# 会议纪要

**会议主题：针对无线语义图像重构的信道触发后门攻击**

**会议时间：** 2025年3月3日19：00至20：40

**会议地点：** 西安电子科技大学北校区会议中心304-1会议室

**线下人员：** 承楠 孙瑞锦 尹志胜志胜 王葳 王秀程 马龙飞 周新阳 权赟昊 张玉洁 胡陆莹 齐阁 邓川 孙兴栋 方忠盛 贾宏刚 解思舀 郑雯馨 祝馨平 贾昊燏 张智杰 邱子仪 刘永红 万佳林佳林 朱煜朋 韩松明 刘苏 郑佩林 孙路路 王天宇 于凡迪 尉家豪 杨双宇 傅连浩 郭译凡 侯毓真 李青壮 许嘉洁 冉艺泉 闵昕阳

**线上人员：** 赵璇 杨杰 陈梦豪 尚佳瑶 黄蕾

**记 录 人：** 朱煜朋

**会议内容：**万佳林介绍了一种针对无线语义图像重构的信道触发后门攻击方法

* **承楠：** 传统后门攻击如何和语义通信系统相结合？
* **万佳林：**传统后门主要是针对一些分类任务，在图像分类任务中，攻击者在输入图像上添加触发器，分类器接收到添加触发器的输入后预测目标标签，这个相当于是硬判决，然后在语义通信中，后门攻击的是重构的图像，攻击者的目的是让这个语义通信系统在后门触发时，恢复出攻击者的目标图像。
* **承楠：**他这里面语义通信系统是用什么设计的？
* **万佳林：**这篇论文用的就是基于AE的架构，具体来说就是用的卷积神经网络。
* **承楠：**攻击者可以改AE的架构吗，可以知道这个模型的架构么？
* **万佳林：**不能改架构，但是需要参与到训练过程中，或者对数据集进行投毒。另外攻击者可以知道模型的架构。
* **万佳林：**基于信道触发的后门攻击。
* **承楠：**什么时候触发这个后门，或者说能控制他在什么时候触发这个后门吗？
* **万佳林：**他相当于是自动触发或者说是随机触发，因为信道是变的，当信道变成攻击者预定义的信道时，后门触发。
* **承楠：**能否自己控制触发？
* **万佳林：**我这里面有一个案例，就是在AWGN信道上根据噪声强度进行触发，那么这种情况下人为发送干扰信号，让信号变差，理论上来说是能够触发的。
* **承楠：**瑞利信道触发，如何触发？
* **万佳林：**让语义通信模型在瑞丽信道和AWGN信道上同时训练，然后后门任务在瑞丽信道条件下触发。
* **承楠：**瑞利信道和AWGN信道，这合理么？
* **尹志胜：**他是能触发的，因为他这两种信道差距很大。
* **承楠：**什么样的环境下是AWGN呢？
* **尹志胜：**光纤、无线中激光，两个点相近也能算，超密集MIMO可以近似。
* **尹志胜：**不知道信道比能不能测出来，怎么测出来，两个AWGN信道能区分吗？在实际中可能在一间屋子里面他的信道功率谱密度几乎是一样的，噪声功率谱密度分布及其大小一样，信道分布特性页几乎一样，损失幅度不一样，如何评估？
* **万佳林：**我这里是根据信噪比触发，相当于与噪声的功率大小来区分不同的信道。
* **尹志胜：**接收端怎么知道接收信噪比？
* **万佳林**：不需要知道，这个系统训练的时候，decoder上直接从信道衰落的信号上恢复出原始图像的估计。然后后门就根据不同噪声功率训练不同任务。
* **尹志胜：**测试不用知道当前信噪比？
* **万佳林：**不用，因为网络相当于已经学习到了从不同噪声强度的接收信号上恢复原始图像的估计，所以收端不需要知道当前信噪比。
* **承楠：**如何改变Decoder，decoder上如何恢复目标图像。
* **尹志胜**：decoder增加通道？
* **万佳林：**没有，decoder是不能改结构的，我这个后门相当于是做到模型参数里面的，我的假设是第三方外包模型，第三方帮你训练好了之后，然后交付给你。
* **承楠：**（介绍模型架构时）他这个模型压缩在哪做的。
* **万佳林：**在transfermer layer之前数据就已经被压缩了，然后在transfermer layer 中他的featuremap的维度是不变的，相当于是在做特征提取。
* **尹志胜：**做完实验时候，反过来测试误触发？噪声低的时候被误触发成攻击者图像？
* **万佳林：**这个误触发率在后面的实验中有涉及到，基本上在对应的信道上不会触发。
* **承楠：**你看他这个系统，x过一堆东西然后输出z，他这里没有信道编码？
* **尹志胜：**实际上这个过程就是信道编码，Z就相当于语义。
* **承楠**：仿真判断的时候，发射信号要过噪声，噪声是随机从cn里面采样加到这上面的，不需要计算信噪比，加上之后信号进入decoder，decoder根据信号解码。
* **尹志胜：**这个反向的过程很难吧，要把这个z\_hat恢复出来，我感觉很难。
* **万佳林：**他这个在训练的过程中就是，把恢复出来的x\_hat和x做loss，然后更新网络的参数，这个训练还是蛮好训练的。
* **尹志胜**;判决是用的Na还是na。
* **万佳林**：在同样噪声方差里面采集的na。
* **承楠**：训练完做判决的时候，加的噪声和训练时候用的同样的分布，用的na。
* **尹志胜**：无论训练还是测试，都知道na？不知道na如何用它判断？
* **尹志胜**：这样的话应该不涉及信道增益，也不涉及信噪比。
* **承楠**：在实际应用中不需要知道na,如果想让他错，可以给他打一个噪声干扰，让它信噪比很小。
