Adversarial Attacks Targeting Point-to-Point Wireless Networks

Ahmad Ghasemi
ECE department
University of Massachusetts Amherst
Amherst, USA
aghasemi@umass.edu

Majid Moradikia
Department of Data Science
Worcester Polytechnic Institute (WPI)
Worcester, MA, USA
mmoradikia@wpi.edu

Seyed (Reza) Zekavat
Department of Data Science
Worcester Polytechnic Institute (WPI)
Worcester, MA, USA
rezaz@wpi.edu

Hossein Pishro-Nik
ECE department
University of Massachusetts Amherst
Amherst, USA
pishro@engin.umass.edu

2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring) | 979-8-3503-8741-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/VTC2024-SPRING62846.2024.10683209

***Abstract*—This paper introduces a novel adversarial attack targeting Graph Neural Network (GNN)-based radio resource management in point-to-point networks. Our proposed attack, executed during the test phase, manipulates the system's input by exploiting specific constraints. Formulated as an optimization problem, the attack aims to maximize resource stealing, thereby degrading the quality of communication. We assess the attack's efficacy with respect to the number of users, signal-to-noise ratio, and the adversary's power budget. The results demonstrate that our proposed attack approaches the performance of an established upper-bound adversarial benchmark while maintaining lower complexity, highlighting its effectiveness and potential for real-world applicability.**

***Index Terms*—graph neural network, adversarial attack, P2P wireless networks**

I. Introduction

In the realm of wireless communications, Machine Learning (ML) and its subset, deep learning (DL), have been transformative, addressing complex tasks like radio resource management [1], [2], beam prediction [3], and channel estimation [4]. Despite their advancements, DL algorithms struggle with generalization and scalability, necessitating substantial data and diminishing in effectiveness with larger problem sizes. To mitigate these issues, Graph Neural Networks (GNNs) have been introduced, combining graph theory with DL, and have shown success in diverse fields such as Computer Vision [5], [6] and Natural Language Processing [7], [8].

GNNs have recently been applied to wireless communications [1], [2], [9]–[15], but they share a common vulnerability with other ML algorithms to adversarial attacks during training or testing [16], [17]. Unlike jamming or spoofing attacks, adversarial attacks subtly manipulate DL inputs to induce errors, presenting a significant risk to P2P wireless communication systems like device-to-device, machine-to-machine, and vehicle-to-vehicle communications [18].

These systems, integral to modern wireless networks and crucial in various sectors, including IoT [19] and 5G mobile communications [20], are vulnerable to such attacks. Adversarial attacks can lead to degraded performance and increased risks in IoT networks [21], affect communication quality in vehicular ad hoc networks [22].

Despite the significance of these threats, research on adversarial attacks against GNN-based P2P wireless communications remains limited. This paper aims to address this gap by exploring practical adversarial attack strategies on GNN-based P2P systems. Specifically, we introduce a novel adversarial attack targeting the vertices of a trained GNN model during the testing phase. The design of this attack adheres to two constraints: 1) *Channel-Bounded Constraint*: This limits the adversary to a certain number of simultaneous perturbations. 2) *Min-Detectable Constraint*: The adversary aims to perturb information in a way that minimizes detection likelihood by the system.

Considering these constraints, we propose a new optimization problem where the adversary targets the channel information of a subset of active pairs in the network. The objective of this approach is to minimize the total Quality of Communication (QoC) within the network, defined as a weighted sum rate, by maximizing the QoC of the targeted subset of users. Moreover, we introduce a heuristic algorithm to solve this optimization problem. The results demonstrate the effectiveness of the proposed adversarial attack, which succeeds to steal almost all resources by perturbing merely half of the available channels.

Related Works

I) Graphs and GNNs in Wireless Communications

Graphs and GNNs play a crucial role in wireless communications due to their ability to efficiently utilize domain knowledge, specifically the graph structure. Previous research [1], [2], [9], [10], [12] extensively employs GNNs for wireless communication problem-solving. For instance, Transmitter-Receiver (TX-RX) channels and channel correlations are modeled as vertices and edges, respectively, in [1], [2] to

