基于 CPU-FPGA 协同架构的 VoIP 数据加密系统设计与实现

李　斌[1]，杨　欢[1]，李德阳[2]，杨志明[1]，姬胜凯[1]

（1. 中国电子信息产业集团有限公司第六研究所，北京　100083；2. 96901 部队，北京　100094）

摘　要：针对通信网络飞速发展背景下 VoIP 面临日益严峻的数据安全风险，提出并实现了一种基于 CPU + FPGA 软硬件协同处理架构的 VoIP 数据加密方案，高效集成 AES 算法，对实时采集的 VoIP 话音流进行加密处理。搭建专用话音测试环境，采用主观、客观相结合方法对所设计的加密模块在正常工作状态下的性能及通话质量进行全面评估。测试结果表明，该加密方案在保障安全性的同时有效维持了通话质量。

关键词：FPGA；软硬件协同；AES；VoIP；加密

中图分类号：TN918. 4；TP309　　**文献标识码：**A　　**DOI：**10. 19358/j. issn. 2097 - 1788. 2025. 10. 007

引用格式：李斌，杨欢，李德阳，等. 基于 CPU-FPGA 协同架构的 VoIP 数据加密系统设计与实现［J］. 网络安全与数据治理，2025，44(10)：40 - 45.

Design and implementation of a VoIP data encryption system based on a CPU-FPGA collaborative architecture

Li Bin[1]，Yang Huan[1]，Li Deyang[2]，Yang Zhiming[1]，Ji Shengkai[1]

（1. The 6th Research Institute of China Electronics Corporation，Beijing 100083，China；2. Unit 96901，Beijing 100094，China）

Abstract：In the context of the rapid development of communication networks and the increasingly severe data security risks faced by VoIP，this paper proposes and implements a VoIP data encryption scheme based on a CPU + FPGA software-hardware collaborative processing architecture. The scheme efficiently integrates the AES algorithm to encrypt real-timely captured VoIP voice streams. A dedicated voice test environment was established，and a combined subjective and objective approach was used to comprehensively evaluate the performance and call quality of the designed encryption module under normal operating conditions. The test results demonstrate that the proposed encryption scheme effectively maintains call quality while ensuring security.

Key words：FPGA；hardware/software co-design；AES；VoIP；encryption

0　引言

现代通信网络的飞速发展深刻改变了信息交互模式。传统电话网络基于严格的时分复用（Time Division Multiplexing，TDM）机制构建，其核心问题在于带宽利用率低下[1]。TDM 要求为每个通话固定分配一条 64 kb/s 的信道，导致网络必须承载大量静默期服务（例如交谈中的思考停顿、轮流讲话间隙），造成了显著的带宽资源浪费。近年来，IP 语音（Voice over IP，VoIP）技术凭借其高效性在通信领域获得了广泛应用，其利用分组交换和模拟 - 数字转换技术，能够在仅需 4. 8 kb/s ~ 8 kb/s 的较低带宽下，建立高质量的话音通道，相比于传统 TDM 电话节省了大量带宽资源，为用户带来了极大的便利，并推动了通信服务的普及和升级[2]。然而，VoIP 基于开放的 IP 网络传输语音数据包，其固有的开放性也引入了严峻的安全挑战。随着网络电话的普及，针对 VoIP 通信的威胁日益凸显，包括但不限于：通话内容窃听、语音数据篡改、身份伪装欺骗以及服务拒绝攻击[3]。这些威胁不仅侵犯用户隐私，也可能导致重要信息泄露或通信中断，造成严重后果。在信息全球化和数据价值日益提升的今天，保障 VoIP 语音数据的机密性、完整性和实时可用性已成为通信安全领域的关键需求。特别是，实时语音通信对端到端延迟和语音质量有着严苛的要求，因此，开发能够满足高安全性与低处理延迟双重目标的 VoIP 数据保护方案，具有重要的理论意义和实际应用价值。

当前，保障 VoIP 数据安全的主要技术路线包括：（1）软件加密方案：在通用处理器 CPU 上运行加密算法。其优势在于灵活性强、易于部署和更新。然而，纯软件方案在处理高吞吐量、低延迟的实时语音流时面临

