

DOI:10.3969/j.issn.1671-0673. 2025. 06. 001

自然冗余译码下的帧定界与协议识别联合处理方法

于国贤, 冉晓旻, 于宏毅, 沈智翔, 鞠易祥

(信息工程大学, 河南 郑州 450001)

摘要: 针对高误码条件下协议容错接收中, 传统分步级联处理并非最优且易引发错误扩散的问题, 提出一种链路层帧定界与网络层协议识别的跨层容错联合估计方法。该方法基于帧定界与协议识别两者共有的数据分组长度这一自然冗余规律, 构建统一的联合自然冗余译码模型, 并基于 MAP 准则推导理论最优解。为降低计算复杂度, 依据潜在似然最大化原则, 设计基于 top- k 准则的快速算法, 有效控制解空间搜索规模。在常见卫星协议栈和公开数据集上的仿真结果表明, 当误码率为 0.1 时, 该方法相比传统方案定界错误降低约 30%, 协议识别错误率下降近一个数量级, 在可接受的计算开销下显著提升系统在高误码环境下的容错恢复能力。

关键词: 容错估计; 自然冗余; 跨层联合; 路径搜索; 最大似然准则

中图分类号: TN911.22

文献标识码: A

文章编号: 1671-0673(2025)06-0631-08

Joint Processing Method for Frame Delineation and Protocol Identification Based on Natural Redundancy Decoding

YU Guoxian, RAN Xiaomin, YU Hongyi, SHEN Zhixiang, JU Yixiang

(Information Engineering University, Zhengzhou 450001, China)

Abstract: To address the suboptimality and error propagation issues of traditional cascaded processing in protocol fault-tolerant reception under high bit error rates, a cross-layer fault-tolerant joint estimation method for link-layer frame delineation and network-layer protocol identification is proposed. By exploiting the natural redundancy of data packet length common to both frame delineation and protocol identification, a unified joint natural redundancy decoding model is established, and the theoretically optimal solution is derived based on the MAP criterion. To reduce computational complexity, a fast algorithm based on the top- k criterion is designed according to the principle of maximizing potential likelihood, effectively controlling the search space. Simulation results on common satellite protocol stacks and public datasets show that, when the bit error rate is 0.1, compared with those of the traditional scheme, the frame delimitation error is reduced by approximately 30%, and the protocol identification error rate is reduced by nearly an order of magnitude through the proposed method, significantly enhancing fault-tolerant recovery capability of the the system under high error rates environments with acceptable computational overhead.

Key words: fault-tolerant estimation; natural redundancy; cross-layer joint; path search; maximum likelihood criterion

在战场通信^[1]、高速移动^[2]、偏远^[3]或封闭区域^[4]等复杂环境下, 无线链路受强干扰、衰减等影响,

常出现高误码。常规应对方法依赖降低码率^[5]、提升功率^[6]或高增益天线^[7]改善接收, 但代价高昂。

收稿日期: 2025-09-08; 修回日期: 2025-10-31

基金项目: 国家自然科学基金(61501517)

作者简介: 于国贤(2001—), 女, 硕士生, 主要研究方向为网络协议解析。E-mail: 17836957263@163.com

尤其在军事监听等非合作通信中,接收方无法建立反馈重传机制,且信号捕获受限于旁瓣强度与高成本传感资源,迫切需要不依赖反馈的高效容错接收技术。

为提升高误码环境下的传输可靠性,协议容错控制处理主要包括聚合帧定界、网络层协议识别、会话流分离与乱序重排以及应用层协议识别等环节。文献[8]从自然冗余译码的角度,提出广义信道前向编译码的概念,研究已分离会话流的多种冗余的联合译码处理。在终端接收机方面,文献[9-12]通过挖掘信源的自然冗余,结合深度学习实现了含错信源的部分恢复。在路由节点方面,文献[13]提出的基于维特比的路径搜索可实现聚合帧的容错定界,文献[14]提出的改进型均值漂移聚类算法可依据五元组字段在高误码条件下实现含噪会话流的分离。

但目前协议容错处理多局限于单一环节,类似通信系统中噪声系数的级联效应,层次化处理易导致误差逐级传播并放大。尤其在处理链较长时,系统整体性能将高度依赖前端的准确性。因此,对多个环节进行统一建模与协同优化的联合处理,是实现最优性能的有效途径。由于帧定界和协议识别均高度依赖长度字段这一共同特征,二者存在潜在耦合,因此本文将对这两个阶段开展研究。在容错定界方面,文献[15]将其建模为Markov过程,采用BCJR算法实现了理论最优,但

较高的计算复杂度使得部分步骤难以执行;文献[16]则从序列译码角度出发,通过粗定界提升了整体效率,但其定义的经验性路径度量准则损失了部分性能;文献[13]在此基础上优化路径度量,采用基于Viterbi的路径搜索方法从最大似然角度逼近了理论最优。在网络层协议容错识别上,低误码时仅根据IP协议字段即可直接确定,但在高误码环境下该字段易出错,传统方法失效,目前公开文献中尚缺乏有效的容错识别研究。

鉴于数据包容错定界与协议识别均依赖长度字段,针对高误码下物理层输出含错的问题,提出一种基于“广义信道前向译码”的联合估计方法,无需重传即可同步实现边界恢复与协议识别。在常见卫星协议栈及WIDE MAWI数据集^[17]上的实验结果表明,误码率为0.1时,该方法相比分步处理性能提升近一个数量级,具有良好的鲁棒性。

1 帧定界与协议识别容错联合处理

图1给出了 η 个不同类型的数据分组在发送端经链路层封装、接收端进行容错定界并提取IP分组的简化模型。蓝色部分为同步标志位,高级数据链路控制(HDLC)中为0x7E。每个分组都由负载和多个上层的首部及尾部组成。低误码时,仅通过同步标志位和协议号硬判决即可准确完成定界与识别;但在高误码下,这些关键信息易受干扰失效,需挖掘新的冗余辅助估计。