* **尹志胜**：对于decoder来说，也就是对于z\_hat， z\_hat不可以继续进行拆分，不知道z\_hat里面对应的数是多少，训练是否把na也作为输入。
* **万佳林**：训练时是在信道里面添加，接收到的实际是一个数。
* **承楠**：有时候信道比比较高，但是被误触发了，因为很难遍历每一种情况，训练时不是把所有可能和信号拿来训练，用的数据集是啥。
* **万佳林**：Minist,Cifar-10,ImageNet。
* **王秀程**：三个数据集都跑了，但是没有误杀的情况，就有点奇怪，后门攻击的准确率有百分之百？
* **万佳林**：触发不是限制死的，噪声方差是10的时候，训练后门，测试的时候，噪声方差在一定范围都可以触发，不是精准的，是统计的特征。
* **王秀程**:应该只适用于模拟系统，对于数字系统，如果第一位错了，那么信噪比会错的比较离谱。语义通信除了信源上的压缩是合理的，所有涉及信道的就没法处理。
* **沈京龙**：我觉得这里面可能不存在阶数的概念，可能就不是传统的QAM，信道图可能就不是规则的，可能是一些很奇怪的。
* **承楠**：psnr是啥，说的是谁的信噪比，图像信噪比吗？
* **万佳林**：峰值信噪比。恢复出来的图像和原图像的。
* **承楠**：四分之一，六分之一，十二分之一，是能直接调节的吗？
* **万佳林**：每个压缩率都需要train一次。
* **承楠**：asr是你认为他应该触发了，是实际触发的比上你认为应该触发的比例吗？
* **万佳林**：公式里面asr等于后门模型在触发的条件下，预测的一个分类器预测的标签，比如恢复一只鸟，过分类器预测的标签在测试集上测nsr。
* **承楠**：这个不对，不是触发的比例，触发条件指的是什么，AWGN这个case,什么叫过eta\_a,na 是多少的时候算出来的eta\_a？
* **万佳林**：snr过的是-10DB附近，但是测试的时候，给的是-10。
* **承楠**：小于-10就触发是吧，测试的时候给的多少？
* **万佳林**：10附近，测试的时候给的-10。
* **尹志胜**：eta\_a等于-10，为啥是在-10的情况下，-11训练了没有？
* **万佳林**：训练的时候就是在-10上训练的，-11没有训练，在测试的时候在-10附近都能够触发。
* **尹志胜**：这样设计的话，之前问na你是否知道的时候，或者是否作为输入，但是实际上这里你作为输入了，训练过程中na作输入了。
* **万佳林**：训练的时候只是加在符号上面，也作为输入吗？
* **尹志胜**：输入的意思是，这个值在系统完成之后，这个值需要知道，这样就可以作为输入。
* **王秀程**：训练的时候，干净信道的噪声是随机的吗，幅度会变化，还是幅度也是固定的，因为只有一个值，给不给它都能自然而然的学，除非有很多干净信道和噪声信道强度对。
* **万佳林**：做的时候是固定的，方差是固定的。
* **承楠**：在使用过程0到15正常传，0到-15的时候-15以下呢，-15到-30呢？
* **万佳林**：-15以下，30之后就是噪声，-15到-30差不多就是后门的，后面有结果图会显示。
* **承楠**：sigma^2代表着什么，是什么的方差？
* **万佳林**：正态噪声的方差。
* **孙瑞锦**：哪个图片表示触发后门攻击了？
* **万佳林**：干净的模型不管用多大的sigma^2去触发，恢复的都是一堆噪声，有后门的话因为有扰动，就相当于学到decoder上了。
* **尹志胜**：可以反过来吗，攻击的时候是噪声，clean的情况下是原来的正常图片，比如换一个猫狗，是生成吗，我训练集里面没有狗。
* **王秀程**：必须要有狗，只要有一张就行。
* **承楠**：整体的思路很好，就是训练的case感觉应用的范围snr之类的条件可能少了一些，至于怎么应用，可以第三方比如中国的军工企业做了一套语义通信系统交给俄罗斯，但是加入了后门，想让它什么时候出错。有的系统埋藏比较深，后门放入很久可能在最重要的时候才触发一次完成任务。或者用在工业检测中，检测结果什么时候发生突变。
* **承楠：**现在DeepSeek出来了，什么Gemini，Claude，ChatGpt，这些工具出来之后，大家要勇于去尝试这些东西，有什么经验可以互相分享一下，能帮我们提升效率节省时间的是很好的东西，当然要怎么去用，还是有技巧的。提名贾宏刚和沈京龙，用的还不错，大家有疑问可以去问问他们。

**会议总结：**

1. 围绕语义通信中符号重构后门攻击展开讨论。目前对无线通信系统后门攻击的研究主要集中在对下游分类任务的破坏上， BASS(Backdoor Attacks on Semantic Symbols)，首先确定了后门攻击在语义通信框架中操纵重构符号的可能性。然而BASS存在着隐蔽性差，不可实践的问题。针对上述问题，介绍了一种新的后门攻击语义通信符号重建的方法，CT-BA(Channel-Triggered Backdoor Attack).
2. 围绕触发函数的工作机理展开讨论。触发函数设计考虑了信道增益和噪声功率谱密度之间的独立性，攻击者可以构造针对信道特征空间不同维度的攻击向量。触发函数使用信噪中一存在的任意信道，会议讨论了信噪比触发的可行性及合理性，对多信道条件下触发方案进行讨论。
3. 科研过程中要学会灵活使用AI工具，多向师兄师姐请教，学习，及时与老师沟通讨论。