979-8-3503-8741-4/24/$31.00 ©2024 IEEE

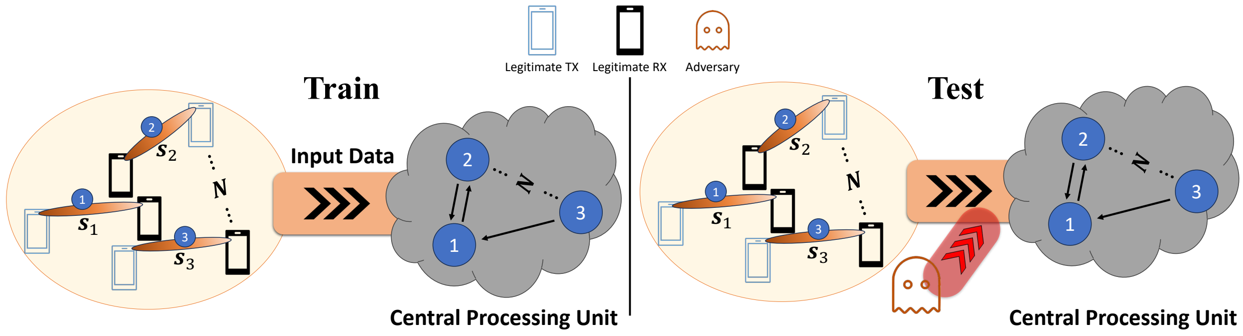

Fig. 1: System model illustration: (left) Training phase, (right) Testing phase

optimize antenna allocations. Additionally, GNNs are utilized in [9], [10] for centralized modeling of P2P communications, representing transceiver pairs and inter-user interference as vertices and edges, respectively. Another application is seen in [12], where GNNs are employed for power control in cellular systems, determining an optimal power allocation strategy based on estimated channel matrices.

II) Adversarial Attacks in ML/DL-based Wireless Communications

The realm of adversarial attacks on Machine Learning (ML)/Deep Learning (DL)-based wireless communications is explored in literature [16], [23]. These attacks facilitate signal mis-classification, as demonstrated in [23], where a small perturbation is added to test data during the testing phase, resulting in mis-classification at the receiver. In [16], different adversarial attacks against an autoencoder communication system are investigated, showing the destructive potential even when the adversary lacks perfect knowledge of the DL model or synchronization with the transmitter. Further, adversarial attacks against power allocation scenarios are considered in [17], where a Deep Neural Network (DNN) allocates transmit power to orthogonal subcarriers. The adversary perturbs input data to the DNN, affecting user sum rates by perturbing pilot signals or transmitting perturbed channel estimations to the base station.

A. Organization and Notation

The remainder of the paper is organized as follows: Section II presents the system model and problem definition. The proposed adversarial attacks are introduced in Section III. The proposed approaches are evaluated in Section IV. Finally, Section V concludes the paper.

Notation: In this paper, vectors are denoted by small bold-italic face letters $\mathbf{a}$, and capital bold-italic face letters $\mathbf{A}$ represent matrices. $\mathcal{A}$ is a set, and a is a scalar. The i^{th} element and the number of elements of set $\mathcal{A}$ or the cardinality of this set are denoted by $\mathcal{A}[i]$ and $|\mathcal{A}|$, respectively. $|a|$ and $\measuredangle a$ represent the magnitude and phase of the complex number a. Two new element-wise operators for vectors are defined as follows: $|\dot{\mathbf{a}}| \triangleq [|a_0|, |a_1|, \ldots, |a_{N-1}|]^{\mathrm{T}}$ and $\dot{\measuredangle}\mathbf{a} \triangleq [\measuredangle a_0, \measuredangle a_1, \ldots, \measuredangle a_{N-1}]^{\mathrm{T}}$, where $a_i, \forall\ i = 0, \ldots, N-1$, are the elements of the vector $\mathbf{a}$. The transpose and Hermitian (conjugate transpose) of a matrix/vector are denoted by $(.)^{\mathrm{T}}$ and $(.)^{\dagger}$, respectively. $\|\cdot\|_2$ denotes the l_2-norm of a vector. $\mathbb{D}^{l\times l}$ and $\mathbb{C}^{m\times n}$ represent a diagonal matrix of dimension $l \times l$ and a complex matrix of dimension $m \times n$. The n^{th} diagonal element of a diagonal matrix $\mathbb{D}$ is denoted by $\mathbb{D}_n$. $\mathbb{R}$ denotes the set of all real numbers. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. $\mathbf{0}_N$ and $\mathbf{1}_N$ are the N-dimensional all-zeros and all-ones vectors, respectively. We use $\mathcal{CN}(\mu, \sigma^2)$ to denote a circularly symmetric complex Gaussian random vector with mean μ and variance σ^2. Finally, $P(\cdot)$, $(\cdot)^*$, and $\mathbb{E}(\cdot)$ denote probability, optimum value, and expectation, respectively.

II. System Model and Problem Definition

This study examines a multi-user multi-input single-output (MISO) wireless network consisting of N active transceiver pairs, denoted by $\mathcal{N} = \{1, 2, \ldots, N\}$. Each transmitter (TX) is equipped with N_t antenna elements, while receivers (RX) have a single antenna (see Fig. 1 left).

