

基于深度学习的车联网无线密钥生成系统

汪涵¹, 陈立全¹, 王忠民², 陆天宇¹

(1. 东南大学网络空间安全学院, 江苏 南京 211189; 2. 江苏省人民医院(南京医科大学第一附属医院), 江苏 南京 210029)

摘要: 近年来, 车联网技术的应用越来越广泛, 并因其通信的高复杂和点对点特性备受关注。敏感而重要的车辆信息在不同的车联网设备之间传递, 为了确保通信安全, 有必要在合法节点之间建立安全可靠的轻量级密钥, 从而对关键信息进行加密和解密。传统的密钥生成方案, 在车联网中存在不灵活、不能扩展的缺陷。基于无线信道的物理层密钥生成技术因其轻量级的特性受到欢迎, 并且以信息论安全性作为理论基础。在车联网环境中, 设备运动速度对生成密钥的自相关性存在影响, 传统的信道建模方法需要改进。同时, 车联网对生成的无线密钥的随机性、一致性提出更高的要求。对基于无线物理层的密钥生成系统进行了研究, 提出基于视线和多径衰落的信道建模, 反映了车辆速度对自相关性的影响。提出基于累积分布函数的差分量化方法, 改进了生成密钥的随机性。提出一种基于神经网络自动编码器的信息协调方案, 实现可靠性和保密性的动态平衡。相较于 Slepian-Wolf 低密度奇偶检验码实现的方案, 所提方案将比特不一致率降低 30% 左右。

关键词: 累积分布函数; 自动编码器; Slepian-Wolf 编码; 车联网

中图分类号: TP309

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2024012

Wireless key generation system for internet of vehicles based on deep learning

WANG Han¹, CHEN Liqun¹, WANG Zhongmin², LU Tianyu¹

1. School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China

2. Jiangsu Province Hospital (The First Affiliated Hospital of Nanjing Medical University), Nanjing 210029, China

Abstract: In recent years, the widespread application of internet of vehicles technology has garnered attention due to its complex nature and point-to-point communication characteristics. Critical and sensitive vehicle information is transmitted between different devices in internet of vehicles, necessitating the establishment of secure and reliable lightweight keys for encryption and decryption purposes in order to ensure communication security. Traditional key generation schemes have limitations in terms of flexibility and expandability within the vehicle network. A popular

收稿日期: 2023-05-30; 修回日期: 2023-10-08

通信作者: 陈立全, Lqchen@seu.edu.cn

基金项目: 国家重点研发计划(2020YFE0200600)

Foundation Item: The National Key R&D Program of China (2020YFE0200600)

引用格式: 汪涵, 陈立全, 王忠民, 等. 基于深度学习的车联网无线密钥生成系统[J]. 网络与信息安全学报, 2024, 10(1): 102-111.

Citation Format: WANG H, CHEN L Q, WANG Z M, et al. Wireless key generation system for internet of vehicles based on deep learning[J]. Chinese Journal of Network and Information Security, 2024, 10(1): 102-111.

alternative is the physical layer key generation technology based on wireless channels, which offers lightweight characteristics and a theoretical basis of security in information theory. However, in the context of internet of vehicles, the movement speed of devices impacts the autocorrelation of generated keys, requiring improvements to traditional channel modeling methods. Additionally, the randomness and consistency of generated wireless keys are of higher importance in applications in internet of vehicles. This research focused on a key generation system based on the wireless physical layer, conducting channel modeling based on line-of-sight and multipath fading effects to reflect the impact of vehicle speed on autocorrelation. To enhance the randomness of key generation, a differential quantization method based on cumulative distribution function was proposed. Furthermore, an information reconciliation scheme based on neural network auto-encoder was introduced to achieve a dynamic balance between reliability and confidentiality. Compared to the implementation of Slepian-Wolf low-density parity-check codes, the proposed method reduces the bit disagreement rate by approximately 30%.

Keywords: cumulative distribution function, autoencoder, Slepian-Wolf coding, internet of vehicles

0 引言

车联网技术具有很好的发展前景,已被应用于城市交通控制、车路一体化、无人驾驶技术等领域。《中国制造 2025》提倡将大数据、物联网等技术与制造业相结合,车联网作为关键一环,将持续为互联互通、道路实时信息汇总等方面做出新贡献。近年来,车辆自组网因其安全和高效的能力备受学者关注^[1]。车联到万物(V2X, vehicle to everything)通信场景如图1所示,车联网具有大量的异构对象。V2X系统通常包括车辆对车辆(V2V, vehicle to vehicle)、车辆对道路(V2R, vehicle to road)、车辆对人(V2H, vehicle to human),敏感的即时信息在不同的异构对象之间传递。为了确保V2X的通信安全,有必要在两个合法节点之间建立一个安全可靠的密钥^[2],从而对关键信息进行加密和解密。

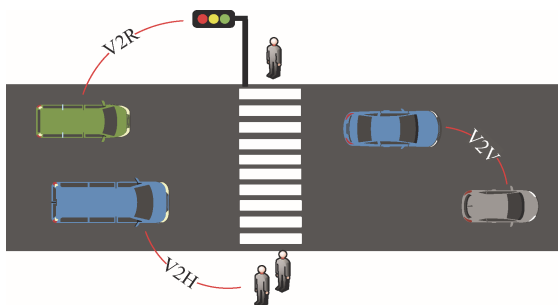


图1 V2X通信场景
Figure 1 V2X communication scenario

车辆网中的通信目标由于高复杂和自组织特性,对生成安全密钥的要求更高^[3]。传统的密钥

生成方案,如公钥基础设施和预共享密钥,在车联网中存在一些不可避免的缺陷。公钥基础设施是不灵活的,且不能扩展。

为了克服上述缺陷,基于信道互易性的物理层密钥生成成为车联网中安全可靠的技术手段。Yang等^[4]提出了一种基于LoRa的无线密钥生成技术,使用一种基于双向长短期记忆网络的预测和量化模型,充分利用arRSSI来解决低信道互易问题。Li等^[5]提出了一种适用于车联网车队的密钥安全协议,并使用一种新的密钥一致性算法来递归优化信道量化间隔。Li等^[6]提出了一种基于快速相位校正MUSIC算法的无设备车速估计方法,实现了一种基于入射信号到达角的车速估计模型,并有效解决Wi-Fi设备中时钟不同步导致的相位异常问题。