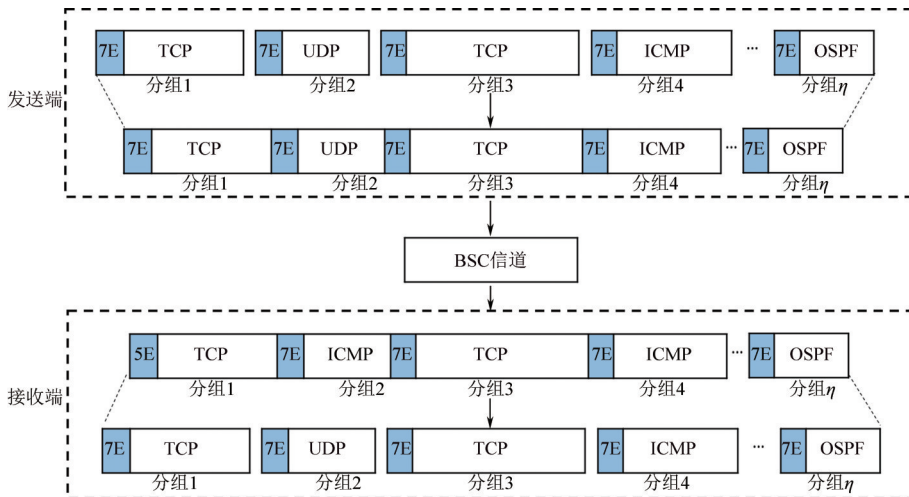


图1 链路层容错定界与IP分组提取模型

从协议栈设计角度考虑,各层首尾部字段的功能开销,受语法和语义约束形成的结构规律可视为一种自然冗余。依据功能可大致分为固定字段 k 、长度字段 u 、特征字段 r 、校验字段 c (包括首部

校验和整个数据分组的校验)、其他首部字段 o 和信源负载数据字段 b 等^[13]。值得注意的是, k 、 u 、 r 和 c 中的结构性冗余信息通过适当的编解码策略,均可有效辅助协议控制容错处理过程。

假设第 i 个分组的长度为 ξ_i bit($\xi_{\min} \leq \xi_i \leq \xi_{\max}$), η 个分组的总长度为 $\ell = \sum_{i=1}^{\eta} \xi_i$ bit,则各分组的真实起始边界 $\zeta_1, \zeta_2, \dots, \zeta_{\eta}$ 可分别记作 $1, \xi_1 + 1, \xi_1 + \xi_2 + 1, \dots, \sum_{i=1}^{\eta-1} \xi_i + 1$,其IP协议类型可记为 $\lambda_1, \lambda_2, \dots, \lambda_{\eta}$ 。帧定界与协议容错联合估计问题就是需要在数据分组总数 η 未知的情况下,仅根据接收序列 $\mathbf{y}$ 和部分已知的协议自然冗余信息来同时找出每个分组的边界和协议类型。

本文在Viterbi状态空间路径搜索^[13]的基础上,构建帧定界与协议识别的联合估计模型。通过在 $\mathbf{y}$ 上与 $\mathbf{k}$ 匹配,可滑动搜索出所有分组的潜在边界(粗节点) $\mathbf{g} = \{g_1, g_2, \dots, g_m, g_{m+1}\}$,其中 $g_{m+1} = \ell + 1$ 为帧真实终止边界。对 $\forall g_i$ (下游节点),当满足长度约束 $\xi_{\min} \leq g_i - g_j \leq \xi_{\max}, j \in [1, 2, \dots, i-1]$ 时,可确定其上游节点集合 $\{g_i\}_{\text{up}}$ 。其中,任意上下游节点对 $(g_a, g_b), g_a \in \{g_b\}_{\text{up}}$ 不仅唯一确定了某预测分组的起止位置,还可唯一提取其中的 $\mathbf{k}, \mathbf{u}$ 和 $\mathbf{r}$,并拼接为已知矢量 $\tilde{\mathbf{k}}_{\text{ku}}^a$ 。同时, g_a 到 g_b 之间的节点都不构成真实边界,对应字段非冗余信息,不要求可靠传输,可将其建模为多个与 $\tilde{\mathbf{k}}_{\text{ku}}^a$ 大小相同的0/1等概率序列 $\tilde{\mathbf{R}}_{\text{ku}}^i$ 。因此可定义字段选择变量 $\tilde{s}_{\text{ku}}^i \in \{\tilde{\mathbf{k}}_{\text{ku}}^i, \tilde{\mathbf{R}}_{\text{ku}}^i\}$ 来表示对粗节点的判别,若 $\tilde{s}_{\text{ku}}^i = \tilde{\mathbf{k}}_{\text{ku}}^i$,则认为该节点为有效边界,否则为无效边界。当且仅当 (g_i, g_j) 满足 $\tilde{s}_{\text{ku}}^i = \tilde{\mathbf{k}}_{\text{ku}}^i, \tilde{s}_{\text{ku}}^j = \tilde{\mathbf{k}}_{\text{ku}}^j$,且 $\tilde{s}_{\text{ku}}^h = \tilde{\mathbf{R}}_{\text{ku}}^h, h = i+1, \dots, j-1$ 时, $[g_i, g_j]$ 才可被判定为一个完整的分组,其类型可记为 $\tau(g_i, g_j) = \lambda_i \in T$,其中 T 为所有类型可能情况。

因此,最佳路径即联合最大后验解为

$$\left(\hat{\mathbf{s}}_{\text{ku}}^1, \dots, \hat{\mathbf{s}}_{\text{ku}}^m, \lambda_1, \dots, \lambda_{\eta} \right) = \arg \max_{\substack{\tilde{s}_{\text{ku}}^i \in \{\tilde{\mathbf{k}}_{\text{ku}}^i, \tilde{\mathbf{R}}_{\text{ku}}^i\}, \forall i \in [1, m], \\ \lambda_j \in T, \forall j \in [1, \eta]}} P\left(\tilde{s}_{\text{ku}}^1, \dots, \tilde{s}_{\text{ku}}^m, \lambda_1, \dots, \lambda_{\eta} | \mathbf{y}\right) \quad (1)$$

