learning an adversarially robust deep neural network classifier. It has been shown that the HSIC bottleneck enhances robustness to adversarial attacks both theoretically and experimentally. Besides exploring the fundamental power of data sharing, it is also an interesting future direction to design frameworks to incentivize data sharing mechanisms that not only improve performance but also satisfy the privacy requirements.

D. AI for Network Security

Resourceful adversaries (e.g., nation-states, terrorists, cyber criminals) can create havoc in network ecosystems and carry out malicious activities by exploiting network vulnerabilities. New wireless networks, like 5G, WiFi 6 and the envisioned 6G, bring certain security advantages. For example, in mmWave communications, directionality can help mitigate spoofing and eavesdropping attacks. However, securing next-generation networks is a formidable challenge because of inherent security requirements, increased network complexity, huge numbers of connected devices, and low latency requirements. Security must cover network design, operation, and evolution across both anticipated uses (common cases) and unanticipated-yet-possible uses (corner cases). To this end, AI tools can be leveraged to build a comprehensive network security framework that provides secure foundations for the networks, protects the networks through a security life-cycle and enhances security mechanisms.

1) Systematic Analysis of Network Protocol Specification: A major issue with protocol specifications, especially for the ones expressed in natural language as in the case of standardization documents, is the underspecification of steps in protocols. As shown in [48], underspecifications can lead to a number of vulnerabilities in protocol implementations, for which the develops do not always take the most secure solutions. As a formal verification, a systematic framework named as VMAnalyzer was proposed in [49] to perform security analysis on the Voice over WiFi (VoWiFi) protocol. By modeling the VoWiFi protocol as finite state machines (FSMs), VMAnalyzer can generate model variants to remove the underspecifications. However, extracting FSMs requires huge manual effort, especially for natural language specifications as in standardization documents. By leveraging domain-specific patterns, sentence-level natural language processing (NLP) techniques can be developed as a promising solution to this problem.

2) Systematic Analysis of Network Protocol Implementations: Network protocol implementations often have various bugs [50]–[52], which makes the analysis of these implementations, e.g., model checking, clearly more important to guarantee the network security. Particularly, several implementation flaws have been identified in cellular protocols by analyzing the security and noncompliant behaviors with the cellular standard [48], [53]–[55]. However, traditional approaches hinge heavily upon manual analysis and the quality of the properties being tested, which can be error-prone and time-consuming. To address this, [56] developed a testing framework to enable automated noncompliance checking for commercial 4G LTE device implementations, by utilizing black-box automata learning to extract input-output FSMs. Besides, since most implementations are proprietary and the corresponding source codes not completely open, causality inference can be leveraged to examine hidden implementations and policies behind the observed consequences.

3) Learning and Adapting Network Security Policies: Fine-grained policies, such as policies for firewalls, access control, security service chaining, and dynamic definitions of virtual networks, are critical defense mechanisms. Correspondingly, the generation of those policies is a key security life-cycle activity, where manual specifications however are often infeasible because the policies are attribute-based and complex. AI techniques can be used to learn the security policies from logs of data and other information sources. Take the learning of attribute-based access control (ABAC) policies [57] as an example. An evolutionary approach was proposed in [58] to learn a single rule per group of decision examples based on a divide-and-conquer algorithm. [59] learned a Restricted Boltzmann Machine using logs in order to generate candidate rules. However, the learning process should be able to carefully balance between overfitting and safe policy generalization. Towards this end, [60] proposed a generic framework that incorporates available context information with logs to improve accuracy, while generating ABAC rules that can be expressed in propositional logs. Beyond learning security policies from scratch, how to quickly adapt the policies with limited logs is an interesting direction, where transfer learning can play an important role.

III. Networks for AI

In the current and future large-scale distributed AI problems we envision, data is private and not colocated with learners, computation is distributed, applications are real-time, and interactions occur between both human and AI agents. Mechanisms that enable distributed AI have the added benefit of AI-democratization, making it accessible to all rather than only large corporations. This section lays the foundation of new AI algorithms that are robust to immutable network and computing constraints, are adaptive to heterogeneity, and robust to failures. New network architectures and algorithms could be developed for AI, accounting for human-AI-network interactions, privacy, security, and fairness.

A. Network-Aware AI Operation

Distributed ML algorithms presently run on centrally partitioned datasets over high-performance, reliable data centers, connected by fast communication links. In the edge networks, edge nodes have limited computational resources with potential straggler problems, communication links are slow and unreliable, and data is imbalanced, highly heterogeneous, and comes with privacy constraints. These new challenges point to the need of scalable and network-aware distributed ML algorithms, accounting for computation, communication, and data constraints at the edge.