巨大挑战。复杂的加解密运算会显著增加 CPU 负载，导致处理时延增大、抖动加剧，甚至可能因资源不足而丢包，最终严重劣化通话体验质量（QoS/QoE）[4]。（2）专用硬件加密方案：采用专用集成电路（ASIC）实现加密功能。ASIC 通常能提供极高的吞吐量和极低的固定延迟。但其主要缺陷在于缺乏灵活性和可重构性。一旦芯片设计完成，算法通常难以更新升级，且开发周期长、成本高，难以满足 VoIP 快速演进的安全需求和多样化的应用场景[5]。（3）基于 FPGA 的硬件加速方案：利用 FPGA 的并行处理能力实现加密加速。FPGA 方案在性能和灵活性之间取得了一定平衡[6]，然而，现在许多 FPGA 加速方案未能充分考虑在完整 VoIP 终端或网关系统中，如何最优协调安全性、实时性（低延迟、低抖动）与资源（功耗、逻辑单元）消耗之间的关系。此外，如何高效管理 CPU 与 FPGA 之间的数据交互和控制，也是影响整体系统性能和复杂度的关键因素。综上所述，现有 VoIP 数据安全方案在应对实时性、灵活性、系统集成度以及性能 - 资源平衡等方面仍存在明显局限。亟需一种能够有效协同高性能硬件加速与灵活软件控制的新型架构，在确保强安全性的同时，严格满足 VoIP 对低延迟和高语音质量的苛刻要求。

针对上述挑战与需求，本文创新性地提出并实现了一种基于 CPU + FPGA 软硬件协同处理架构的 VoIP 数据加密方案。本方案的核心思想在于根据任务特性进行合理的软硬件功能划分：CPU 及其存储模块组成了主控调度单元，负责核心控制与调度功能，包括各类通信接口数据的解析与协议处理、系统资源的动态配置与管理，同时可根据命令类型，调度和控制算法运算单元的工作流程；FPGA 作为算法运算单位的核心部件，利用其强大的并行处理能力和可重构特性，高效实现关键数据的加解密运算，其通过硬件逻辑直接处理实时捕获的 VoIP 语音流，显著提升处理速度并降低延迟。通过这种协同架构，本方案旨在深度融合软件的灵活控制优势与硬件的高效并行处理能力，突破纯软件方案在实时性上的瓶颈和纯硬件方案在灵活性上的不足，最终实现在强安全保障下，对 VoIP 语音质量的有效维持。为验证方案的有效性，本文搭建了专用的语音测试环境，采用主观与客观相结合的评价方法，对加密模块的性能和通话质量影响进行了全面评估。

1 VoIP 概述

VoIP 又称网络电话，是伴随 IP 技术和电信业务的发展而产生的一种使用 IP 分组数据来承载语音业务的新型应用。和传统的基于公共线路交换网络（PSTN）的语音业务相比，VoIP 具有运营成本低、功能扩展性强的优势[7]。

1.1 VoIP 基本原理

VoIP 通过 IP 网络实现语音信息传递，其基本原理是：话音信号经发送端发送后进行模拟信号到数字信号的转换，再通过压缩编码技术对转换完的数字信号进行压缩处理，最后按照网络协议对压缩编码的话音数据进行封包，封包后的网络数据将通过 Internet 发送至接收方；接收方依次按照 IP 解包、解压缩编码、数字信号转换模拟信号的顺序将语音信号还原，从而实现双方正常的话音通话[8]。VoIP 基本原理如图 1 所示。

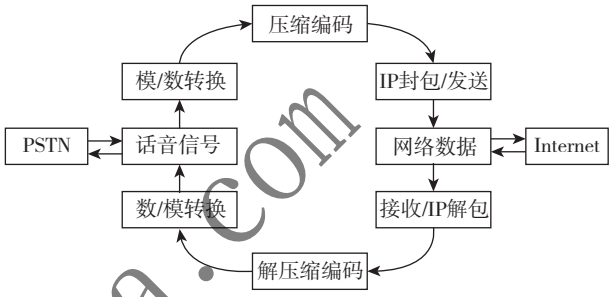

图 1 VoIP 基本原理图

1.2 VoIP 常用协议

随着 VoIP 的发展，一些相关的 VoIP 协议也应运而生，目前被广泛接受的 VoIP 控制信令体系包括 ITU-T 的 H.323 系列和 IETF 的会话初始化协议 SIP[9]，可保证电话呼叫的顺利实现和话音质量，其中 SIP 协议是 VoIP 使用比较统一的协议，它支持 TCP 和 UDP 两种传输协议，扩展性较强。关于媒体流传输，IETF 定义的实时传输协议（Real Time Protocol，RTP）作为传输载体，使用 UDP 传输协议，同时实时传输控制协议（Realtime Transmission Control Protocol，RTCP）对话音传输过程中的数据状态进行实时监控，二者配合使用，保证实时话音数据的正确性。VoIP 通信协议网络如图 2 所示。

<table>
<tr><td colspan="5">工具、媒体等</td></tr>
<tr><td colspan="2">SDP</td><td colspan="3">媒体流</td></tr>
<tr><td>H.323</td><td>SIP</td><td>RTP</td><td>RTCP</td><td>RTSP</td></tr>
<tr><td colspan="2">TCP</td><td colspan="3">UDP</td></tr>
<tr><td colspan="5">IP 网络</td></tr>
<tr><td colspan="5">底层传输</td></tr>
</table>