Consider $\{s_n\}_{n=1}^{N}$ as the unit-norm signals transmitted from the n^{th} TX to the n^{th} RX. Define the precoding matrix $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N]^{\mathrm{T}} \in \mathbb{C}^{N\times N_t}$, where $\{\mathbf{q}_n\}_{n=1}^{N}$ represents the precoder for the n^{th} transmitter. The estimated signal received at the n^{th} RX is given by

$$y_n = \mathbf{h}_{n,n}^{\dagger}\mathbf{q}_n s_n + \sum_{\substack{i=1\\ i\neq n}}^{N} \mathbf{h}_{i,n}^{\dagger}\mathbf{q}_i s_i + n_n, \tag{1}$$

where $\mathbf{h}_{i,n} \in \mathbb{C}^{N_t}$ denotes the channel vector from the i^{th} TX to the n^{th} RX, and $n_n \sim \mathcal{CN}(0, \sigma_n^2)$ is the Additive White Gaussian Noise (AWGN) at the n^{th} RX.

Furthermore, the channel characteristics are encapsulated within a channel tensor $\mathbf{H} \in \mathbb{C}^{|\mathcal{V}|\times|\mathcal{V}|\times N_t}$. The tensor elements $\mathbf{H}_{i,n,:} = \mathbf{h}_{i,n} \in \mathbb{C}^{N_t}$, for $\{i, n\} \in \mathcal{N}$, distinguish between desired and interference channels of transceiver pairs through diagonal and off-diagonal elements, respectively. This channel tensor is accessible to both the central processing unit (CPU) and potential adversaries. The CPU is tasked with constructing and continually updating the Deep Learning (DL) model.

A. Graph Modeling of P2P Wireless Communications

As depicted in Fig. 1, we model the P2P wireless network under consideration as a *directed graph*. In this graph, each transceiver pair is represented by a vertex, specifically the n^{th} transceiver pair corresponding to the n^{th} vertex. The features of these vertices encapsulate the characteristics of the transceivers. A directed edge from vertex i to vertex j signifies interference from TX i to RX j, with the edge feature encompassing the properties of the respective interference channel. Significantly, interference only occurs if the distance between TX i and RX j falls below a predefined threshold T_d.

The graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is formally defined, where $\mathcal{V}$ and $\mathcal{E}$ represent the sets of vertices and edges, respectively. The vertex feature matrix $\mathbf{Z} \in \mathbb{C}^{|\mathcal{V}|\times(N_t+2)}$ is described by $\mathbf{Z}_{n,:} = [\mathbf{h}_{n,n}, w_n, \sigma_n^2]^{\mathrm{T}}$, where $|\mathcal{V}|$ is the size of set $\mathcal{V}$, and w_n denotes the weight of the n^{th} vertex or pair. The term $N_t + 2$ arises from N_t being the length of $\mathbf{h}_{n,n}$ and w_n and σ_n^2 being scalar values.

The adjacency feature tensor $\mathbf{A} \in \mathbb{C}^{|\mathcal{V}|\times|\mathcal{V}|\times N_t}$ is defined as follows:

$$\mathbf{A}_{i,n,:} = \begin{cases} \mathbf{0}_{N_t}, & \text{if } \{i,n\} \notin \mathcal{E}, \\ \mathbf{h}_{i,n}, & \text{otherwise}, \end{cases} \tag{2}$$

where $\mathbf{h}_{i,n} \in \mathbb{C}^{N_t}$ for $\{i,n\} \in \mathcal{N}$ represents the channel vector from TX i to RX n. Utilizing the variables $\mathbf{Z}$ and $\mathbf{A}$, we redefine the received signal at the n^{th} RX as:

$$y_n = \underbrace{\mathbf{Z}^{\dagger}_{n,1:N_t}\mathbf{q}_n s_n}_{\text{Desired Signal}} + \underbrace{\sum_{\substack{i=0 \\ i\neq n}}^{N} \mathbf{A}^{\dagger}_{i,n,:}\mathbf{q}_i s_i}_{\text{Interference}} + \underbrace{n_n}_{\text{Noise}}, \tag{3}$$

Consequently, the signal-to-interference-plus-noise ratio (SINR) for the n^{th}, $n \in \mathcal{N}$, RX of the corresponding vertex (transceiver pair) is:

$$\text{SINR}_n = \frac{|\mathbf{Z}^{\dagger}_{n,1:N_t}\mathbf{q}_n|^2}{\sum_{i=1, i\neq n}^{N} |\mathbf{A}^{\dagger}_{i,n,:}\mathbf{q}_i|^2 + \mathbf{Z}_{n,N_t+2}}, \tag{4}$$

where $\mathbf{Z}_{n,N_t+2}$, as per the definition of $\mathbf{Z}$, denotes the noise power.