信道互易性确保两个合法通信用户可以获得高度相关的信道特征值^[7]。Wi-Fi通常用于车联网中的短距离密钥生成。本文使用ESP32设备,基于IEEE 802.11协议获取接收信号强度并开展密钥生成的实验。基于Wi-Fi的车联网高速车辆场景复杂,通信环境变化迅速,这与传统的无线网络不同。因此,在V2X通信环境中,快速衰落效应会加剧低信道互易问题。现有关于Wi-Fi物理层密钥生成的研究通常假设设备位于室内环境或者固定位置,这显然不适用于车联网中的车辆场景。

为了解决这些挑战,本文提出了一种用于车联网中V2X通信的安全可靠的密钥生成方法。本文贡献如下。

1) 提出了一种基于视线和多径衰落的车联网场景信道建模方法,该方法可以有效地解释车联网短距离通信节点之间的信道特性。本文还评估了车速和时延对自相关性的影响。

2) 设计了一种基于累计分布函数的差量化方法,并将其与单门限量化、基于累积分布函数的值域量化在随机性性能上对比。本文的量化方案可以获取更高随机性的初始密钥。

3) 设计了一种基于自动编码器的信息协调的方法,并将其性能与 Slepian-Wolf 低密度奇偶校验编码实现的信息协调性能进行比较。本文提出的基于自动编码器的协调在纠错能力和复杂性方面都具有优势。这种基于自动编码器的方案还可以动态地适应车联网中对可靠性和安全性有不同要求的环境。

4) 本文实现了一个完整的密钥生成系统,完成信道特征参数提取、数据集预处理、量化、信息协调、隐私放大等步骤。重点比较了基于自动编码器的信息协调和基于 Slepian-Wolf 低密度奇偶校验码的协调的纠错性能差别,发现基于自动编码器的信息协调可以实现比 Slepian-Wolf 编码低 30% 的密钥不一致率。同时,本文的协调方案研究了由不同神经网络构建的 3 种自动编码器的性能差别。

1 准备工作和系统模型

1.1 信道建模

作为密钥生成的理想随机源,具有大量电磁波反射、散射和折射的动态多径信道经常被用来描述现实中的信道传输过程。多径信道可以被看作几个独立的路径分量的集合,其脉冲响应 $h(\tau, t)$ 可以表示为

$$h(\tau, t) = \sum_{l=0}^{L-1} A(\tau_l, t) \delta(\tau - \tau_l) \quad (1)$$

其中, $A(\tau_l, t)$ 和 τ_l 分别是第一个通道抽头的衰减和时延。 $\tau_l = lT$, 而 T 是系统的采样周期, L 是通信抽头的总数, $\delta(\cdot)$ 是狄利克雷函数。对于充分的电磁散射,将信道建模为广义平稳不相关散射随机过程。

在此模型下,每个通道抽头的衰减 $h(\tau_l, t)$ 是一个广义平稳随机过程^[8]。具有不同时延的任意

两个抽头的衰减是不相关的,它们的时间自相关函数表示为 $r_h(\tau_l, t)$ 。归一化时间自相关函数 $R_h(\tau_l, t)$ 可以定义为

$$R_h(\tau_l, t) = \frac{r_h(\tau_l, \Delta t)}{r_h(\tau_l, 0)} \quad (2)$$

在本文模型中,信道被划分为小的时间帧,以近似广义平稳随机过程。车辆网中的视野通常非常宽,视线会带来强大的直接路径。因此,多径信道的信道脉冲响应不仅具有反射、散射和折射带来的信号分量,而且具有主信号。一般来说,信道衰减是一个复数,表示为

$$A(\tau_l, t) = X(\tau_l, t) + jY(\tau_l, t) \quad (3)$$

其中, $X(\tau_l, t)$ 和 $Y(\tau_l, t)$ 是具有相同方差 σ_0^2 和均值 $\mu_{x,l}$ 、 $\mu_{y,l}$ 的独立同分布的平稳时变高斯随机变量^[9]。对于室外环境, $l=0$ 表示有视距情况, $l=1, \dots, L-1$ 表示无视距情况。在此基础上, $h(\tau_l, t)$ 可以进一步表示为

$$h(\tau, t) = \sum_{l=0}^{L-1} A(\tau_l, t) = \sum_{l=0}^{L-1} (X_l(\tau_l, t) + jY_l(\tau_l, t)) = X_0(\tau_0, t) + \dots + X_{L-1}(\tau_{L-1}, t) + jY_0(\tau_0, t) + \dots + jY_{L-1}(\tau_{L-1}, t) = X(\tau, t) + jY(\tau, t) \quad (4)$$

其中, $X(\tau, t)$ 和 $Y(\tau, t)$ 是具有相同方差 $\sigma^2 = L\sigma_0^2$ 的平稳时变高斯随机变量。它们的均值分别是 $\mu_x = \mu_{x,0}$ 、 $\mu_y = \mu_{y,0}$ 。

这意味着多径信道的振幅 $\bar{h}(\tau, t)$ 遵循莱斯分布,使用概率密度函数。

$$\frac{\bar{h}(\tau, t)}{\sigma_0^2} e^{-\frac{\bar{h}^2(\tau, t) + \mu_{x,0}^2 + \mu_{y,0}^2}{2\sigma_0^2}} I_0 \left(\frac{\bar{h}(\tau, t) \sqrt{\mu_{x,0}^2 + \mu_{y,0}^2}}{\sigma_0^2} \right)。$$