图 2 VoIP 通信协议网络

1.3 SIP 工作流程

SIP 消息可大致分为 SIP 请求和响应，其中，SIP 请求主要包括 INVITE、ACK、BYE、CANCEL 等请求类型；SIP 响应主要包括 Trying_100、Ringing_180、200_OK、

400_Bad Request、404_Not Found、408_Request Timeout 等响应状态[10]，相关状态如表 1 所示。

表 1　SIP 呼叫响应相关状态

消息状态	含义
SIP 请求	
INVITE	发起呼叫，邀请对端加入会话
ACK	确认收到 INVITE 呼叫请求
BYE	终止呼叫
CANCEL	取消还未 OK 的呼叫、请求
SIP 响应	
Trying_100	尝试呼叫
Ringing_180	振铃确认
200_OK	成功呼叫
400_Bad Request	普通错误消息，客户端不能识别消息
404_Not Found	请求的用户未找到
408_Request Timeout	请求超时

SIP 终端有多种方式建立呼叫，本文以双方之间已知对端地址，直接发起呼叫流程[11]为例，具体如图 3 所示。

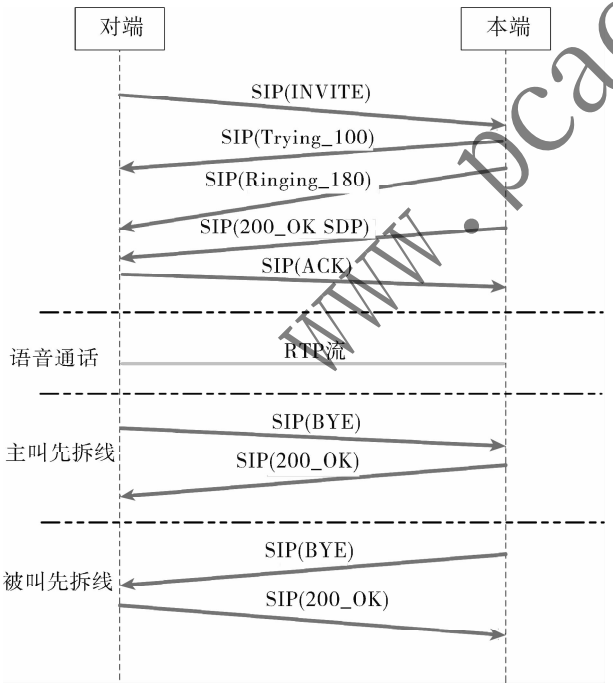

图 3　SIP 呼叫流程

（1）对端发起 INVITE 请求，直接将本端 VoIP 的地址信息添加至 INVITE 请求帧的标题头，邀请本端加入会话；

（2）本端接受 INVITE 并回发 Trying_100 给对端；本端开始振铃，提醒有来电，同时回发 Ringing_180；

（3）本端应答呼叫，将对端地址信息添加至 200_OK 应答帧的标题头；

（4）对端根据 200_OK 响应中的标题头，发送一个 ACK 给本端。至此，SIP 对话被建立起来。

（5）话音在 RTP 流上承载，双方进行交流；

（6）断开呼叫时，一端发送一个 BYE 的请求给另一端，通话双方的 RTP 流被停止，另一端通过发送 200_OK 确认 BYE 的事务。至此，双方呼叫被正式终结。

1.4　VoIP 话音面临的安全问题

由于 IP 网络的开放性，使得 VoIP 在 IP 网络上运行时易遭受攻击，这样就会导致在进行一些重要通话过程时容易被人监听，造成信息泄露等[12]。

本文基于 VoIP 的 G. 729 协议编码方式，采用 AES 算法对 RTP 流进行加密，保证其完整性和机密性，防止窃听和破译。

2　系统硬件架构设计

VoIP 加密模块采用一体化设计，硬件架构主要由 VoIP 话音单元、主要控制单元、算法处理单元、电源时钟单元和接口单元组成，系统硬件架构如图 4 所示。

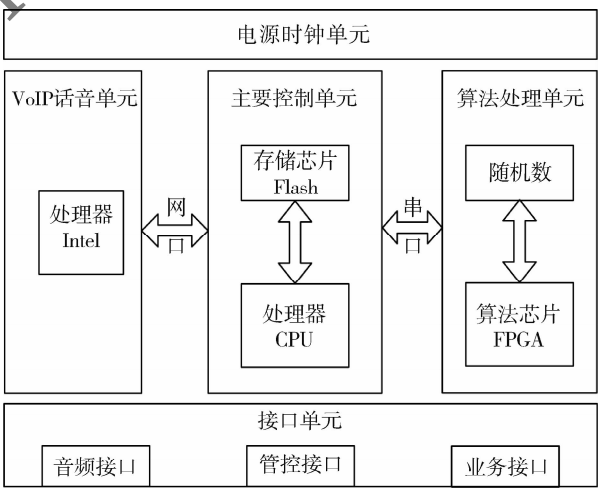

图 4　系统硬件架构

2.1　VoIP 话音单元

VoIP 话音单元主要由 Intel 研制的 PXB4389 组成，其主要实现 SIP 相关协议的逻辑处理、通信报文的逻辑处理、RTP 语音包处理、语音编解码、回声处理以及呼叫流程处理等。