1) Communication-Efficient and Network-Aware Distributed Optimization: Traditional distributed ML algorithms assume reliable communication between a central aggregating server and worker nodes. In contrast, in edge networks, edge nodes are usually communication-constrained, resource-constrained, and with heterogeneous computing speeds. Therefore, it is important to achieve good performance comparable to the traditional methods while simultaneously reducing the communication overhead [61]–[63]. Although fundamental and practical work are still missing, there have been some studies in this direction. For example, [64] developed a federated learning algorithm, which can leverage the spatial and temporal correlations in the data to reduce communication costs, to improve the efficiency and accuracy of distributed mean estimation. Moreover, a novel server-based variance-reduced algorithm was proposed in [65], which simply uses the most recent update for each client, to achieve reduced convergence error. Besides the trade-off between the utility performance and the communication overhead, (extreme) mobility, privacy and security are other three important factors that need to be taken into consideration.

2) Scalable, Network-Aware Distributed Inference: Future edge networks tend to be distributed and with large scale. In the literature for dealing with such a network, scalability has received considerable attention. This requires us to address the challenges resulted by stragglers [66], non-linear or even more complex settings [67], heterogeneous constraints like security requirements [68]. An efficient parallel resource allocation method has been proposed in [69] that can not only learn a highly-accurate complex neural architectures, but also satisfy underlying systems constraints. Nonetheless, it is still open how to leverage network information to (i) design coding schemes that use partial computations by slow nodes to handle stragglers, (ii) make coded computation secure and adaptive to the heterogeneous, dynamic, and malicious nature of edge computing systems and resources, and (iii) reduce communication overheads and improve inference and energy efficiency at the edge.

3) Meta-Learning and Active Learning: Meta-learning significantly increases training efficiency by leveraging pre-trained models to perform only light-weight fine-tuning for new tasks [70], [71]. However, existing meta-learning algorithms cannot resolve the challenges due to constrained resources at the edge and the limited supervision. The latter is also the reason that active learning emerges as an another important learning idea. Active learning is a method-agnostic approach to extend meta-learning with active data selection at training time to yield improved performance. An active learning algorithm has been shown in [72] to achieve improved accuracy and reduced annotation cost in various settings with extreme class-imbalance.

B. Network Operation for Distributed AI-Applications

Networks have traditionally been designed for the purposes of communication, and the main functionality is to be reliable "bit pipelines". Yet, future networks, need to be re-engineered to better serve the new needs of distributed AI. Thus, an important question is how networks should adaptively allocate communication, computing, and storage resources to optimize information freshness, diversity, fidelity, etc., for distributed and diverse AI applications.

1) Network Operation for Managing AI-Side Uncertainty and Dynamic: AI services for processing edge networks can vary drastically in size, scale, and urgency. One of the crucial questions in the design of networks for AI is how to address the partial observations in the design of online algorithms for distributed AI processing. To partially resolve this question, an RL algorithm that achieves a sub-linear regret by considering latent (unobserved) variables in the case with short time horizon was developed in [73]. In [73], it is shown that how many episodes a general instance of LMDPs requires to approximate the optimal policy. Beyond this, it is also important to understand how to schedule and process dynamically arriving AI services with varying and typically unknown service requirements.

2) Network Operation for Managing Network-Side Uncertainty and Dynamics: In addition to the uncertainty from the AI-side, another important issue is to address the network-side uncertainty and dynamics, especially when the network size grows. For example, efficient computing resource allocation and scheduling algorithms were developed in [74] for minimizing training job completion time. Moreover, a rate allocation strategy was proposed in [75] to distribute data for distributed learners with provable performance guarantee under realistic network constraints. However, more comprehensive work for addressing the key issues, such as various sources of the network uncertainty, time scales of the network uncertainty, and the heterogeneity in the interaction and collaboration of network nodes, is still needed.

3) Unified, Distributed Network Operation for AI Applications: A key challenge here is that many distributed ML and AI jobs, including both training and inference, have a complex structure and are iterative in nature. For example, the training time of ML/AI models, if not managed carefully, could be orders of magnitude higher than traditional computing jobs. As a result, AI-aware network operation is necessary and important for achieving good performance. For example, [76] proposed a feasible distributed algorithm with convergence guarantees for resource allocation when the utility function is unknown in a distributed AI network. The proposed algorithm solves a distributed resource allocation lower-level problem, together with a user-specific quantity tuning upper-level problem.

C. Human, AI and Networks: Research at the Interface

A core function of the Internet is to connect people and the data they produce/ observe. Increasingly, AI mediates these interactions in ways ranging from AI content/product-recommendation engines to semi-automatic AI systems that help operators manage networks. Current AI systems have two serious limitations: 1) AI services do not dynamically scale capabilities in response to client/user network resources; 2) AI systems are black-boxes to humans and vice-versa. These issues can be addressed at the interface of humans, AI, and networks.