其中, $I_0(\cdot)$ 是第一类零阶修正贝塞尔函数。以 dB 或者 dBm 为单位的接收信号强度可以表示为^[9]

$$RSS = 20 \log \bar{h}(\tau, t) + P_x \quad (5)$$

P_x 是发射功率和时间符号。为简单起见, $\bar{h}(\tau, t)$ 中的时间符号被忽略。

1.2 自相关性

本文根据时变多径衰落信道和广义平稳不相关散射随机过程的标准进行建模。每个信道抽头的平均功率遵循 IEEE^[10]推荐的指数衰减功率

时延曲线和钟形多普勒功率谱。本文中应用的室外交通环境拥有在 $l=0$ 时的视距成分和 $l=1, \dots, L-1$ 时的非视距成分。因此，信道模型是莱斯衰落信道模型，它是瑞利衰落信道和纯多普勒信道模型的组合。它的功率谱密度分布函数在瑞利衰落信道的功率谱分布的基础上增加了纯多普勒模型的直视路径。通常情况下，直视路径分量被设置在有效频带的零频移处。归一化多普勒功率谱密度公式如下。

$$S(f) = \frac{2\pi\sqrt{A}}{2\pi^2 f_d + k\sigma_0^2 \sqrt{A}} \left(\frac{1}{1 + A \left(\frac{f}{f_d} \right)^2} + \frac{k\sigma_0^2 \delta(f)}{2\pi} \right) \quad (6)$$

其中， A 是一个常数。在 IEEE 802.11 信道中， A 的值被设置为 9。 σ_0 是在散射波的每个入射角处入射波的平均功率。 k 是莱斯因子，它表示直接路径和间接路径的功率比。归一化多普勒功率谱密度和归一化时间形成快速傅里叶逆变换对，信道模型的自相关函数如式(7)所示。

$$R(\Delta t) = \frac{2\pi\sqrt{A}}{2\pi^2 f_d + k\sigma_0^2 \sqrt{A}} \left(\frac{\pi f_d}{\sqrt{A}} e^{-\frac{2\pi f_d}{\sqrt{A}} \Delta t} + \frac{k\sigma_0^2}{2\pi} \right) \quad (7)$$

为了简化分析，在建模过程中，假设所有信道抽头都具有相同的归一化多普勒功率谱密度。这里，多普勒频移 f_d 取决于 Alice 和 Bob 之间的速度， $f_d = \frac{\Delta v}{c} f_0 = \frac{|v_A - v_B|}{c} f_0$ ，其中 c 是光速。 f_0 是被设置为 2.4 GHz 的载波频率。在计算中，莱斯因子取 0.1。在此基础上，相关时间可以进一步推导为

$$R(\Delta v, \Delta t) = \frac{2\pi c \sqrt{A}}{2\pi^2 \Delta v f_0 + k\sigma_0^2 c \sqrt{A}} \left(\frac{\pi \Delta v f_0}{c \sqrt{A}} e^{-\frac{2\pi \Delta v f_0}{c \sqrt{A}} \Delta t} + \frac{k\sigma_0^2}{2\pi} \right) \quad (8)$$

如图 2 所示，自相关函数 (ACF) 与时延和速度差相关，并且随着时延和速度差的增加而下降，其值域为 0 到 1。本文认为上述相关性大于 60%，设置时延为 10 ms 时，速度差应该小于 60.5 km/h。下文中，将进一步研究在速度差小于 60.5 km/h 的假设下无线密钥生成的性能。

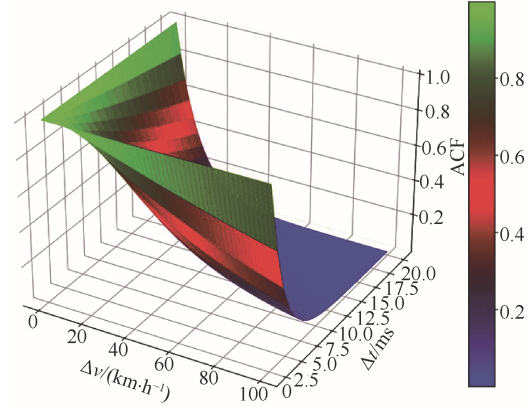


图 2 关于时延和速度差的自相关函数
Figure 2 Autocorrelation function related to time delay and speed difference

2 安全车辆密钥生成系统

基于车辆网的物理层密钥生成流程（如图 3 所示）由 4 个阶段组成：信道探测、数据预处理和量化、基于 AE 的信息协调和隐私放大^[11]。

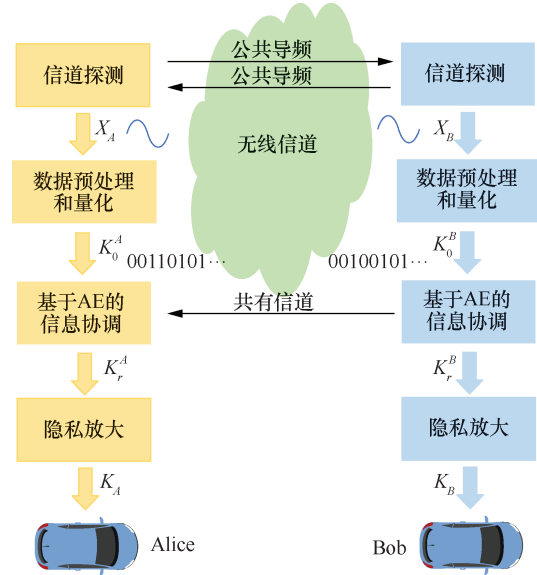


图 3 基于车辆网的物理层密钥生成流程
Figure 3 Physical layer key generation process based on vehicle network