To train the graph-based wireless communication system, we employ the GNN model proposed in [9], comprising three layers. The channel states $\{\mathbf{Z}^{\dagger}_{n,1:N_t}\}_{n=1}^{N}$ and users' weights $\{w_n\}_{n=1}^{N}$ serve as inputs to the GNN. The CPU leverages information from the transceivers to train a centralized GNN model of the system (refer to Fig. 1 right). The GNN outputs the beamforming vectors for users, aiming to minimize the loss function l_Θ at the final layer, as shown:

$$l_\Theta = -\mathbb{E}\left(\sum_{n=1}^{N} \mathbf{Z}_{n,N_t+1}\log_2(1 + \text{SINR}_n(\Theta))\right), \tag{5}$$

where $\mathbf{Z}_{n,N_t+1}$ indicates the user weights per definition of $\mathbf{Z}$, and $\text{SINR}_n(\Theta)$ is defined by:

$$\text{SINR}_n(\Theta) = \frac{|\mathbf{Z}^{\dagger}_{n,1:N_t}\mathbf{q}_n(\Theta)|^2}{\sum_{\substack{i=1 \\ i\neq n}}^{N} |\mathbf{A}^{\dagger}_{i,n,:}\mathbf{q}_i(\Theta)|^2 + \mathbf{Z}_{n,N_t+2}}, \tag{6}$$

as previously defined after Equation (4).

III. Adversarial Attack

In this section, we delineate the assumptions about the adversary relevant to this study.

Remark III.1. ***(Adversary's Assumptions):***

1) *The adversary operates as a white box, having access only to channel information.*
2) *It executes Evasion attacks.*
3) *Equipped with multiple antenna elements, the adversary can simultaneously target multiple channels, unlike with a single antenna which restricts it to one channel at a time.*
4) *It moderates perturbation power to diminish the probability of detection by the CPU and legitimate users.*
5) *All transceiver pairs fall within the adversary's transmission range [24], [25].*
6) *The adversary can eavesdrop and learn information from transceiver pairs.*
7) *It has the flexibility to select different channels and frequencies to disrupt at any given time.*
8) *As a reactive adversary, it engages in physical carrier sensing (as part of standards like 802.11) to discern if a channel is idle or busy.*
9) *It can transmit malicious messages using address spoofing techniques [24], [25].*

■

During the testing phase, the adversary aims to maximize the weighted sum-rate of a subset of transceiver pairs which are under its attack. This in turn degrades the network's performance QoC. The adversary perturbs the channel information $\mathbf{h}_{i,n}$ transmitted by each pair included in the subset to the CPU. From this point, we use the terms 'channel' and 'channel information' interchangeably to refer to the information relayed from the transceiver pairs to the CPU. To acquire this data, the adversary might:

1) Impersonate a fake CPU temporarily, tricking users into sending data and injecting malicious packets to extract necessary information [26].
2) Persistently monitor the data to learn about the transceiver pairs and their channel information.

Leveraging this acquired knowledge and considering network power constraints, the adversary seeks optimal $\widehat{\mathbf{h}}_{i,n}$ values to maximize the sum-rate of the selected subset of transceiver pairs. The corresponding optimization problem is formulated as:

$$\max_{\widehat{\mathbf{H}}} \quad \sum_{i\in\mathcal{A}} \mathbf{Z}_{i,N_t+1}\log_2(1 + \widehat{\text{SINR}}_i), \tag{7a}$$

$$\text{s.t.} \quad \|\mathbf{q}_n\|_2^2 \leq P_{\max}, \forall\ n \in \mathcal{N}, \tag{7b}$$

here, $\widehat{\text{SINR}}_i$ indicates the distorted SINR, with the perturbed channel information $\widehat{\mathbf{h}}_{i,n}$ substituted in (4). The constraint expresses the power budget limitation at transmitters.

Remark III.2. ***(Channel-bounded (B_c) Constraint):*** *This constraint limits the number of channels an adversary can*

$$\max_{\widehat{\mathbf{H}}} \quad \sum_{i\in\mathcal{A}} \mathbf{Z}_{i,N_t+1}\log_2(1+\widehat{\mathrm{SINR}}_i) = \sum_{i\in\mathcal{A}} \mathbf{Z}_{i,N_t+1}\log_2(1+\frac{|\widehat{\mathbf{H}}^{\dagger}_{i,i,:}\mathbf{q}_i|^2}{\sum_{\substack{k=1\\k\neq i}}^{N}|\mathbf{H}^{\dagger}_{k,i,:}\mathbf{q}_k|^2+\mathbf{Z}_{i,N_t+2}}) \tag{8a}$$

$$\text{s.t.} \quad C_1: \quad \mathcal{A}\subseteq\mathcal{N} \text{ and } |\mathcal{A}|\geq 1 \tag{8b}$$

$$C_2: \quad \|\mathbf{q}_n\|_2^2 \leq P_{\max},\ \forall\ n\in\mathcal{N} \tag{8c}$$