最优解需在满足帧长约束的路径中搜索,并联合遍历相邻有效粗节点界定的协议类型,但穷举所有组合的暴力搜索复杂度高达 $2^m \times 3^{\eta}$,难以应用。

2 基于top- k 准则的快速算法

实际上,式(1)可分解为两阶段概率乘积,即

$$\left(\hat{\mathbf{s}}, \hat{\lambda} \right) = \arg \max_{\tilde{s}, \lambda} P\left(\tilde{s}, \lambda | \mathbf{y}\right) = \arg \max_{\tilde{s}, \lambda} P\left(\tilde{s} | \mathbf{y}\right) P\left(\lambda | \mathbf{y}, \tilde{s}\right) \quad (2)$$

假定定界结果 $\tilde{s}$ 和协议类型 λ 先验等概,则可

写为最大似然(Maximum Likelihood Criterion, ML)形式,即

$$\left(\hat{\mathbf{s}}, \hat{\lambda} \right) = \arg \max_{\tilde{s}, \lambda} P\left(\tilde{s} | \mathbf{y}\right) P\left(\lambda | \mathbf{y}, \tilde{s}\right) \propto \arg \max_{\tilde{s}} \left(P\left(\mathbf{y} | \tilde{s}\right) \arg \max_{\lambda} P\left(\mathbf{y}, \tilde{s} | \lambda\right) \right) \quad (3)$$

式(3)表明,求解需先遍历所有 $\tilde{s}$,并在每个 $\tilde{s}$ 下进一步遍历所有可能的 λ 。但这种双重遍历的方式仍会导致计算复杂度居高不下。

为此,本文引入top- k 搜索策略,提出一种快速算法:首先,在定界阶段保留高概率候选路径,缩小搜索空间;其次,为每条路径筛选各分组的高概率协议类型;最后,组合生成最优联合估计路径,可在保证性能的同时显著降低计算开销。下文介绍两阶段独立处理方法,作为后续算法的铺垫。

2.1 基于Viterbi路径搜索的容错定界

文献[13]基于分组累积长度的Markov性,将帧定界建模为变步数的类卷积码:消息为分组长度 $\xi_i, \xi_{\min} \leq \xi_i \leq \xi_{\max}$,状态为累积长度,合法状态转移则由长度约束决定。通过Viterbi路径搜索,在粗节点序列 $\mathbf{g}$ 上从 g_1 节点逐步递推至 g_{m+1} ,可得到最佳路径,即最终定界结果 $\mathbf{g}_{\text{opt}}$ 。

对 $\forall (g_a, g_b), g_a \in \{g_b\}_{\text{up}}$,可定义从 g_a 转移至 g_b 的路径单步度量为各 $\mathbf{s}_{\text{ku}}^i$ 的似然之和:

$$M_{a \rightarrow b}^1 = \sum_{i=a}^{b-1} \log_2 \left(P\left(\mathbf{y}_{\text{ku}}^i | \mathbf{s}_{\text{ku}}^i\right) \right) \quad (4)$$

式中, $\mathbf{y}_{\text{ku}}^i$ 表示该分组中 $\mathbf{k}$ 和 $\mathbf{u}$ 的观测值。当 $i=a$ 时, $\mathbf{s}_{\text{ku}}^i = \mathbf{k}_{\text{ku}}^a = [\mathbf{k}, \mathbf{r}]$;否则, $\mathbf{s}_{\text{ku}}^i = \mathbf{R}_{\text{ku}}^i$ 。

则从 g_1 经 g_a 递推至 g_b 的路径累积度量值满足如下递推关系:

$$M_{1 \rightarrow b}^a = M_{1 \rightarrow a} + M_{a \rightarrow b}^1 \quad (5)$$

式中: $M_{1 \rightarrow a}$ 表示节点 g_1 至 g_a 的累积度量; $M_{1 \rightarrow b}^a$ 表示由 g_a 扩展至 g_b 后的累积度量。

由于递推时考虑了所有节点,因此即使分组数不同,各路径的似然计算次数也相同,累积度量具有可比性。可借鉴Viterbi思想,每步仅保留一条最优路径,如式(6)所示。递推至终点 g_{m+1} 后即可确定 $\mathbf{g}_{\text{opt}}$ 。

$$M_{1 \rightarrow b} = \max_{g_a \in \{g_b\}_{\text{up}}} (M_{1 \rightarrow b}^a) \quad (6)$$

2.2 基于ML的IP协议分类

鉴于IP协议种类繁多,而TCP和UDP占主流,其余多为控制报文,可将协议识别简化为3个分类:TCP、UDP和其他。IP首部中的协议字段是指示上层协议的直接冗余,长度字段也有一定的结

构约束。不同协议首部还包含特定冗余,如TCP保留位通常为0,UDP长度与IP总长度对应。除 u 外,其余固定冗余均可视作 r ,具有固定偏移和取值,可用于高误码下的协议容错识别。

由于各分组内部独立,该问题可被建模为非线性等长分组码译码:消息为协议类型 T ,发送码字由 r 和 u 拼接而成,即 $v=[r,u]$ 。然而,不同协议的冗余字段数量和位置不同,导致 v 长度不一,直接比较会因长度差异而产生偏差。为此,通过填充使所有 v 等长——在 v_A 不包含协议 B 特有的 r_B 上填充服从0/1等概率分布的随机序列,从而确保不同协议度量值之间具有可比性。

ML判决准则如式(7)所示:

$$\hat{\lambda} = \underset{\lambda \in T}{\operatorname{argmax}} P(y|\lambda) = \underset{\lambda \in T}{\operatorname{argmax}} \sum_v P(y|\lambda, v) = \underset{\lambda \in T}{\operatorname{argmax}} \sum_r \sum_u P(y|\lambda, r, u) \quad (7)$$