2.1 信道探测值的获取

ESP32 是集成 Wi-Fi 功能的主控芯片，可用于模拟车内的通信节点网络。本文使用 ESP32 作为无线接入点和站。在 Visual Studio Code 集成开发环境中将 PlatformIO 添加为扩展，通过 USB 数据线烧录代码至 ESP32 开发板。如图 4 所示，两台笔记本电脑控制 ESP32 通信，并在串口处获

取源源不断的接收信号强度。

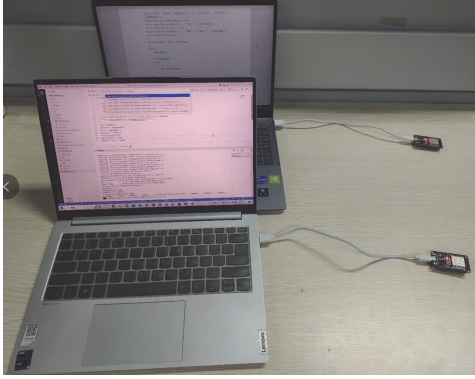


图 4 ESP32 模拟通信节点
Figure 4 Communication nodes simulated by ESP32

2.2 数据预处理

Alice 和 Bob 之间的通信具有大尺度衰落，因此在探测过程中检测到的接收信号强度会增加不稳定的偏差，导致量化后初始密钥中常出现大量连续的 0 或 1。可以将检测到的接收信号强度减去计算出的滑动平均值 v_t ，如式(9)所示。

$$v_t = \beta \cdot v_{t-1} + (1 - \beta)\theta_t \quad (9)$$

其中， θ_t 是第 t 个序列数据，而 β 是衰减系数。由于滑动平均值和实际数据之间的差异在一开始就很大，因此使用式(10)中的 v_{biased_t} 来校正平均值的估计。

$$v_{\text{biased}_t} = \frac{v_t}{1 - \beta^t} \quad (10)$$

2.3 基于累积分布函数的差分量化

预处理后，使用基于累积分布函数的差分方法进行量化，以获得初始密钥 K_q^A 和 K_q^B 。

基于累积分布函数，将 Alice 的测量值 $X_A(i)$ 量化至 $K \triangleq 2^{m_i+2} = 4 \times 2^{m_i}$ 个同样可能的量化等级，在数值空间上形成 $4 \times 2^{m_i} - 1$ 个阈值。同理，对 Bob 的测量值做相同操作。累积分布函数量化算法的阈值是根据测量值的分布来计算的，因此它可以确保量化密钥中 1 和 0 的比例是均匀的。为了达到同样可能的量化水平，需要 $X_A(i)$ 的分布。一般来说，设 $F_i(X_A) = P(X_A(i) \leq x)$ 为 $X_A(i)$ 的累积分布函数，常用的模拟分布类型是正态分布。通过使用累积分布函数的逆函数，借助式(11)，获取各量化仓的阈值。

$$\eta_k = F_i^{-1}\left(\frac{k}{4 \times 2^{m_i}}\right), 1 \leq k \leq K-1 \quad (11)$$

其中， η_k 表示第 k 个量化仓的阈值，默认 $\eta_0 = -\infty$ ， $\eta_{K-1} = +\infty$ 。

基于累积分布函数的差分量化，以量化仓序号为比较标准，将值域的变化趋势量化为初始密钥，其算法如下。

算法 1 基于累积分布函数的差分量化

输入 测量值 $X_A(i)$ 、量化等级 m 、量化长度 N

输出 Alice 得到的密钥序列 K_q^A

- 1) $F_i(X_A) = P(X_A(i) \leq x)$
- 2) **for** $j=1$ **to** N :
- 3) $\eta_j = F_j^{-1}\left(\frac{j}{4 \times 2^m}\right)$ //量化仓的阈值
- 4) 根据 η_j 判断 Alice 的量化仓序号 $k_A(i)$
- 5) **end for**
- 6) **for** $m=1$ **to** $N-1$
- 7) **if** $k_A(m) < k_A(m+1)$ **then**
- 8) $K_q^A(m) = 1$
- 9) **else if** $k_A(m) > k_A(m+1)$ **then**
- 10) $K_q^A(m) = 0$
- 11) **else** $K_q^A(m) = -1$
- 12) **end if**
- 13) **end for**
- 14) **for** $n=1$ **to** $N-1$
- 15) **for** $k=n$ **to** $N-2$
- 16) **if** $K_q^A(n) = -1$ 且 $k_A(k+1) < k_A(k+2)$
- 17) **then** $K_q^A(n) = 1$
- 18) **else if** $K_q^A(n) = -1$ 且 $k_A(k+1) > k_A(k+2)$
- 19) **then** $K_q^A(n) = 0$
- 20) **else continue**
- 21) **end if**
- 22) **end for**
- 23) **end for**

基于累积分布函数的值域量化通过量化仓序号直接赋值，受接收信号强度在值域空间分布的影响较大。单门限量化仅计算一个量化阈值，简

单地比较接收信号强度与其大小关系来进行量化步骤。相比单门限量化、基于累计分布函数的值域量化，基于累积分布函数的差分量化具有更好的初始密钥随机性。

2.4 信息协调方案对比

2.4.1 Slepian-Wolf LDPC 码

低密度奇偶校验 (LDPC, low-density parity-check) 码是一种传统的纠错码，它使用稀疏矩阵来检查和纠正错误。Baksi 等^[12]提出的 Slepian-Wolf LDPC 码基于信道估计中差异的对数似然比 (LLR, logarithm likelihood ratio) 计算，