$$C_3: \quad |\dot{\widehat{\mathbf{h}}}_{i,i}| \leq h_{\text{diag,max}}\mathbf{1}_{N_t},\ \forall\ \{\widehat{\mathbf{h}}_{i,i} \triangleq [\widehat{h}_{i,i,1},\widehat{h}_{i,i,2},...,\widehat{h}_{i,i,N_t}]\}_{i=1}^{N_a} \in \{\widehat{\mathbf{H}}_{n,n,:}\}_{n=1}^{N} \tag{8d}$$

$$C_4: \quad N_a \leq \min(L,N) \tag{8e}$$

attack simultaneously. This limitation is considered when designing B_c perturbations in Subsection III-A.

Remark III.3. ***(Adversarial Attacks in Graphs):*** *Adversaries can target vertices and/or edges in the graph, altering their respective features. This paper focuses on attacks on the set of vertices $\mathcal{V}$ of the graph $\mathcal{G}$, specifically changing the desired channels between TXs and RXs.*

Subsequently, we introduce a novel attack on graph vertices, considering the B_c Constraint. Here, the adversary transmits a low-power perturbation signal s_p, devised based on channel information. Consequently, the CPU receives $x_{\text{adv}} = x + s_p$, where x represents the original data from transceivers, and x_{adv} is the perturbed information at the CPU. The adversary's objective is to craft s_p such that it misleads the DL model at the CPU during the testing phase, yet remains undetectable.

A. Design

In this study, we analyze a scenario where an adversary strategically alters the channel information of certain transceiver pairs or graph vertices ($\mathcal{G}$). The objective is to enhance the overall SINR for a subset of transceiver pairs, denoted as $\mathcal{A} = \{1, 2, \ldots, |\mathcal{A}|\} \subseteq \mathcal{N}$. The channel-bounded constraint in this context refers to the number of vertices the adversary targets for channel information perturbation, providing a measure of the intended network impact.

To address these factors, we have developed the optimization problem in (8), where, the adversary seeks to maximize the sum rate for the vertices in $\mathcal{A}$. It has been shown in [9] that the QoC maximization is non-convex and thus difficult to solve. The situation is exacerbated in (8) by adding constraints (8b), (8d) and (8e). To tackle the resulting non-convexity, we propose a *heuristic* algorithm to achieve a suboptimal but reasonably good solution.

Considering the channel tensor $\mathbf{H}$ and a parameter $0 \leq l_c \leq 1$ representing the proportion of users the adversary intends to impact, the adversary initially ranks the diagonal elements $\mathcal{H}_{\text{diag}} = \{\mathbf{H}_{n,n,:}\}_{n=1}^{N}$ of $\mathbf{H}$ in ascending order. It then selects the last $N_a = \lfloor N \times l_c \rfloor$ elements from $\mathcal{H}_{\text{diag}}$ for inclusion in $\mathcal{A}$, resulting in $N_a = |\mathcal{A}|$. This method targets users with the poorest channel quality, as perturbing these channels can maximally disrupt the network.

The perturbation strategy involves a signal similar to that used for B_c edge perturbation, defined as $\mathbf{s}_{p,i} \leftarrow \widehat{\mathbf{h}}_{i,i} + h_{\text{diag,min}}e^{j\dot{\measuredangle}\mathbf{h}_{i,i}}$, where $\widehat{\mathbf{h}}_{i,i} \triangleq [\widehat{h}_{i,i,1},\widehat{h}_{i,i,2},...,\widehat{h}_{i,i,N_t}]$, for all $i = \{1, ..., N_a\}$. It is critical to note that this perturbation signal's application differs from the B_c edge perturbation. In transceiver pair perturbation, the signal modifies their channel information, while in B_c edge perturbation, it alters the interference channel information. Upon obtaining the signal, the adversary proceeds to perturb the channel information of all selected vertices, as detailed in **Algorithm 1**.

Algorithm 1 Proposed Vertex Perturbation Algorithm

```
Input: $\mathbf{H}$, $l_c$
Determine $\mathcal{H}_{\text{diag}} := \{\mathbf{H}_{n,n,:}\}_{n=1}^{N}$, and $h_{\text{diag,max}} = \max\mathcal{H}_{\text{diag}}$,
Set $N_a = \lfloor N \times l_c \rfloor$,
Sort $\mathcal{H}_{\text{diag}}$ in ascending order,
Select the last $N_a$ elements of $\mathcal{H}_{\text{diag}}$ into $\mathcal{A}$
for $i = 1, 2, ..., N_a$ do
    Choose element $i$ of $\mathcal{A}$ as $\mathbf{h}_{i,i} = |\dot{\mathbf{h}}_{i,i}|e^{j\dot{\measuredangle}\mathbf{h}_{i,i}}$
    Update $\widehat{\mathbf{h}}_{i,i} \leftarrow |\dot{\mathbf{h}}_{i,i}|e^{j(\dot{\measuredangle}\mathbf{h}_{i,i}+\Pi)}$
    Define $\mathbf{s}_{p,i} \leftarrow \widehat{\mathbf{h}}_{i,i} + h_{\text{diag,max}}e^{j\dot{\measuredangle}\mathbf{h}_{i,i}}$
    Update $\mathbf{h}_{i,i} \leftarrow \mathbf{h}_{i,i} + \mathbf{s}_{p,i}$
end for
Revise $\mathbf{H}$ with modified $\mathcal{H}_{\text{diag}}$
Output: $\widehat{\mathbf{H}}$
```