1) Human-AI Interface: To support highly interactive AI systems that learn on-the-fly, the AI models need to be both observable (experts understanding AI agents) and directable (AI agents revising themselves based on explicit commands and implicit intent of human partners). Towards this end, the mining and learning algorithms should be aware of the architectural capabilities of the networked system of interest (e.g., immersive sensing, unmanned aerial systems), while being cognizant of humans in the loop. Specifically, a right balance should be achieved between the sample rate, communication costs, decision and modeling needs, and data storage and processing requirements [77]. The representations and algorithms may also need to be adaptive, auto-tuned and resource-cognizant for applications at the network edge [78], [79]. In the meanwhile, new theory, methods, and applications should be developed for allowing people to provide explanations, feature relevance, and feedback to AI systems [80]–[82].

2) AI-Network Interface: To optimize the AI models such that it can scale seamlessly while reflecting current distributions and services, the network must provide information to AI systems to facilitate latency tradeoffs between servicing clients and updating data freshness [83], [84]. Simultaneously, AI agents, possibly in consultation with humans-in-the loop, must provide hints to the systems layers on near-term quality of service requirements (e.g., communication and data freshness requirements) across the different phases of the complex AI learning cycle. One promising solution to achieve this is to develop a secure, low-overhead, bi-direction, cross-layer monitoring and introspection interface to support a range of next generation AI technologies. A key requirement of this interface layer is that it must require minimal changes to the underlying physical and transport layers of the networking fabric to ensure compatibility with existing network protocols. This dictates the need to develop specialized AI-centric application layer protocols to support such an interface.

D. Security and Privacy in AI

In distributed learning in edge networks, information about the models being learned flows across the network. For example, in a federated learning algorithm, network users receive intermediate versions of a model and perform local updates to the model based on their own, possibly sensitive, data, before the result is forwarded to the parameter server. This opens the door to a host of attacks that aim at inferring users' private and sensitive data from the observed (intermediate) model. The last decade has witnessed the rise of a rich theory to deal with privacy threats. This theory is centered around a sound and rigorous definition for privacy, known as differential privacy (DP) [85]. Intuitively, a differentially private computation is one where no individual's data has significant influence on the outcome. However, there are several major challenges facing the design and implementation of differentially private AI algorithms in edge networks. Some of these challenges require developing new concepts, some require new algorithmic and networking techniques, and others require building secure and trusted execution environments (TEEs) to perform differentially private computations.

1) Handling the Absense of Trusted Curators: A fundamental challenge herein stems from whether a secure and trusted computationally capable curator (server) is available to perform differentially private computations on sensitive data collected from the network users. Depending on the state of the network and the users' privacy preferences and trust model, such a curator might not always be available. The decentralized (a.k.a., local) model of differential privacy [86], [87] offers a way to address this challenge. This model does not require a trusted curator as it is up to the users to locally apply a privatizing mechanism to their raw data before sending the result across the network. However, compared to the centralized model of differential privacy (that assumes the presence of a trusted curator), this model has several fundamental limitations on the attainable accuracy and communication efficiency [87], [88]. Recently, it was shown that such limitations can be circumvented if the user's identity is masked via a shuffling mechanism. Particularly, [89] initiated the analytic study of the shuffled model as an augmentation of the standard local differentially private model, by leveraging an anonymous channel to collect, premute and forward messages to data collectors. [90] showed that random shuffling of data reports can be utilized to achieve strong central differential privacy guarantees from weak privacy guarantees in the local model. To eventually leverage the power of the shuffling mechanism in distributed AI systems, one critical question here is how to use existing network infrastructure to provide the functionality of the shuffling mechanism.

2) Protecting Data-in-Computing via Trusted Execution Environments: While data-intransit and data-in-storage can be protected through well-developed cryptographic techniques (e.g., AES), protecting data-in-computing still faces enormous challenges. In addition to differential privacy, there have been explorations of other fundamental techniques such as multi-party computation, homomorphic encryption, and recently confidential computing via Trusted Execution Environments (TEEs), e.g., ARM TrustZone, Intel SGX, and AMD SEV, to protect data-in-computing. Therefore, it is of great interests to develop practical and efficient mechanisms based on those fundamental techniques so as to secure data-in-computing in both the edge and the cloud. Take the application of connected and autonomous vehicles (CAVs) as an example. One key security challenge therein is how to efficiently authenticate a vehicle in the ad-hoc CAV network and ensure its accountability and non-repudiation. To address this challenge, certificate-based authentication has been leveraged in [91], [92], which however requires frequent asymmetric key encryption and decryption and hence is not practical. Recently, a new Vehicle-to-Vehicle (V2V) protocal was proposed in [93] based on the TEE of in-vehicle processors, where many security demands can be satisfied naturally by using the Daily Symmetric keys protected by TEE. Considering the promising advantages of TEE, how to provide integrity validation of the run-time execution states of AI models by using TEE is also an interesting direction.

IV. Applications

The research directions that we discussed in Section II and Section III are closely related to many practical applications. In this section, we provide three critical use cases with existing or promising future applications.