以实现信息协调。

如图 5 所示，Slepian-Wolf LDPC 码使用两个独立的 LDPC 码块。Alice 使用第一个 LDPC 编码器矩阵生成辅助信息 p ，并使用第二个 LDPC 编码器矩阵进行封装。其输出矢量由二进制相移键控 (BPSK, binary phase shift keying) 调制并发送给 Bob。Bob 根据自己的观测结果计算 Alice 测量的 LLR。同时，Bob 对 BPSK 信号进行解调，并计算辅助信息 p 的 LLR。Bob 将信道测量的 LLR 和辅助信息的 LLR 连接起来。第一个 LDPC 编码器矩阵用于 Slepian-Wolf 解码以确定密钥 K_A 关联的密钥 K_B ^[12]。

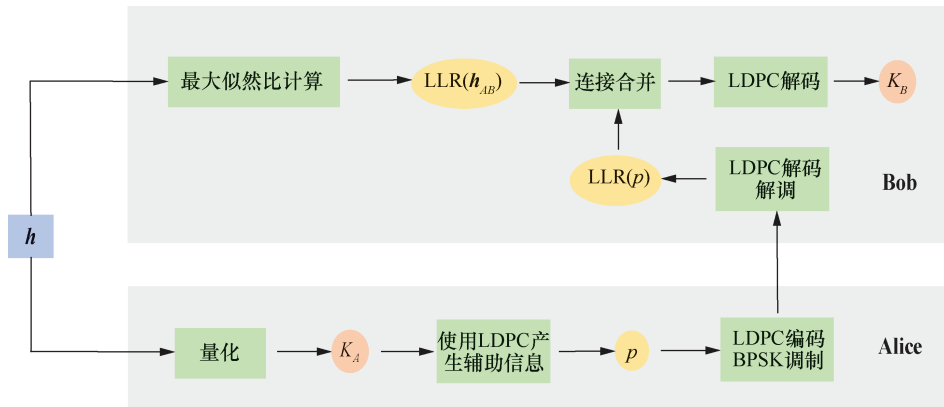


图 5 使用 Slepian-Wolf LDPC 码实现信息协调
Figure 5 Reconciliation using Slepian Wolf LDPC code

2.4.2 自动编码器

Yang 等^[4]提出的基于 LoRa 的车辆密钥方案使用自动编码器实现密钥纠错。信息协调过程中的信息传递将带来安全性隐患^[13]。在纠错模型的基础上，本文增加了一个新的损失函数来测量 Eve 可以窃听的信息泄露，期望在可靠性和保密性之间取得平衡。

基于自动编码器的信息协调流程如图 6 所示，本文设计了一个具有两个输入和两个输出的自动编码器。第一个矢量输入是 Alice 的消息位 K_{Alice} 和随机位 S_{Alice} 的结合。对于 Alice 来说， K_{Alice} 包含 k 位信息，可以表示为 $K_{\text{Alice}} = \{m_A^{(1)}, m_A^{(2)}, \dots, m_A^{(k)}\}$ 。 S_{Alice} 包含 r 个随机位，可以表示为 $S_{\text{Alice}} = \{s_A^{(1)}, s_A^{(2)}, \dots, s_A^{(r)}\}$ 。这同时适用于 Bob。网络的两个输出是失配估计 $\Delta(\text{Alice}, \text{Bob})$ 和高维码字 y_{Alice} 。前者可用于计算比特不一致的损失^[14]，而后者用于计算 Eve 可以获取的密钥泄露 $I(M; Z)$ 。

本文设计的自动编码器有两个编码器部分与一个解码器部分，它们被设置为前向反馈神经网络。除最后一层外，所有层都具有 ReLU 激活功能。Alice 和 Bob 将初始密钥输入它们的全连接层中，并通过归一化层实现高维码字 $y_{\text{Alice}} = f_1(K_{\text{Alice}})$ 和 $y_{\text{Bob}} = f_2(K_{\text{Bob}})$ 。本文将加性高斯白噪声信道添加到码字传输信道中。因此，相应的噪声层被添加到神经网络中。在接收到受噪声影响的码字 $y'_{\text{Bob}} = f_2(K_{\text{Bob}}) + f$ 之后，Alice 将它和自己的码字 y_{Alice} 作差来获取失配码字向量 $h = y'_{\text{Bob}} - y_{\text{Alice}}$ ，然后将其用作解码器的输入。传递 h 而不是 K_{Bob} 的目的是提高传输安全性。其次，向量 h 的尺寸比 K_{Bob} 更小、更加节能^[4]。密集连接层的输出是失配的估计 $\Delta(\text{Alice}, \text{Bob})$ ，这也是 $K_{\text{Alice}} \oplus K_{\text{Bob}}$ 的估计值。Alice 计算 $\Delta(\text{Alice}, \text{Bob}) \oplus K_{\text{Alice}}$ 来获取 Bob 的初始密钥 K_{Bob} 。Alice 的码字 y_{Alice} 还将用于计算第二个损失函数中在 Eve 处可获取的信息泄露量。由于