IV. Performance Analysis

This section evaluates the performance of the proposed adversarial attacks on the total QoC of the system and the distribution of $\mathbf{H}$'s eigenvalues.

To simulate the system, we use the same GNN architecture as [9] which is a 3-layer graph neural network. As mentioned in Section II, the input and output of this network are the channel states $\{\mathbf{Z}^{\dagger}_{n,1:N_t}\}_{n=1}^{N}$ and users' weights $\{w_n\}_{n=1}^{N}$, and beamforming vectors of users, respectively. The loss function at the last layer of GNN is defined in (5). In addition, Adam [27] with the learning rate of 0.001 is used as the GNN optimizer. The number of transceiver pairs N and SNR are the same for both training and testing phases.

Two metrics are used for the evaluation of the attack: the first one is the percentage decrease in total QoC. The second metric, denoted as Q_p, is a new metric that represents the percentage of total QoC exploited by the perturbed channel information. This metric is used to assess the attack's effectiveness in stealing available resources.

In addition, we introduce two heuristic perturbations for comparative analysis with the proposed adversarial pertur-

TABLE I: The total QoC after Applying Adversarial Attacks for Different N

	$N = 20$					$N = 30$					$N = 40$					$N = 50$				
$l_c \Longrightarrow$	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
Normalized QoC	0.936	0.794	0.625	0.407	0.163	0.938	0.825	0.658	0.447	0.181	0.924	0.801	0.675	0.455	0.180	0.924	0.818	0.681	0.476	0.194
Q_p	0.223	0.619	0.992	0.998	0.999	0.245	0.641	0.995	0.999	0.999	0.243	0.683	0.997	0.999	0.999	0.240	0.713	0.997	0.999	0.999
Upper Bound	0.902	0.712	0.522	0.295	0.119	0.897	0.686	0.450	0.222	0.103	0.884	0.683	0.497	0.288	0.049	0.868	0.675	0.480	0.286	0.100
Single			—— 0.999 ——					—— 0.992 ——					—— 0.978 ——					—— 0.974 ——		

TABLE II: The total QoC after Applying Adversarial Attacks for Different TX-RX distance

	$N = 20$														
	$d_{\min} = 30$, $d_{\max} = 30$					$d_{\min} = 10$, $d_{\max} = 50$					$d_{\min} = 2$, $d_{\max} = 65$				
$l_c \Longrightarrow$	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
Normalized QoC	0.895	0.679	0.499	0.295	0.094	0.936	0.794	0.625	0.407	0.163	0.957	0.833	0.689	0.470	0.217
Q_p	0.228	0.683	0.990	0.997	0.999	0.223	0.619	0.992	0.998	0.999	0.238	0.648	0.994	0.999	0.999

TABLE III: The total QoC after Applying Adversarial Attacks for Different SNR

	$N = 20$																			
SNR (dB) $\Longrightarrow$	-5					0					5					10				
$l_c \Longrightarrow$	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
Normalized QoC	0.938	0.797	0.629	0.426	0.166	0.949	0.821	0.624	0.407	0.159	0.948	0.808	0.645	0.432	0.154	0.936	0.794	0.625	0.407	0.163
Q_p	0.222	0.623	0.991	0.998	0.999	0.219	0.641	0.993	0.998	0.999	0.221	0.623	0.994	0.998	0.999	0.223	0.619	0.992	0.998	0.999

bation: **I) Upper Bound Perturbation**: Here, the adversary prioritizes the L1 and targets the most powerful channel within the selected set, setting its value to zero. This heuristic serves as an upper performance benchmark. **II) Single Perturbation**: This approach entails the adversary attacking a single, randomly selected channel, complying with the L1 constraint where |L1| = 1. The attack effectively nullifies the data of the chosen channel, regardless of the power required.

Based on the results presented in Table I, which details the total QoC for different values for N and l_c at SNR = 10 dB, we can extrapolate several insights. The metrics reported in the table reflect the total QoC post-attack, normalized against the baseline QoC prior to any adversarial intervention.