A. Ubiquitous and Immersive Sensing and Networking in 6G+ Systems

The modern wireless-network platforms [94], [95] tend to serve various types of end-users, including sensors, human beings, distributed antennas, edge compute clouds, and base stations associated with automated devices, etc. To provide AI-driven sensing and networking in these and future platforms for 6G+ systems, we encounter a pressing need to develop new AI methods. This type of ubiquitous and immersive sensing and networking can immediately benefit from new results in most of the above-mentioned directions, e.g., advancing deep-learning-based communication algorithms (see Section II-A3), providing systematic understanding of protocol specifications (see Section II-D1), performing network-aware distributed optimization (see Section III-A1), and considering both network-side and AI-side uncertainties (see Section III-B1 and Section III-B2). For example, to achieve better inference performance and data processing in autonomous driving, a key step is to design new efficient bandit and federated learning algorithms for selecting beams, fusing multi-modal data from different sources (e.g., vehicle-mounted LiDAR, camera images and GPS units), and collaborating with edge compute clouds [96], [97]. These rely on the advances for problems in multiple directions, e.g., designing sample-efficient AI algorithms in Section II-B1 and learning from incomplete data set in Section II-B3.

B. Connecting Machines and Humans Under Extreme Mobility

Due to the involvement of moving edge nodes, e.g., human-related devices, and sensors on either the terrestrial or aerial vehicles, (extreme) mobility would become an essential factor that needs to be handled. For example, for the terrestrial vehicular and unmanned aerial system (UAS) mounted systems, mobile elements complement resource provisioning by fixed towers. In a system with such (extreme) mobility, as we discussed in Section II-A1 and Section II-A3, it would be helpful and important to leverage deep learning and physical knowledge to develop new communication algorithms. Moreover, it is also important to extend the multi-agent learning theories and algorithms for handling the mobility (see Section II-C1), to consider mobility when leveraging the shared data from mobile edge devices (see Section II-C3), and to understand fundamentally how the mobility affects when dealing with uncertainties (see Section III-B1 and Section III-B2). For example, [98] has shown the importance of 5G mmWave mobility support, e.g., ML-based policy reconfiguration for performance-centric mobility management, in practical problems on radio access control and mobility management. In addition, it is of great importance to take into consideration the mobility when designing reinforcement learning algorithms for frequency selection in wireless sensor deployments [99].

C. End-to-End Programmable and Virtualized 6G+ Cellular Networks

Softwarization, virtualization, interoperability, and separating the data plane from control functionalities have been widely considered in current cellular networks [100]–[103]. These principles would also be important for 6G+ cellular networks, e.g., to accommodate diverse network services, tenants, traffic on-demand, interoperability of different components, and full programmatic control of the network fabric. However, these features require fundamental improvements in re-engineering the physical environment (see Section II-A2), designing networks as a multi-agent system (see Section II-C1), systematic analysis of network protocol implementations (see Section II-D2), and a unified distributed network operation for AI applications (see Section III-B3). To facilitate research in these directions, an OpenRAN Gym platform has been provided in [104]. This toolbox can be used for end-to-end design, data collection, testing workflows for intelligent control, and testing applications on a softwarized RAN.

V. Conclusions

In this paper, we set the stage for leveraging the synergies between AI and Networking research to design future edge networks and distributed intelligence. To resolve such a grand challenge, we propose a divide and conquer approach that involves research across meta challenges in developing AI for Networks and Networks for AI. To further the research in AI for networks, we proposed four important research directions (each with several more specific open problems) on designing AI for re-engineering the physical layer of the wireless edge network, single-agent and multi-agent learning methods for resource allocation and control, and security against adversaries in edge networks. Similarly, to further research in networks for AI, we also proposed four important research directions on leveraging the feedback (such as dynamic conditions and constraints) from networks, addressing uncertainties from networks, AI and human, and protect the privacy. This paper is a ”call to arms” for developing new AI theories and solutions to spur the development of intelligent XG edge networks and new adaptive network designs for bringing to life new and exciting distributed intelligence based services.

Acknowledgment

This work was supported in part by an award from the National Science Foundation CNS-2112471.

References

[1] F. de Avila Belbute-Peres, K. Smith, K. Allen, J. Tenenbaum, and J. Z. Kolter, “End-to-end differentiable physics for learning and control,” *Advances in neural information processing systems*, vol. 31, 2018.

[2] D. Mrowca, C. Zhuang, E. Wang, N. Haber, L. F. Fei-Fei, J. Tenenbaum, and D. L. Yamins, “Flexible neural representation for physics prediction,” *Advances in neural information processing systems*, vol. 31, 2018.

[3] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, “Parameterized neural networks for high-energy physics,” *The European Physical Journal C*, vol. 76, no. 5, pp. 1–7, 2016.

[4] G. Carleo and M. Troyer, “Solving the quantum many-body problem with artificial neural networks,” *Science*, vol. 355, no. 6325, pp. 602–606, 2017.