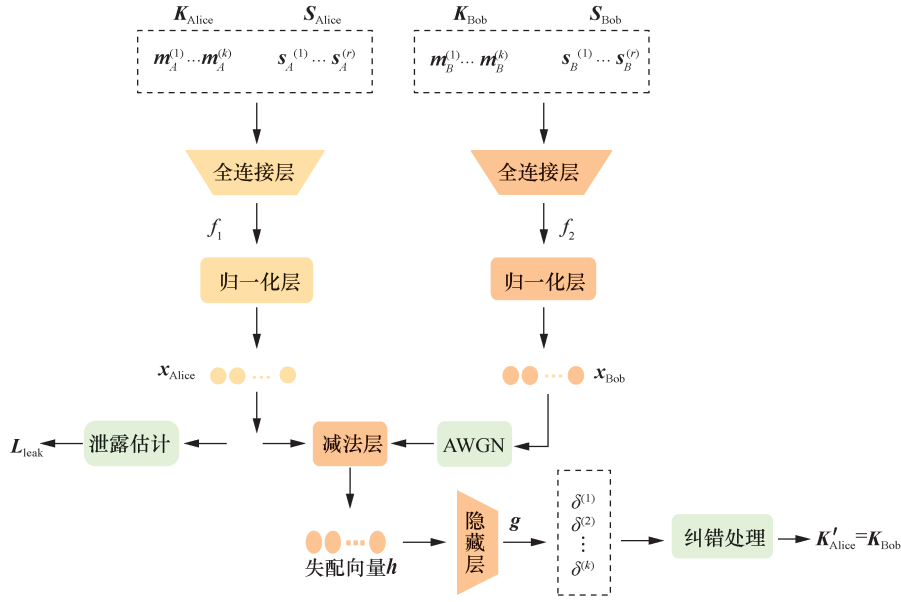


图 6 基于自动编码器的信息协调流程
Figure 6 Reconciliation process based on AE

Alice 和 Bob 是车辆网络中的独立设备，因此可以在 Alice 和 Bob 的两端独立执行训练过程。

在神经网络损失函数的构造中，两个损失函数 L_{message} 和 L_{leak} 被用于描述神经网络中的惩罚机制；引入多目标规划问题 $\min_C w_b L_{\text{message}} + w_e L_{\text{leak}}$ 。此外，可以使用如下组合损失函数的方法^[15]。

$$\min_C -(\log(w_b - L_{\text{message}}) + \log(w_e - L_{\text{leak}})) \quad (12)$$

w_b 和 w_e 分别是分配给比特不一致率（BDR, bit disagreement rate）和信息泄露 $I(M; Z)$ 的损失函数的权重。在训练过程中，Alice 和 Bob 的输入集都用于最小化损失函数 L_{message} ，其表达式如下。

$$\arg \min_{f_1, f_2, g} \|\Delta(\text{Alice}, \text{Bob}) - (K_{\text{Alice}} \oplus K_{\text{Bob}})\|^2 \quad (13)$$

Eve 的接收信号 Z 的分布可以被认为高斯混合的。在 Eve 处可获得的信息泄露可以由 Jensen 不等式来约束^[16]。

$$I(M; Z) \leq k + n - \frac{1}{2^{k+r}} \sum_{\mu_i \in C} \log \sum_{\mu_j \in C} p_{\mu_j}(\mu_i) + \frac{1}{2^{k+r}} \sum_{m \in M} \sum_{\mu_j \in C_M} \log \sum_{\mu_j \in C_M} \tilde{p}_{\mu_j, a}(\mu_i) \quad (14)$$

其中， k 是真实信息位数， n 是码字的长度。 $\mu_i \in C$ 是 Z 的每个高斯分量的平均值， p_{μ_j} 和 $\tilde{p}_{\mu_j, a}$ 是与

Alice 和 Eve 之间噪声方差有关的概率函数。

一旦建立了合适的模型，就可以在预测阶段使用神经网络编码器，从而减少计算负荷。

2.5 隐私放大

密钥纠错后，由于在纠错过程中 Alice 和 Bob 相互通信的信息在一定程度上反映了密钥的特性，即泄露了密钥的信息量，故在提取安全密钥时需将这些泄露的信息量减除^[17]。

3 实验结果和数据分析

本文实验利用 ESP32 与笔记本电脑，在道路上模拟车辆短距离通信。实验获取的接收信号强度作为无线信号的“指纹”，携带反映环境特征的信息^[18]。ESP32 的工作频率为 2.4 GHz，通信双方 Alice 和 Bob 分别放置在移动的小车上，Eve 和 Alice 的距离超过半个波长。在通道探测阶段，时延 $\Delta t = 10 \text{ ms}$ 。从 ESP32 提取到的接收信号强度将通过基于累积分布函数的差分方法实现量化，并与单门限量化、基于累积分布函数的值域量化在初始密钥随机性上对比性能。随后，对初始密钥进行基于 LDPC 码和自动编码器的信息协调方案。本文着重关注信息协调步骤的最终密钥不一致率。实验发现，基于安全车辆密钥生成系统，不同尺寸的自动编码器可以实现安全性和可靠性

的权衡。与 Slepian-Wolf LDPC 编码相比，基于自动编码器的协调方案纠错性能获得明显提升。

3.1 量化性能比较

对接收信号强度预处理后，利用基于累积分布函数的差分方法对接收信号强度量化。

随机性测试结果如表 1 所示，选取美国国家标准与技术研究院的 5 种随机性测试，对户外移

动场景下采集的接收信号强度分别进行单门限量化、基于累积分布函数的值域量化和基于累积分布函数的差分量化。可以发现，基于累积分布函数的差分量化通过了所有的测试，而单门限量化和基于累积分布函数的值域量化都通过 2 种测试。在 3 种量化方案都通过的测试中，本文提出的量化方法拥有相对更高的 p 值。

表 1 随机性测试结果
Table 1 Randomness test results

随机性测试	二元矩阵秩检验测试值	二元矩阵秩检验通过情况	重叠模块匹配检验测试值	重叠模块匹配检验通过情况	线性复杂度检验测试值	线性复杂度检验通过情况	随机游动检验测试值	随机游动检验通过情况	随机游动状态频率检验测试值	随机游动状态频率检验通过情况
单门限量化	2.338×10^{-6}	不通过	0.000 2	不通过	1.360×10^{-37}	不通过	0.699 9	通过	0.617	通过
基于累积分布函数的值域量化	2.338×10^{-6}	不通过	0.000 18	不通过	1.5247×10^{-8}	不通过	0.776 5	通过	1	通过
基于累积分布函数的差分量化	0.413 5	通过	0.126 1	通过	0.410 8	通过	0.849 1	通过	1	通过