式中, $T = \{\text{TCP}, \text{UDP}, \text{其他}\}$ 。

理论上需在 2^4 的码字空间内暴力搜索,但 u 的真实值 L 可通过计数确定,式(7)可简化为仅遍历 r ,搜索空间可降至256种:

$$\hat{\lambda} = \underset{\lambda \in T}{\operatorname{argmax}} P(y, L|\lambda) = \underset{\lambda \in T}{\operatorname{argmax}} \sum_v P(y, L|\lambda, v) = \underset{\lambda \in T}{\operatorname{argmax}} P(y|\lambda, r, u) \quad (8)$$

式(8)基于单分组进行协议类型的估计,但在联合估计框架下, $\tilde{s}$ 为已知条件,因此可推广为

$$(\hat{\lambda}_1, \dots, \hat{\lambda}_\eta) = \underset{\lambda \in T}{\operatorname{argmax}} P(y_i, L_i|\lambda_i) = \underset{\lambda \in T}{\operatorname{argmax}} \prod_{i=1}^{\eta} \sum_r P(y_i|\lambda_i, r_i, u_i) \quad (9)$$

将2.1节与2.2节的单步处理串联后,即构成仿真实验中对比的级联方法。

2.3 快速算法

2.3.1 构建定界路径的高概率集

文献[18]在HDLC/MAC/IPv4协议栈的研究表明,误码率 $p_e = 0.1$ 时,定界错误率约为 $10^{-2} \sim 10^{-3}$ 。其中最典型的错误事件是单分组被误分为两段,即多定1帧。在错误独立假设下,1帧内两个及以上分组同时出错的概率降至 $10^{-4} \sim 10^{-6}$ 。因此可合理假设:1帧中最多仅有1个分组发生一分为二的错误。

假定Viterbi算法输出的最优路径 P_{Path_1} 中包含 s 个真节点,其中帧头和帧尾均为已确定的真实边界,其余 $(s-2)$ 个中间节点都可能为冗余插入的假节点。为构建合理的候选定界路径集合,依次移除每个中间节点,并在满足帧长约束 $\xi_{\min} \leq \xi_i \leq \xi_{\max}$

的前提下,最多可获得 $(s-2)$ 条备选路径。将原始路径 P_{Path_1} 与度量值最高的 $(K-1)$ 条备选路径合并,可构成包含 K 条路径的候选集合 $C_{\text{Path}} = \{P_{\text{Path}_1}, P_{\text{Path}_2}, \dots, P_{\text{Path}_K}\}$ 。假设第 k 条路径划分出 p 个分组,其度量值可计算为

$$M^k = \sum_{i=1}^p M_{a_i \rightarrow b_i}^1 = \sum_{i=1}^p \sum_{j=a_i}^{b_i-1} \log_2 \left(P(\tilde{y}_{ku}^j | \tilde{s}_{ku}^j) \right) \quad (10)$$

式中: M^k 为第 k 条路径的累积度量值; $M_{a_i \rightarrow b_i}^1$ 为该路径下第 i 个分组的单步度量。 $s=5$ 、 $K=2$ 时的示意图如图2所示。

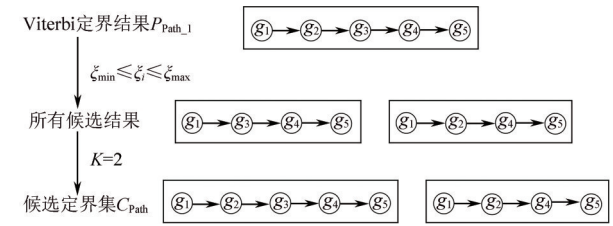


图2 C_{Path} 获取示例

该方法保留定界路径中的高概率候选集,避免对所有可行解进行穷举带来的高昂计算代价。

2.3.2 构建协议类型高概率集

获取 C_{Path} 后,对各预测分组引入其候选类型集,可生成联合估计路径 $C_{\text{Joint}} = \{P_{\text{JPath}_1}, P_{\text{JPath}_2}, \dots, P_{\text{JPath}_A}\}$,其中每条 P_{JPath_i} 不仅包含帧定界结果,还包含一组具体的协议类型分配。

具体而言,对于 P_{Path_i} 中的每个分组 P_{Pkt_i} ,依据式(9)计算各协议类型的对数似然值,并将其归一化为概率分布。设当前分组 P_{Pkt_i} 的归一化概率为 $p_1 > p_2 > p_3$,且满足 $p_1 + p_2 + p_3 = 1$ 。引入置信度阈值 θ ,若最大概率与次大概率之比 $p_1/p_2 < \theta$,则认为这两个协议类型具有较高可能性,保留为候选类型;否则,仅保留最大概率对应的类型。

该方法基于以下核心假设:当误判发生时,数据分组的类型概率主要集中在其中两个协议类型上,而在第3种类型上的概率可忽略不计。这是因为协议特征字段之间的结构性差异通常会导致映射出的码字之间具有较大的汉明距离($d_{\min} > 1$),从而使得似然值呈现指数级差异。即使汉明距离偏移较小(如 $\Delta d = 1$),也会导致归一化概率值呈现高度的非均匀特性。3类概率均匀分布的情况仅在理论边界 $d_{v_1} = d_{v_2} = d_{v_3}$ 下才可能出现。

设每条路径中第 i 个数据分组的候选类型数目为 $\varepsilon_i \in \{1, 2\}$,其选择规则为

$$\varepsilon_i = \begin{cases} 2, & p_1/p_2 < \theta; \\ 1, & \text{其他}. \end{cases}$$

若共构建出 K 条候选定界路径, 每条路径中包含 η_j 个数据分组, 则最终候选联合路径总数为

$$N_{\text{Path}} = \sum_{j=1}^K \prod_{i=1}^{\eta_j} \varepsilon_i^j \quad (11)$$

随着 p_e 的升高, $p_1/p_2 < \theta$ 的事件概率 q 逐渐增大, 将导致候选联合路径数目呈指数增长:

$$N_{\text{Path}} = k \times \prod_{i=1}^{\eta} \varepsilon_i \approx k \times (1+q)^{\eta} \quad (12)$$