[5] A. Bakshi, Y. Mao, K. Srinivasan, and S. Parthasarathy, “Fast and efficient cross band channel prediction using machine learning,” in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–16.

[6] M. Hashemi, A. Sabharwal, C. E. Koksal, and N. B. Shroff, “Efficient beam alignment in millimeter wave systems using contextual bandits,” in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2393–2401.

[7] Y. Deng, X. Zhou, A. Ghosh, A. Gupta, and N. B. Shroff, “Interference constrained beam alignment for time-varying channels via kernelized bandits,” *arXiv preprint arXiv:2207.00908*, 2022.

[8] B. Chen, G. K. Tummala, Y. Qiao, and K. Srinivasan, “In-band wireless cut-through: Is it possible?” in *Proceedings of the 1st ACM Workshop on Hot Topics in Wireless*, 2014, pp. 1–6.

[9] L. Chen, F. Wu, J. Xu, K. Srinivasan, and N. Shroff, “Bipass: Enabling end-to-end full duplex,” in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 2017, pp. 114–126.

[10] H. Yang, P. Qiu, J. Liu, and A. Yener, “Over-the-air federated learning with joint adaptive computation and power control,” *arXiv preprint arXiv:2205.05867*, 2022.

[11] J. Mao, H. Yang, P. Qiu, J. Liu, and A. Yener, “Charles: Channel-quality-adaptive over-the-air federated learning over wireless networks,” *arXiv preprint arXiv:2205.09330*, 2022.

[12] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near shannon limit error-correcting coding and decoding: Turbo-codes. 1,” in *Proceedings of ICC’93-IEEE International Conference on Communications*, vol. 2. IEEE, 1993, pp. 1064–1070.

[13] R. Gallager, “Low-density parity-check codes,” *IRE Transactions on information theory*, vol. 8, no. 1, pp. 21–28, 1962.

[14] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Transactions on information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[15] H. Kim, Y. Jiang, S. Kannan, S. Oh, and P. Viswanath, “Deepcode: Feedback codes via deep learning,” *Advances in neural information processing systems*, vol. 31, 2018.

[16] Y. Jiang, H. Kim, H. Asnani, S. Kannan, S. Oh, and P. Viswanath, “Learn codes: Inventing low-latency codes via recurrent neural networks,” *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 207–216, 2020.

[17] A. V. Makkuva, X. Liu, M. V. Jamali, H. Mahdavifar, S. Oh, and P. Viswanath, “Ko codes: inventing nonlinear encoding and decoding for reliable wireless communication via deep-learning,” in *International Conference on Machine Learning*. PMLR, 2021, pp. 7368–7378.

[18] I. Tariq, R. Sen, T. Novlan, S. Akoum, M. Majmundar, G. de Veciana, and S. Shakkottai, “Auto-tuning for cellular scheduling through bandit-learning and low-dimensional clustering,” *IEEE/ACM Transactions on Networking*, vol. 29, no. 5, pp. 1933–1947, 2021.

[19] S. Krishnasamy, R. Sen, R. Johari, and S. Shakkottai, “Regret of queueing bandits,” *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[20] M. Shi, X. Lin, and L. Jiao, “Power-of-2-arms for bandit learning with switching costs,” in *Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 2022, pp. 131–140.

[21] A. Modi, N. Jiang, A. Tewari, and S. Singh, “Sample complexity of reinforcement learning using linearly combined model ensembles,” in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 2010–2020.

[22] S. Buccapatnam, F. Liu, A. Eryilmaz, and N. B. Shroff, “Reward maximization under uncertainty: Leveraging side-observations on networks,” *arXiv preprint arXiv:1704.07943*, 2017.

[23] F. Liu, J. Lee, and N. Shroff, “A change-detection based framework for piecewise-stationary multi-armed bandit problem,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[24] D. Roy, Y. Li, T. Jian, P. Tian, K. R. Chowdhury, and S. Ioannidis, “Multi-modality sensing and data fusion for multi-vehicle detection,” *IEEE Transactions on Multimedia*, 2022.

[25] H. Wei, X. Liu, and L. Ying, “A provably-efficient model-free algorithm for constrained markov decision processes,” *arXiv preprint arXiv:2106.01577*, 2021.

[26] S. Cayci, Y. Zheng, and A. Eryilmaz, “A lyapunov-based methodology for constrained optimization with bandit feedback,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, 2022, pp. 3716–3723.

[27] H. Wei, X. Liu, and L. Ying, “Triple-q: A model-free algorithm for constrained reinforcement learning with sublinear regret and zero constraint violation,” in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 3274–3307.

[28] A. Wachi, Y. Sui, Y. Yue, and M. Ono, “Safe exploration and optimization of constrained mdps using gaussian processes,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[29] G. Qu, A. Wierman, and N. Li, “Scalable reinforcement learning for multiagent networked systems,” *Operations Research*, 2022.