3.2 比特不一致率对比

在车辆网络的背景下，本节讨论了 Alice 和 Bob 对于不同比特一致率的初始密钥的纠错性能。经过实验验证，基于自动编码器的方案纠错性能优于 Slepian-Wolf LDPC 码^[12]。

在本文提出的信息协调方案中，设计了一个 (k, r, m) 码，它表示一个具有 k 位信息、 r 位随机位和从编码器输出的 m 位码字的神经网络。本文研究了 $k=6, m=3$ 的情况。这意味着 6 位信息位被压缩成具有 3 个比特位的低维码字。本文设计了全连接层分别为 1 层、2 层、3 层的 3 个自动编码器 AE1、AE2、AE3。神经网络的 Alice 和 Bob 的训练集的初始不一致比特率先后设置为 10%、5%。

选取了 20 000 位的测试集进行方案对比验证。Slepian-Wolf LDPC 码和自动编码器的纠错性能对比如图 7 所示，左侧是初始比特不一致率为 0.1 左右时的性能对比，右侧是初始比特不一致率为 0.05 左右时的情形。这里，使用具有最佳性能的 AE3 方案，其中权重 w_b 设置为 1.0。从图 7 可以看出，基于自动编码器的信息协调方案在不同初始密钥生成场景下都优于 Slepian-Wolf LDPC 码，前者的钥不一致率相比后者降低 30% 左右，相比较于初始密钥实现了 50% 左右的位纠错。

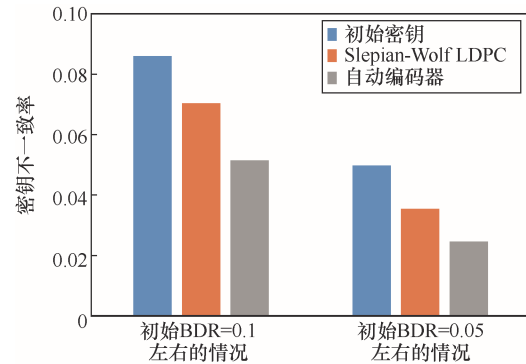


图 7 Slepian-Wolf LDPC 码和自动编码器的纠错性能对比
Figure 7 Comparison of error correction performance between Slepian-Wolf LDPC Code and AE

3.3 可靠性和保密性的平衡

本文提出的信息协调方案的一个优点是对安全性的定量考虑。该方案设计针对比特不一致率和信息泄露的加权损失函数。作为输入的 r 位随机消息用于保护 k 位秘密消息，也可以称为混淆消息。基于双输入、双输出自动编码器的协调方案能够计算 Eve 可用的信息泄露的上界，并在训练过程中寻求泄露少的编码结构。

本文提出的信息协调方案引入了多目标规划问题。因此，可以通过调整两个损失函数的权重比来实现纠错性能和信息泄露之间的平衡，实现对不同需求的环境的自适应信息协调。 w_b 设置为

0.9~1.0, 而 $w_e = 1 - w_b$ 。不同自动编码器的性能平衡如图 8 所示, 考虑了神经网络自动编码器的 3 种结构, 该方案将 6 位信息成功压缩为 3 位, 并且设置训练集的比特不一致率为 0.2。可以发现, 比特不一致率与信息泄露之间存在负相关。作为最复杂的神经网络, AE3 具有最好的综合性能。AE1 是一个具有最简单密集连接层的神经网络, 其相对性能实际上是最差的。

可以认为, 具有更好性能的自动编码器可以由更大尺寸的神经网络规模来实现。然而, 神经网络规模的增加会带来更高的计算复杂度。

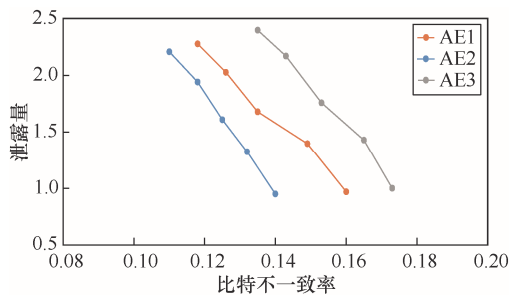


图 8 不同自动编码器的性能平衡

Figure 8 Performance balance of different auto-encoders

3.4 学习曲线研究

图 9 描述了比特不一致率和迭代次数的关系。这里, 选择 AE3 自动编码器进行训练, 以获得迭代次数与比特不一致率之间的关系曲线。可以发现, 比特不一致率随着迭代次数的增加而降低。对于 AE3, 当迭代次数达到大约 500 时, 比特不一致率开始收敛。当迭代次数足够大, 最终纠错后的比特不一致率可以达到 11% 左右。当然, 具有高迭代次数的模型训练取决于计算能力。由于网络结构的限制, BDR 不能无限地下降。

实现 6 维到 3 维压缩的神经网络在迭代次数较小时开始收敛。在实现从 8 维压缩至 3 维的一种神经网络时, 经过实验, 发现其开始收敛的迭代次数在 1 500 左右。这说明, 通过更大尺寸的神经网络实现的更长信息位的传输需要更多的迭代次数, 以实现比特一致率性能的最终稳定提升。