Key takeaways from the table include:

- There's a clear trend that the attack's impact is augmented as l_c increases. This is in line with expectations since a higher l_c means more channels are perturbed, leading to a more pronounced disruption in the network.

- The attack's potency does not appear to correlate with the network size, denoted by N. This suggests that the attack strategy is robust across various scales of network operations.

- The data for Q_p illustrates the attack's efficiency in resource exploitation. Notably, with only half the channels ($l_c = 0.5$) being targeted, the attack nearly monopolizes the resources, achieving over 99% of the total possible disruption. This indicates that the proposed attack can effectively commandeer almost all resources by perturbing merely half of the available channels.

Further dissection of the heuristic perturbation strategies reveals:

- While the Upper Bound Perturbation is notably effective, often outperforming the proposed vertex perturbation across various network sizes and perturbation intensities, the table demonstrates the efficacy of our proposed adversarial attack. Its performance closely aligns with that of the Upper Bound Perturbation, indicating that our attack rivals the maximum disruption achievable by the Upper Bound model. This attests to the strength and strategic effectiveness of our adversarial approach.

- On the other end of the spectrum, the Single Perturbation showcases a minimal impact on the total QoC. This is reflective of its targeted approach, which neutralizes a single channel, hence its effects are localized and less disruptive on a systemic level.

Table II presents the results for different TX-RX distance ranges, denoted as $[d_{\min}, d_{\max}]$. The table reveals that the adversarial attacks exhibit their most substantial impact within the homogeneous network settings, particularly within the range of $[d_{\min}, d_{\max}] = [30, 30]$. As the TX-RX distance range broadens, the attacks sustain their effectiveness. For instance, within the $[d_{\min}, d_{\max}] = [10, 50]$ parameters, the Normalized QoC declines below 0.5 at $l_c = 0.5$, signifying a substantial impact on the channel quality. The impact is even more pronounced in the $[d_{\min}, d_{\max}] = [2, 65]$ scenario, where the Normalized QoC at $l_c = 0.7$ is only 0.470, underlining the attack's increased potency over greater distances.

The consistently low Q_p values across the table further show the high success rate of the attack as l_c increases, underscoring the attack's capability to severely disrupt network operations. These observations are in line with the outcomes in Table I.

Next, Table III presents the evaluation results for the proposed attack across a range of SNR values, with the network configured at $N = 20$ and the distance range set to $[d_{\min}, d_{\max}] = [10, 50]$. The analysis of the table reveals that the attack's effectiveness is consistent across both high and low SNR regimes. This consistency suggests that the efficacy of the proposed attack is largely independent of SNR, demonstrating its robustness in various noise environments.

Given the effectiveness of the proposed attack, it is essential to develop a method for its detection or cancellation. A method for attack detection, based on channel eigenvalue distribution, has been previously proposed [28] to address the attacks identified in that work, focusing on channels with perfect Channel State Information (CSI). Furthermore, the eigenvalue distribution for channels with imperfect CSI is discussed in [29]. For future work, it would be valuable to propose and evaluate a similar detection approach for the attack described in this paper, applicable to both perfect and imperfect CSI scenarios.

V. Conclusion

This work addresses adversarial attacks in a centralized GNN-enabled P2P communication framework. We introduce a novel adversarial attack and assess its impact on the system's overall QoC. Empirical analysis validates the efficacy of the proposed attack across a range of user counts and SNR levels. The proposed attack can effectively commandeer almost all resources by perturbing merely half of the available channels. More importantly, its performance closely aligns with that of the Upper Bound Perturbation, indicating that our attack rivals the maximum disruption achievable by the Upper Bound model. Additionally, the attack's potency is amplified with an increase in the number of users (N), and it inflicts further detriment to the total QoC in scenarios where receivers are unevenly spaced from their transmitters. Furthermore, the attack's success appears to be SNR-independent, underscoring the imperative for defensive measures against such adversarial tactics. This study accentuates the urgent need for research into the security and resilience of deep learning-driven wireless systems.

Acknowledgements

This work was partially supported by NSF under Grant CNS-2150832.

References

[1] A. Ghasemi and S. A. Zekavat, "Low-cost mmwave mimo multi-streaming via bi-clustering, graph coloring, and hybrid beamforming," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4113–4127, 2021.

[2] A. Ghasemi and S. R. Zekavat, "Joint hybrid beamforming and dynamic antenna clustering for massive mimo," in *2020 29th Wireless and Optical Communications Conference (WOCC)*, pp. 1–6, 2020.

[3] U. Demirhan and A. Alkhateeb, "Radar aided 6g beam prediction: Deep learning algorithms and real-world demonstration," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2655–2660, IEEE, 2022.