[30] T. Xu and Y. Liang, “Provably efficient offline reinforcement learning with trajectory-wise reward,” *arXiv preprint arXiv:2206.06426*, 2022.

[31] M. Shi, X. Lin, and S. Fahmy, “Competitive online convex optimization with switching costs and ramp constraints,” *IEEE/ACM Transactions on Networking*, vol. 29, no. 2, pp. 876–889, 2021.

[32] J. Feng, Y. Shi, G. Qu, S. H. Low, A. Anandkumar, and A. Wierman, “Stability constrained reinforcement learning for real-time voltage control in distribution systems,” *arXiv preprint arXiv:2209.07669*, 2022.

[33] S. Amani, C. Thrampoulidis, and L. Yang, “Safe reinforcement learning with linear function approximation,” in *International Conference on Machine Learning*. PMLR, 2021, pp. 243–253.

[34] M. Shi, X. Lin, and L. Jiao, “Combining regularization with look-ahead for competitive online convex optimization,” in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 2021, pp. 1–10.

[35] D. Vial, A. Parulekar, S. Shakkottai, and R. Srikant, “Improved algorithms for misspecified linear markov decision processes,” in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 4723–4746.

[36] Z. Guan, T. Xu, and Y. Liang, “Per-etd: A polynomially efficient emphatic temporal difference learning method,” *ICLR 2022*, 2022.

[37] R. S. Sutton, A. R. Mahmood, and M. White, “An emphatic approach to the problem of off-policy temporal-difference learning,” *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2603–2631, 2016.

[38] V. Syrgkanis, A. Agarwal, H. Luo, and R. E. Schapire, “Fast convergence of regularized learning in games,” *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[39] H. Xiong, L. Zhao, Y. Liang, and W. Zhang, “Finite-time analysis for double q-learning,” *Advances in neural information processing systems*, vol. 33, pp. 16 628–16 638, 2020.

[40] T. Xu, Z. Wang, and Y. Liang, “Improving sample complexity bounds for (natural) actor-critic algorithms,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 4358–4369, 2020.

[41] T. Xu, S. Zou, and Y. Liang, “Two time-scale off-policy td learning: Non-asymptotic analysis over markovian samples,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[42] S. Zou, T. Xu, and Y. Liang, "Finite-sample analysis for sarsa with linear function approximation," *Advances in neural information processing systems*, vol. 32, 2019.

[43] F. Hairi, J. Liu, and S. Lu, "Finite-time convergence and sample complexity of multi-agent actor-critic reinforcement learning with average reward," in *International Conference on Learning Representations*, 2021.

[44] C. Tekin and M. Liu, "Performance and convergence of multi-user online learning," in *International Conference on Game Theory for Networks*. Springer, 2011, pp. 321–336.

[45] ——, "Online learning in decentralized multi-user spectrum access with synchronized explorations," in *MILCOM 2012-2012 IEEE Military Communications Conference*. IEEE, 2012, pp. 1–6.

[46] Y. Li, T. Si Salem, G. Neglia, and S. Ioannidis, "Online caching networks with adversarial guarantees," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 5, no. 3, pp. 1–39, 2021.

[47] Z. Wang, T. Jian, A. Masoomi, S. Ioannidis, and J. Dy, "Revisiting hilbert-schmidt information bottleneck for adversarial robustness," *Advances in Neural Information Processing Systems*, vol. 34, pp. 586–597, 2021.

[48] H. Kim, J. Lee, E. Lee, and Y. Kim, "Touching the untouchables: Dynamic security analysis of the lte control plane," in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 1153–1168.

[49] H. Lee, I. Karim, N. Li, and E. Bertino, "Vwanalyzer: A systematic security analysis framework for the voice over wifi protocol," in *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security*, 2022, pp. 182–195.

[50] C. Peng, C.-y. Li, G.-H. Tu, S. Lu, and L. Zhang, "Mobile data charging: new attacks and countermeasures," in *Proceedings of the 2012 ACM conference on Computer and communications security*, 2012, pp. 195–204.

[51] C.-Y. Li, G.-H. Tu, C. Peng, Z. Yuan, Y. Li, S. Lu, and X. Wang, "Insecurity of voice solution volte in lte mobile networks," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, pp. 316–327.

[52] G.-H. Tu, C.-Y. Li, C. Peng, Y. Li, and S. Lu, "New security threats caused by ims-based sms service in 4g lte networks," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 1118–1130.

[53] M. Chlosta, D. Rupprecht, T. Holz, and C. Pöpper, "Lte security disabled: misconfiguration in commercial networks," in *Proceedings of the 12th conference on security and privacy in wireless and mobile networks*, 2019, pp. 261–266.

[54] S. Hussain, O. Chowdhury, S. Mehnaz, and E. Bertino, "Lteinspector: A systematic approach for adversarial testing of 4g lte," in *Network and Distributed Systems Security (NDSS) Symposium 2018*, 2018.