3.5 隐私放大

本文实验采用普适哈希函数族实现隐私放大, 分别选用 SHA3、SM3、BLAKE2b 这 3 种哈希算法。Alice 从公共且已知的普适哈希函数族中

随机选取一个哈希函数, 并通过一个可信的公共信道传输给 Bob。Alice 和 Bob 利用哈希函数作用于信息协调后的密钥以得到足够安全的密钥。

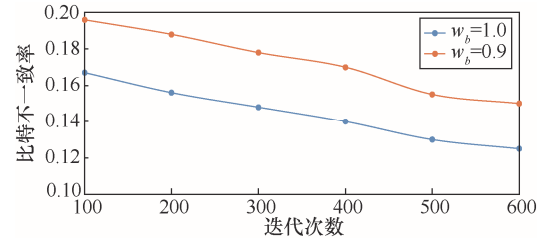


图 9 比特不一致率和迭代次数的关系

Figure 9 Relationship between BDR and number of training epochs

4 结束语

本文研究了车联网中基于深度学习的自适应无线密钥生成系统。首先, 基于多径衰落和视距对车辆网络进行建模, 并分析了速度差和时延对自相关性的影响。实验数据基于 ESP32 设备, 经过数据预处理后供量化步骤使用。本文提出了一种基于累计分布函数的差分量化方法, 并与单门限量化、基于累积分布函数的值域量化在生成密钥随机性上对比。本文提出的量化方案以优异的结果通过了最多的随机性测试。在密钥生成的信息协调步骤中, 本文提出了一种基于神经网络自动编码器的信息协调方案, 与 Slepian-Wolf LDPC 码相比, 该方案降低 30% 左右的密钥不一致率。基于自动编码器的协调方案还可以实现动态平衡纠错性能和信息泄露的多环境需求。

参考文献:

- [1] SHAWKY M A, USMAN M, IMRAN M A, et al. Adaptive chaotic map-based key extraction for efficient cross-layer authentication in VANETs[J]. Vehicular Communications. 2023, 39:4-5.
- [2] CHENG L, STANCIL D D, BAI F, et al. A roadside scattering model for the vehicle-to-vehicle communication channel[J]. IEEE Journal on Selected Areas in Communications, 2013, 31(9): 449-459.
- [3] MARGELIS G, FAFOUTIS X, OIKONOMOU G, et al. Physical layer secret-key generation with discrete cosine transform for the Internet of Things[C]//Proceedings of IEEE International Conference on Communications (ICC). 2017: 2-4.
- [4] YANG H Q, LIU H B, LUO C W, et al. Vehicle-key: a secret key establishment scheme for LoRa-enabled IoV communications[C]//Proceedings of IEEE 42nd International Conference on Distributed Computing Systems. 2022: 5-7.

- [5] LI K, LU L Y, NI W, et al. Secret key agreement for data dissemination in vehicular platoons[J]. IEEE Transactions on Vehicular Technology, 2019, 99: 1.
- [6] LI S, MA Y F, GU X, et al. Wi-Fi-based device-free vehicle speed measurement using fast phase correction MUSIC algorithm[C]//Proceedings of 2020 International Conferences on Internet of Things (iThings). 2020: 9060-9062.
- [7] LIU H, WANG Y, REN Y, et al. Bipartite graph matching based secret key generation[C]//Proceedings of IEEE Conference on Computer Communications. 2021.
- [8] ZHANG J Q, MARSHALL A, WOODS R, et al. Efficient key generation by exploiting randomness from channel responses of individual OFDM subcarriers[J]. IEEE Transactions on Communications, 2016, 64(6): 2579-2580.
- [9] PAN Y J, XU Z Q, LI M, et al. Man-in-the-middle attack resistant secret key generation via channel randomization[C]//Proceedings of International Symposium on Mobile Ad Hoc Networking and Computing. 2021: 231-240.
- [10] ERCEG V, SCHUMACHER L, KYRITSIS P, et al. TGn channel models[S]. IEEE Standard 802.11-03/940r4, 2004.
- [11] ZHANG J Q, LI G Y, MARSHALL A, et al. A new frontier for IoT security emerging from three decades of key generation relying on wireless channels[J]. IEEE Access, 2020, 8(99): 138410-138412.
- [12] BAKSI S, SNOAP J, POPESCU D C. Secret key generation using one-bit quantized channel state information[C]//Proceedings of Wireless Communications Networking Conference. 2017.
- [13] RANA V, CHOU R A. Design of short block length wiretap channel codes: deep learning and cryptography working hand in hand[C]//Proceedings of IEEE Information Theory Workshop (ITW). 2021.
- [14] LIN Q, XU W T, LIU J, et al. H2b: heartbeat-based secret key generation using piezo vibration sensors[C]//Proceedings of the International Conference on Information Processing in Sensor Networks(IPSNS). 2019:265-267.
- [15] BESSER K L, LIN P H, JANDA C R, et al. Wiretap code design by neural network autoencoders[J]. IEEE Transactions on Information Forensics and Security. 2019, 5(99): 3377-3378.
- [16] ARTEMY K, BRENDAN T. Estimating mixture entropy with pairwise distances[J]. Entropy, 2017, 19(7): 361-362.
- [17] 李默, 张春梅, 银振强, 等. 量子密钥分配后处理概述[J]. 密码学报, 2015(2): 9.
- LI M, ZHANG C M, YIN Z Q, et al. Overview of post-processing for quantum key distribution[J]. Journal of Cryptologic Research,

2015(2):9.

- [18] 杨铮, 刘云浩. Wi-Fi 雷达: 从 RSSI 到 CSI[J]. 中国计算机协会通讯, 2014, 10(11): 55-57.

YANG Z, LIU Y H. Wi-Fi radar: from RSSI to CSI[J]. Communications of the China Computer Federation, 2014, 10 (11): 55-57.

[作者简介]



汪涵 (1999-), 男, 江苏盐城人, 东南大学硕士生, 主要研究方向为无线物理层密钥生成技术。



陈立全 (1976-), 男, 东南大学教授、博士生导师, 主要研究方向为移动信息安全、物联网系统与安全。



王忠民 (1974-), 男, 博士, 南京医科大学教授, 主要研究方向为医学信息学、医院信息系统、网络安全。



陆天宇 (1994-), 江苏苏州人, 东南大学博士生, 主要研究方向为物理层安全、无线资源分配。