2.3.3 联合搜索估计最优解

相比单纯定界, 在固定字段的基础上引入协议特征字段 $[r, u]$ (如 IP 协议、TCP 保留位、紧急指针及 UDP 长度), 并在已知协议类型的条件下利用这些字段参与度量计算, 从而实现定界与协议的联合估计。获取 $C_{\text{Joint}} = \{P_{\text{JPath}_1}, P_{\text{JPath}_2}, \dots, P_{\text{JPath}_{N_{\text{Path}}}}\}$ 后, 可依据式 (10) 重新计算 P_{JPath_i} 的累积度量值, 评估其作为联合估计结果的可能性。其中 P_{JPath_i} 中由 (g_a, g_b) 组成的第 j 个分组的单步度量可表示为

$$M_{a \rightarrow b}^1 = \sum_{i=a}^{b-1} \log_2 \left(P(\tilde{y}_{ku}^i | \tilde{s}_{ku}^i) \right) \quad (13)$$

最终输出为 C_{Joint} 中累积度量值最大者, 即

$$P_{\text{JPath}_{\text{opt}}} = \underset{P_{\text{JPath}_i} \in C_{\text{Joint}}}{\operatorname{argmax}} M^k \quad (14)$$

帧定界与协议识别联合处理过程如算法 1 所示。

算法 1 帧定界与协议识别联合处理算法

输入: k, r, γ, K, θ 。

输出: $\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_{\eta}, \lambda_1, \lambda_2, \dots, \lambda_{\eta}$ 。

1. 执行 Viterbi 路径搜索获得 P_{Path_1}

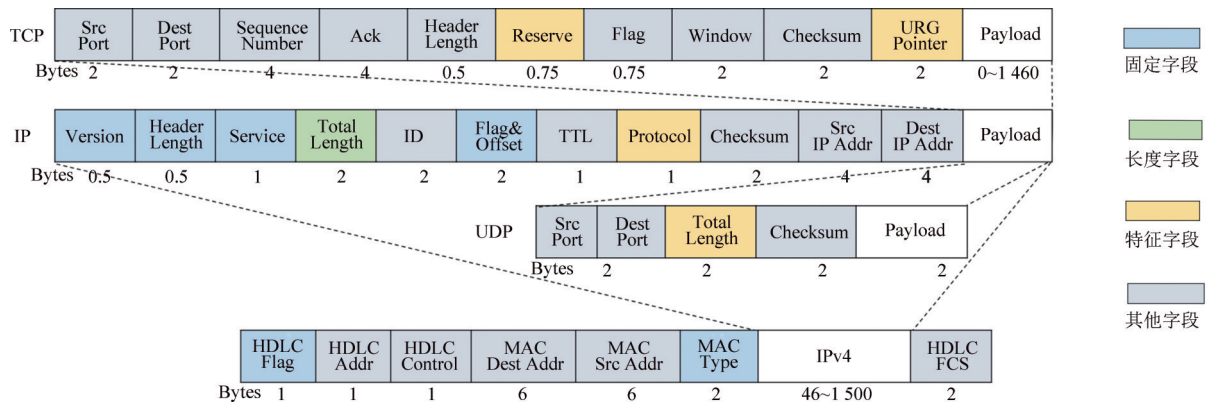


图3 HDLC/MAC/IPv4协议栈

3.1 算法关键参数的优化选择

3.1.1 不同 θ 的影响

为选取合适的 θ , 在 $\theta \in \{1, 3, 5, 7, 9\}$ 的条件下进行了多组仿真实验。式 (12) 指出, θ 直接影响类型候选集的覆盖概率和 N_{Path} 。较大的 θ 会导致 N_{Path} 指数级增长, 增加计算复杂度; 较小的 θ 则会因过度剪枝而遗漏低概率协议组合, 增加漏检风险。实

2. 基于 $\xi_{\min} \leq \xi_i \leq \xi_{\max}$ 与式 (10), 以 P_{Path_1} 为基础生成 $(K-1)$ 条候选路径, 最终形成包含 P_{Path_1} 在内的 C_{Path}

3. for each P_{Path_i} in C_{Path}

4. for each P_{Pkt_i} in P_{Path_i}

5. 依据式 (9) 计算 P_{Pkt_i} 的概率并排序 $\{p_1, p_2, p_3\}$

6. if $p_1/p_2 < \theta$

7. $\tilde{T}_j = \{T_1, T_2\}$

8. else

9. $\tilde{T}_j = \{T_1\}$

10. end

11. end

12. end

13. 基于 C_{Path} 和 $\tilde{T}$ 生成 C_{Joint}

14. for each P_{JPath_i} in C_{Joint}

15. 依据式 (13) 计算 M^i

16. 依据式 (14) 输出 $P_{\text{JPath}_{\text{opt}}}$

17. end

3 仿真结果与性能分析

为验证所提方法在复杂网络环境下的有效性, 构建两类测试场景: 1) HDLC/MAC/IPv4 卫星通信模型^[13] (假设 IP 不分片、首部固定 20 B), 如图 3 所示; 2) 真实网络环境下的 WIDE MAWI 数据集^[17], 其记录了跨太平洋骨干链路中来自日本多所大学与国际机构间通信的长期流量, 其中 TCP 和 UDP 占比超 90%, ICMP 也占一定比例, 具有较强代表性。

验结果如图 4 所示。当 $\theta > 1$ 时, 算法容忍误码率门限可由 0.08 提升至 0.1, 但性能随 θ 的增大趋于饱和。假定 p_3 取最大值, 临界情况时, $p_1/p_2 = \theta$ 、 $p_2 = p_3$ 、 $p_1 = \frac{\theta}{\theta+2} p_1 + p_2 = \frac{\theta+1}{\theta+2}$, 这表示 θ 所对应的最小概率覆盖率。当前仿真设置对应的覆盖率依次为 $\frac{2}{3}$ 、 $\frac{4}{5}$ 、 $\frac{6}{7}$ 、 $\frac{8}{9}$ 、 $\frac{10}{11}$ 。可见, 当 $(p_1 + p_2)$ 占主导时, 高