[55] D. Maier, L. Seidel, and S. Park, "Basesafe: Baseband sanitized fuzzing through emulation," in *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2020, pp. 122–132.

[56] S. R. Hussain, I. Karim, A. A. Ishtiaq, O. Chowdhury, and E. Bertino, "Noncompliance as deviant behavior: An automated black-box noncompliance checker for 4g lte cellular devices," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 1082–1099.

[57] E. Bertino, G. Ghinita, A. Kamra *et al.*, "Access control for databases: Concepts and systems," *Foundations and Trends® in Databases*, vol. 3, no. 1–2, pp. 1–148, 2011.

[58] E. Medvet, A. Bartoli, B. Carminati, and E. Ferrari, "Evolutionary inference of attribute-based access control policies," in *International Conference on Evolutionary Multi-Criterion Optimization*. Springer, 2015, pp. 351–365.

[59] D. Mocanu, F. Turkmen, A. Liotta *et al.*, "Towards abac policy mining from logs with deep learning," in *Proceedings of the 18th International Multiconference, ser. Intelligent Systems*, 2015.

[60] A. Abu Jabal, E. Bertino, J. Lobo, M. Law, A. Russo, S. Calo, and D. Verma, "Polisma-a framework for learning attribute-based access control policies," in *European Symposium on Research in Computer Security*. Springer, 2020, pp. 523–544.

[61] J. Wang and G. Joshi, "Adaptive communication strategies for best error-runtime trade-offs in communication-efficient distributed sgd," in *Proceedings of the SysML Conference*, 2019.

[62] S. Dutta, G. Joshi, S. Ghosh, P. Dube, and P. Nagpurkar, "Slow and stale gradients can win the race: Error-runtime trade-offs in distributed sgd," in *International conference on artificial intelligence and statistics*. PMLR, 2018, pp. 803–812.

[63] J. Wang, H. Liang, and G. Joshi, "Overlap local-sgd: An algorithmic approach to hide communication delays in distributed sgd," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8871–8875.

[64] D. Jhunjhunwala, A. Mallick, A. Gadhikar, S. Kadhe, and G. Joshi, "Leveraging spatial and temporal correlations in sparsified mean estimation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 14 280–14 292, 2021.

[65] D. Jhunjhunwala, P. SHARMA, A. Nagarkatti, and G. Joshi, "Fedvarp: Tackling the variance due to partial client participation in federated learning," in *The 38th Conference on Uncertainty in Artificial Intelligence*, 2022.

[66] J. Dean and L. A. Barroso, "The tail at scale," *Communications of the ACM*, vol. 56, no. 2, pp. 74–80, 2013.

[67] A. Mallick, M. Chaudhari, and G. Joshi, "Fast and efficient distributed matrix-vector multiplication using rateless fountain codes," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8192–8196.

[68] Y. Keshtkarjahromi, R. Bitar, V. Dasari, S. El Rouayheb, and H. Seferoglu, "Secure coded cooperative computation at the heterogeneous edge against byzantine attacks," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[69] E. Jääsaari, M. Ma, A. Talwalkar, and T. Chen, "Sonar: Joint architecture and system optimization search," *arXiv preprint arXiv:2208.12218*, 2022.

[70] K. Ji, J. D. Lee, Y. Liang, and H. V. Poor, "Convergence of meta-learning with task-specific adaptation over partial parameters," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11 490–11 500, 2020.

[71] K. Ji, J. Yang, and Y. Liang, "Bilevel optimization: Convergence analysis and enhanced design," in *International Conference on Machine Learning*. PMLR, 2021, pp. 4882–4892.

[72] J. Zhang, J. Katz-Samuels, and R. Nowak, "Galaxy: Graph-based active learning at the extreme," *arXiv preprint arXiv:2202.01402*, 2022.

[73] J. Kwon, Y. Efroni, C. Caramanis, and S. Mannor, "Rl for latent mdps: Regret guarantees and a lower bound," *Advances in Neural Information Processing Systems*, vol. 34, pp. 24 523–24 534, 2021.

[74] H. Yang, M. Fang, and J. Liu, "Achieving linear speedup with partial worker participation in non-iid federated learning," *arXiv preprint arXiv:2101.11203*, 2021.

[75] Y. Liu, Y. Li, L. Su, E. Yeh, and S. Ioannidis, "Experimental design networks: A paradigm for serving heterogeneous learners under networking constraints," *arXiv preprint arXiv:2201.04830*, 2022.

[76] K. Ji, M. Liu, Y. Liang, and L. Ying, "Will bilevel optimizers benefit from loops," *arXiv preprint arXiv:2205.14224*, 2022.

[77] M. Kim and P. Smaragdis, "Bitwise neural networks," *arXiv preprint arXiv:1601.06071*, 2016.

[78] N. Sedaghati, T. Mu, L.-N. Pouchet, S. Parthasarathy, and P. Sadayappan, "Automatic selection of sparse matrix representation on gpus," in *Proceedings of the 29th ACM on International Conference on Supercomputing*, 2015, pp. 99–108.