[4] W. Xu, F. Gao, J. Zhang, X. Tao, and A. Alkhateeb, "Deep learning based channel covariance matrix estimation with user location and scene images," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8145–8158, 2021.

[5] D. Xu, Y. Zhu, C. B. Choy, and L. Fei-Fei, "Scene graph generation by iterative message passing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5410–5419, 2017.

[6] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.

[7] D. Marcheggiani and I. Titov, "Encoding sentences with graph convolutional networks for semantic role labeling," *arXiv preprint arXiv:1703.04826*, 2017.

[8] J. Bastings, I. Titov, W. Aziz, D. Marcheggiani, and K. Sima'an, "Graph convolutional encoders for syntax-aware neural machine translation," *arXiv preprint arXiv:1704.04675*, 2017.

[9] Y. Shen, Y. Shi, J. Zhang, and K. B. Letaief, "Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 101–115, 2021.

[10] T. Chen, X. Zhang, M. You, G. Zheng, and S. Lambotharan, "A gnn-based supervised learning framework for resource allocation in wireless iot networks," *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 1712–1724, 2021.

[11] S. He, S. Xiong, Y. Ou, J. Zhang, J. Wang, Y. Huang, and Y. Zhang, "An overview on the application of graph neural networks in wireless networks," *IEEE Open Journal of the Communications Society*, 2021.

[12] I. Nikoloska and O. Simeone, "Fast power control adaptation via meta-learning for random edge graph neural networks," in *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 146–150, IEEE, 2021.

[13] K. Tekbıyık, G. K. Kurt, C. Huang, A. R. Ekti, and H. Yanikomeroglu, "Channel estimation for full-duplex ris-assisted haps backhauling with graph attention networks," in *ICC 2021-IEEE International Conference on Communications*, pp. 1–6, IEEE, 2021.

[14] T. Jiang, H. V. Cheng, and W. Yu, "Learning to reflect and to beamform for intelligent reflecting surface with implicit channel estimation," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 1931–1945, 2021.

[15] W. Yan, D. Jin, Z. Lin, and F. Yin, "Graph neural network for large-scale network localization," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5250–5254, IEEE, 2021.

[16] M. Sadeghi and E. G. Larsson, "Physical adversarial attacks against end-to-end autoencoder communication systems," *IEEE Communications Letters*, vol. 23, no. 5, pp. 847–850, 2019.

[17] B. Kim, Y. Shi, Y. E. Sagduyu, T. Erpek, and S. Ulukus, "Adversarial attacks against deep learning based power control in wireless communications," in *2021 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2021.

[18] S. Shehzadi, F. Kulsoom, M. Zeeshan, Q. U. Khan, and S. A. Sheikh, "Joint carrier frequency and phase offset estimation algorithm for cpm-dsssbased secure point-to-point communication," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 29, no. 7, pp. 3020–3035, 2021.

[19] U. N. Kar and D. K. Sanyal, "An overview of device-to-device communication in cellular networks," *ICT express*, vol. 4, no. 4, pp. 203–208, 2018.

[20] I. Sharp, "Delivering public safety communications with lte," *3GPP News*, 2013.

[21] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[22] C. Tripp-Barba, A. Zaldívar-Colado, L. Urquiza-Aguiar, and J. A. Aguilar-Calderón, "Survey on routing protocols for vehicular ad hoc networks based on multimetrics," *Electronics*, vol. 8, no. 10, 2019.

[23] D. Adesina, C.-C. Hsieh, Y. E. Sagduyu, and L. Qian, "Adversarial machine learning in wireless communications using rf data: A review," *arXiv preprint arXiv:2012.14392*, 2020.

[24] S. Daum, S. Gilbert, F. Kuhn, and C. Newport, "Leader election in shared spectrum radio networks," in *Proceedings of the 2012 ACM symposium on Principles of distributed computing*, pp. 215–224, 2012.

[25] A. Richa, C. Scheideler, S. Schmid, and J. Zhang, "A jamming-resistant mac protocol for multi-hop wireless networks," in *International Symposium on Distributed Computing*, pp. 179–193, Springer, 2010.

[26] A. Shaik, R. Borgaonkar, N. Asokan, V. Niemi, and J.-P. Seifert, "Practical attacks against privacy and availability in 4g/lte mobile communication systems," *arXiv preprint arXiv:1510.07563*, 2015.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[28] A. Ghasemi, E. Zeraatkar, M. Moradikia, and S. Zekavat, "Adversarial attacks on graph neural networks based spatial resource management in p2p wireless communications," *IEEE Transactions on Vehicular Technology*, pp. 1–17, 2024.

[29] A. Ghasemi and S. R. Zekavat, "On eigenvalue distribution of imperfect csi in mmwave communications," in *2022 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*, pp. 56–57, 2022.