概率类型集合已充分覆盖,漏检风险极低。

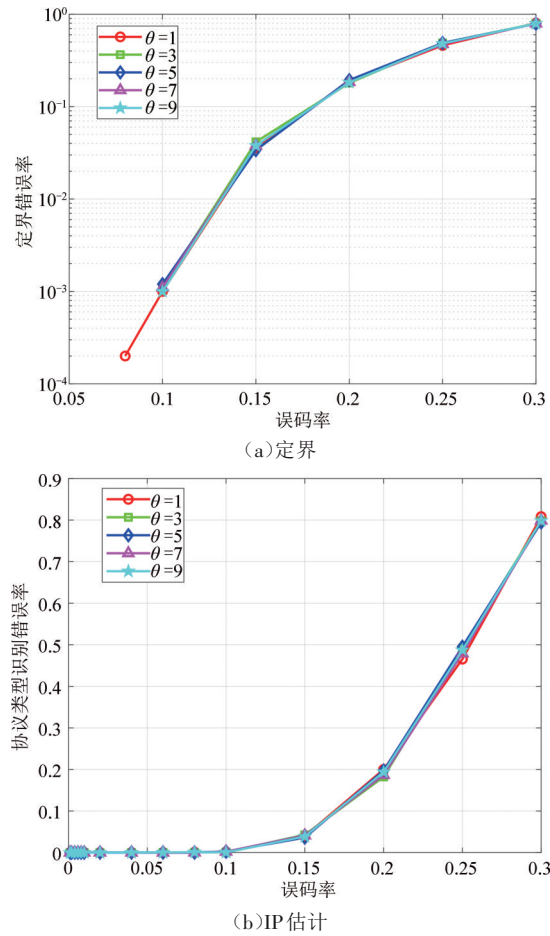


图4 不同 θ 下的联合估计结果

此外,各协议类型字段结构差异较大,汉明距离通常显著。例如, $p_e = 0.1$ 时,若 y 与两类特征之间的汉明距离差仅为2,归一化似然比分别为 $p(y|A) \approx 0.98$ 、 $P(y|B) \approx 0.02$,此时优势类型已很明显,即便 $\theta = 9$,概率覆盖率也仅约91%,继续增大 θ 对性能的提升有限。

综上, $\theta \in \{3,5\}$ 为理想参数范围,能有效平衡 N_{Path} 与准确率。

3.1.2 不同 K 的影响

为选取合适的 K ,在 $K = \{3,5,7,9\}$ 的条件下进行了多组仿真实验。与 θ 类似, K 显著影响复杂度。较大的 K 可提高保留最优路径的概率,但会增加计算开销;较小的 K 则易漏检真实路径,影响性能。实验结果如图5所示。

在当前设定条件下,定界与协议识别的性能并未随 K 的增加而显著提升。这是因为真假粗节点的路径度量差异显著。在定界阶段, $p_e = 0.1$ 时,有效码长 $L_e = 72$ bit,假节点单步度量为 $\ln(0.5) \times 72 = -49.91$ 。 N 长的码字错误位数服从参数为 N

和 p_e 的二项分布^[14],平均错误位数为 $N \times p_e$ 。因此,真节点的度量可计算为 $d \times \ln(p_e) + (L_e - d) \times \ln(1 - p_e) = -23.41$, $d \approx 7.2$ bit。二者差距明显,最优路径易于保存。真假节点度量比较如图6所示。可以看出,随着误码率的升高,两类节点的度量趋于一致,由 P_{Path_1} 构建 C_{Path} 时,最优路径与次优路径之间的区分度下降,此时应适当增加 K 防止漏检。

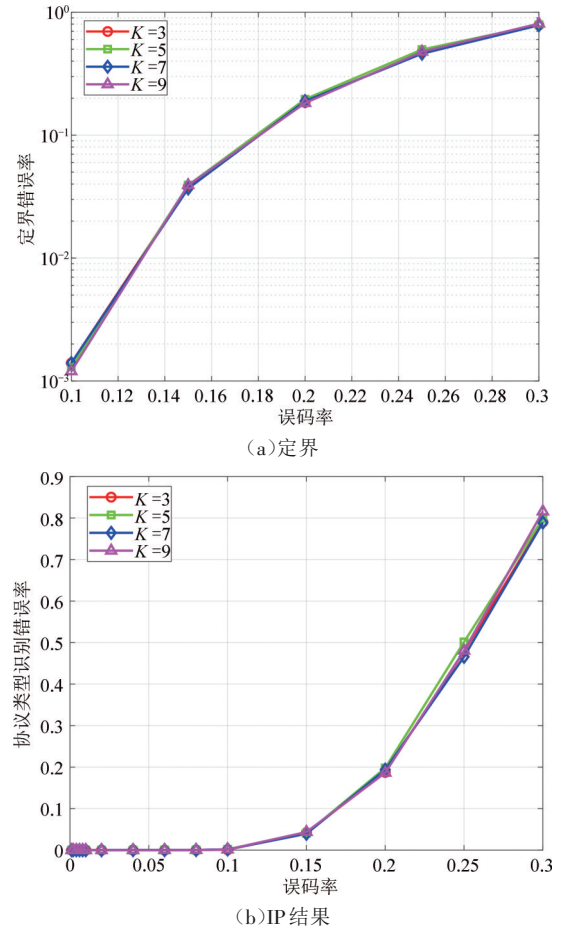


图5 不同 K 值下的联合估计结果

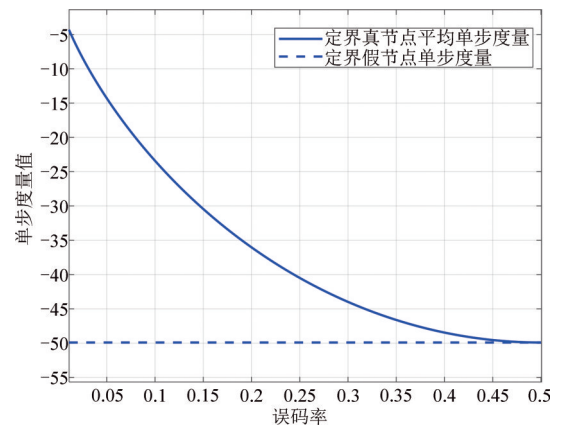


图6 真假节点度量比较

在进一步考虑联合估计时,有效码长扩展至118 bit,假节点单步度量为-81.79。引入 $[r,u]$ 后,

TCP、UDP和其他协议的真节点度量分别为-44.24、-47.84和-52.34。多数情况下,由于 P_{Path_i} 通常比次优路径多1个真节点,其累积度量显著更高,导致后续扩展的最优联合路径的累积度量值相对较低,因而算法难以完全纠正“多定1帧”类型错误。