[79] A. Ashari, N. Sedaghati, J. Eisenlohr, S. Parthasarath, and P. Sadayappan, "Fast sparse matrix-vector multiplication on gpus for graph applications," in *SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE, 2014, pp. 781–792.

[80] T. M. Mitchell, R. M. Keller, and S. T. Kedar-Cabelli, "Explanation-based generalization: A unifying view," *Machine learning*, vol. 1, no. 1, pp. 47–80, 1986.

[81] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behavioral and brain sciences*, vol. 40, 2017.

[82] H. Raghavan, O. Madani, and R. Jones, "Active learning with feedback on features and instances," *The Journal of Machine Learning Research*, vol. 7, pp. 1655–1686, 2006.

[83] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, 2017.

[84] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Age-optimal sampling and transmission scheduling in multi-source systems," in

Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2019, pp. 121–130.

[85] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*. Springer, 2006, pp. 265–284.

[86] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 2013, pp. 429–438.

[87] R. Bassily and A. Smith, "Local, private, efficient protocols for succinct histograms," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 2015, pp. 127–135.

[88] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[89] A. Cheu, A. Smith, J. Ullman, D. Zeber, and M. Zhilyaev, "Distributed differential privacy via shuffling," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2019, pp. 375–403.

[90] Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, K. Talwar, and A. Thakurta, "Amplification by shuffling: From local to central differential privacy via anonymity," in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2019, pp. 2468–2479.

[91] N. Bißmeyer, H. Stübing, E. Schoch, S. Götz, J. P. Stotz, and B. Lonc, "A generic public key infrastructure for securing car-to-x communication," in *18th ITS World Congress, Orlando, USA*, vol. 14, 2011.

[92] S. Chan-Edmiston, S. Fischer, S. Sloan, M. Wong *et al.*, "Intelligent transportation systems (its) joint program office: strategic plan 2020–2025," United States. Department of Transportation. Intelligent Transportation . . . , Tech. Rep., 2020.

[93] M. K. Jangid and Z. Lin, "Towards a tee-based v2v protocol for connected and autonomous vehicles," in *Workshop on Automotive and Autonomous Vehicle Security (AutoSec)*, vol. 2022, 2022, p. 27.

[94] J. Breen, A. Buffmire, J. Duerig, K. Dutt, E. Eide, M. Hibler, D. Johnson, S. K. Kasera, E. Lewis, D. Maas *et al.*, "Powder: Platform for open wireless data-driven experimental research," in *Proceedings of the 14th International Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization*, 2020, pp. 17–24.

[95] D. Raychaudhuri, I. Seskar, G. Zussman, T. Korakis, D. Kilper, T. Chen, J. Kolodziejski, M. Sherman, Z. Kostic, X. Gu *et al.*, "Challenge: Cosmos: A city-scale programmable testbed for experimentation with advanced wireless," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–13.

[96] B. Salehi, G. Reus-Muns, D. Roy, Z. Wang, T. Jian, J. Dy, S. Ioannidis, and K. Chowdhury, "Deep learning on multimodal sensor data at the wireless edge for vehicular network," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7639–7655, 2022.

[97] B. Salehi, J. Gu, D. Roy, and K. Chowdhury, "Flash: Federated learning for automated selection of high-band mmwave sectors," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 1719–1728.

[98] Q. Li and C. Peng, "Reconfiguring cell selection in 4g/5g networks," in *2021 IEEE 29th International Conference on Network Protocols (ICNP)*. IEEE, 2021, pp. 1–11.

[99] J. P. Bello, C. Silva, O. Nov, R. L. Dubois, A. Arora, J. Salamon, C. Mydlarz, and H. Doraiswamy, "Sonyc: A system for monitoring, analyzing, and mitigating urban noise pollution," *Communications of the ACM*, vol. 62, no. 2, pp. 68–77, 2019.

[100] T. Melodia, S. Basagni, K. R. Chowdhury, A. Gosain, M. Polese, P. Johari, and L. Bonati, "Colosseum, the world's largest wireless network emulator," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021, pp. 860–861.

[101] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "A survey on software-defined networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 27–51, 2014.

[102] Q. M. Qadir, T. A. Rashid, N. K. Al-Salihi, B. Ismael, A. A. Kist, and Z. Zhang, "Low power wide area networks: A survey of enabling technologies, applications and interoperability needs," *IEEE Access*, vol. 6, pp. 77 454–77 473, 2018.

[103] L. Cao, P. Sharma, S. Fahmy, and V. Saxena, "{ENVI}: elastic resource flexing for network function virtualization," in *9th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 17)*, 2017.

[104] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Colo-ran: Developing machine learning-based xapps for open ran closed-loop control on programmable experimental platforms," *IEEE Transactions on Mobile Computing*, 2022.