综上,尽管增大 K 可覆盖更多候选路径,但因真假节点度量差异显著,对算法性能提升有限。

3.2 HDLC/MAC/IPv4下的性能比较

为评估本文提出的 $\text{top-}k$ 联合估计方法的性能,实验将其与传统的“先定界后分类”的级联方法^[13]对比:前者联合优化边界与协议识别,充分利用冗余;后者仅依赖Viterbi定界结果独立识别类型。为衡量前序错误对后续的影响,定义仅当起止边界均准确时定界正确;协议识别正确需定界正确且类型判定无误。即识别错误包括定界错误导致的误分和定界正确下的类型误判。

3.2.1 不同误码率下的仿真结果

实验依据图3的HDLC/MAC/IPv4协议栈生成10万个分组,其中TCP、UDP和其他占比分别为65%、30%和5%,用来模拟实际协议分布。每10个分组组成1帧,作为二进制对称信道(Binary Symmetric Channel, BSC)信道的输入。为平衡性能开销,设 $\theta = 5$ 、 $K = 5$ 。在Viterbi定界阶段采用的固定字段长度为61 bit^[13],包括HDLC标志位、MAC协议号、IPv4版本号、首部长度、区分服务、标志位、片偏移以及IP总长度的高5位。不同误码率下的定界结果如图7所示。

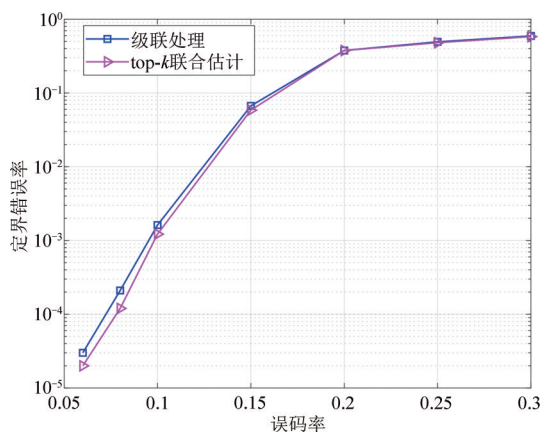


图7 不同误码率下的定界性能

$p_e = 0.1$ 时, $\text{top-}k$ 联合估计将定界错误率由0.17%降至0.12%,修正了约30%的错误。通过利用3层及4层中的自然冗余,有效定界比特从72 bit增至118 bit,有效增强了路径区分能力,部分纠正了“多定1帧”的错误。但由于原始Viterbi定界错误率已较低,因此联合估计算法所带来的性能增

益相对有限。随着误码率升高, $\text{top-}k$ 联合估计与Viterbi的性能差距逐渐缩小并趋于一致。原因是“多定1帧”的错误占比下降,而“一分为三”或边界偏移等其他错误占比显著上升,由于该算法仅针对前者优化,故在高误码下优势有限。

为评估级联处理中的误差传播,协议识别除对比单级级联与联合估计外,还引入了定界完全正确下的理想识别结果作为上界,结果如图8所示。

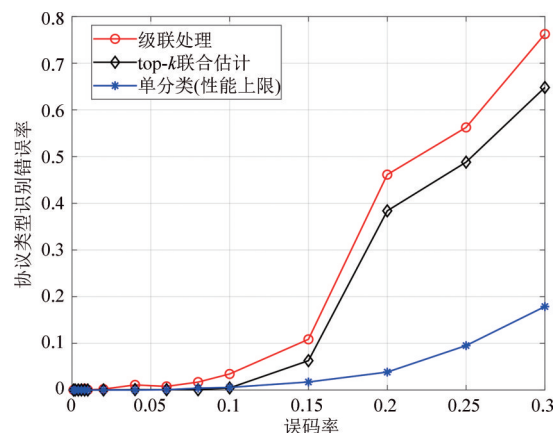


图8 不同误码率下的协议识别性能

$p_e = 0.1$ 时, $\text{top-}k$ 联合估计将IP分类错误率由3.44%降至0.56%,接近理想性能0.44%。通过联合利用冗余, TCP、UDP和其他分组的可用冗余分别从46、40、24 bit提升至102、96、80 bit,显著增强了类型区分能力。此外,结果也表明,前期定界准确性对分类性能具有决定性影响,边界错误将引发后续的误判,联合求解有效抑制了误差扩散。

3.2.2 复杂度分析

为评估 $\text{top-}k$ 联合估计与传统级联方法的计算开销差异,从时间复杂度与空间复杂度两方面进行对比:时间复杂度由冗余字段似然计算次数决定;空间复杂度反映在累积度和协议类型概率(或对数似然)的存储开销。

为简化分析,沿用文献^[13]的假设:粗节点 $\{g_1, g_2, \dots, g_m, g_{m+1}\}$ 在长为 ℓ 的 y 上均匀分布,分组长度 ξ 满足 $\xi_{\min} \leq \xi \leq \xi_{\max}$ 。则平均每下游节点对应 $m_{\text{up}} = (\xi_{\max} - \xi_{\min})m/\ell$ 个上游节点,形成 mm_{up} 个节点对。假设各分组 $\varepsilon_i = 2$,并结合 K 条 P_{Path_i} ,可构造 $N_{\text{Path}} = K \times 2^{\eta_{\max}}$ 条 P_{Path_i} ,其中 $\eta_{\max} = \lfloor \ell/\xi_{\min} \rfloor$ 为最大可能分组数。复杂度对比如表1所示。该分析基于最坏情况假设: $\xi_i = \xi_{\min}$ 、 $\varepsilon_i = 2$,导致 N_{Path} 达到理论上限。但在实际通信中, ξ_i 的不确定性使得帧内

分组数通常小于 $\eta_{\max}$, 且由于协议结构差异显著, 多数分组的类型分布高度集中, 多类型共存概率较低。因此, 实际上 N_{Path} 将远低于理论估计, top- k 方法的真实开销也将低于上述结果。

表1 复杂度对比

算法	时间复杂度	空间复杂度
级联	$2mm_{\text{up}} + 3\eta_{\max}$	$O(m + \eta_{\max})$
top- k	$2mm_{\text{up}} + 3\eta_{\max} + 2mm_{\text{up}}N_{\text{Path}}$	$O(m + \eta_{\max} + N_{\text{Path}})$

3.2 公开数据集下的仿真结果

为验证算法在真实环境中的适用性, 本文采用 WIDE MAWI 公开数据集进行实验。筛选满足 IP 总长度 ≤ 1500 B 的分组以符合固定字段假设。实验结果如图 9 所示。

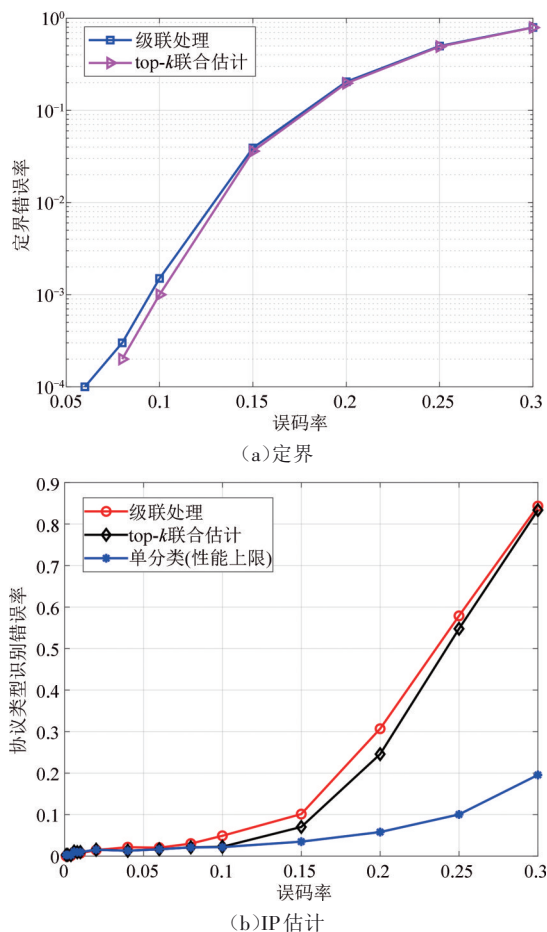


图9 公开数据集下联合估计结果

结果表明, top- k 方法将定界无误的误码率门限从 0.06 提升至 0.08。 $p_e = 0.1$ 时, Viterbi 定界错误率为 0.15%, 后续分类错误率为 4.9%; 联合估计将二者分别降至 0.1% 和 2.26%, 接近理想性能 2.17%。

需注意, 实验对真实数据进行筛选。若直接

处理原始数据, 部分 k 和 r 将不满足假设 (如 TCP 紧急指针非零、分片导致标志位不可用、巨型帧使 IP 长度高位失效), 用于定界字段长度将小于 61 bit, 性能可能略低于仿真, 但总体趋势一致, 仍验证了算法在真实环境中的有效性。

4 结束语

针对高误码下级联处理非最优且易错误扩散的问题, 提出一种联合最优的跨层容错估计方法, 通过挖掘协议栈冗余, 设计 top- k 路径搜索机制, 在保留 K 条最优路径及对应协议组合的基础上, 显著提升了定界与识别性能。但当前方法尚未逼近 MAP 解, 未来将利用数据包长度分布等冗余构建更优模型, 并深化对 θ 、 K 的理论分析, 同时探索其他迭代算法以继续降低复杂度。

参考文献:

- [1] 许彤通. 复杂战场环境中可靠通信接收关键技术研究[D]. 西安: 西安电子科技大学, 2022: 1-130.
- [2] PRIYA P, REDDY C S, SEN D. Channel estimator and nonlinear detector for mmWave beamformed OTFS systems in high mobility scenarios[J]. IEEE Transactions on Vehicular Technology, 2023, 72(9): 11698-11713.
- [3] JI X Y, WANG C X, CHANG H T. Characteristic analysis and modeling of underground space wireless communication channels[C]//Proceedings of 2022 IEEE 95th Vehicular Technology Conference. Piscataway, USA: IEEE, 2022. DOI: 10.1109/VTC2022-Spring54318.2022.9860697.
- [4] MACCARTNEY G R, RAPPAPORT T S. Rural macrocell path loss models for millimeter wave wireless communications[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(7): 1663-1677.
- [5] 包涵, 涂国防, 张灿, 等. 基于可变长纠错码和掺杂调制的联合信源信道编码调制方法[J]. 电子与信息学报, 2023, 45(6): 2045-2053.
- [6] TRANTER W H, SHANMUGAN K S, RAPPAPORT T S, 等. 通信系统仿真原理与无线应用[M]. 肖明波, 许芳, 杨光松, 译. 北京: 机械工业出版社, 2005: 57-87.
- [7] KADIYAM S, JHANSI RANI A. Design and analysis of a high gain millimeter-wave antenna array for dual purpose applications[J]. Wireless Personal Communications, 2023, 130(1): 593-607.
- [8] LI Y B, YU H Y, ZHANG X, et al. Generalized channel coding and decoding with natural redundancy in protocols[J]. IEEE Transactions on Communications, 2024, 72(2): 648-664.

(下转第 651 页)