VoIP 话音单元对外提供音频接口，具备模拟话音接入，单路模拟接口接入电话机功能。

2.2　主要控制单元

主要控制单元由龙芯公司研制的 GSC3280 低功耗

CPU 和存储芯片组成。CPU 实现任务调度、多通道服务、协议解析、管理维护、资源配置管理、数据流向控制等功能；存储模块选用兆易创新研制的 GD25Q128C，为系统程序存储和运行提供足够空间。

2.3 算法处理单元

算法处理单元块基于 Xilinx 研制的高性能芯片 XC4VLX25 提供算法处理，实现数据安全保密通信。随机数芯片选用 WNG5，保证算法运算所需的随机数。

2.4 电源时钟单元

电源时钟单元主要实现电源的稳定和转换以及全局时钟输出功能。电源单元负责进行电压转换，提供工作电压、CPU 所需的 +1.8 V 和 +1.2 V 核心电压以及 FPGA 所需的 +1.5 V 核心电压；时钟单元采用 50 MHz 的晶振通过时钟驱动电路实现。

2.5 接口单元

接口单元主要为音频接口、业务接口、控制接口提供交互功能，其中音频接口具备模拟话音接入，业务接口支持 100 Mb/s 以太网络，控制接口采用标准 232 串口。

3 系统软件体系设计

VoIP 加密模块软件采用分层设计，主要包括话音交互层、话音业务处理、SIP 协议栈、RTP/RTCP 协议栈、CODEC、IP 网络层和系统接入层，具体如图 5 所示。

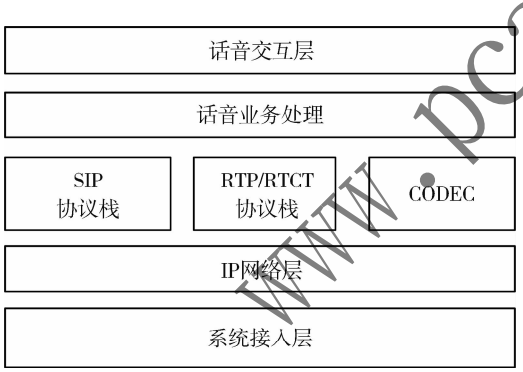

图 5　系统软件架构

话音交互层负责与外部数据通信及话音业务的衔接；话音业务层负责处理呼叫流程；SIP 协议栈负责处理 SIP 相关的协议；RTP/RTCP 协议栈负责处理 RTP 实时语音流；CODEC 负责数据的安全保密；IP 网络层提供 TCP/IP 的接入；系统接入层提供嵌入式系统资源。

（1）话音在 VoIP 加密模块加密传输流程

如图 6 所示，话音数据通过音频接口传到 VoIP 话音单元，在话音单元依次通过音频编解码，对话音进行模拟/数字转换处理，然后根据相应语音编解码技术对其进行压缩编码处理，经过 CPU 添加各种头部后转换为 IP 数据包经网口传递给主要控制单元，主要控制单元 CPU 通过协议解析后，将待加密话音数据发送给算法处理单元 FPGA 进行话音加密，加密完成后的话音数据返给主要控制单元 CPU 进行 IP 封包处理，随后通过业务网口传递给对端。

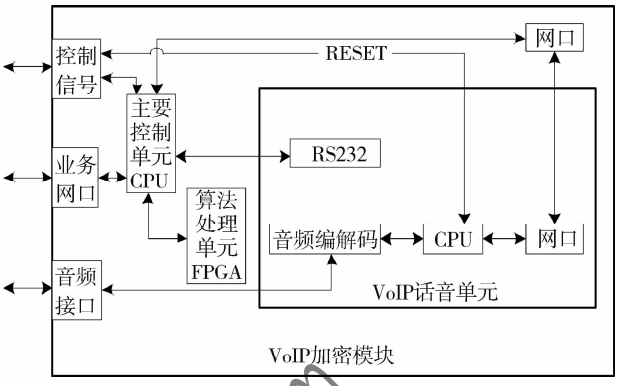

图 6　话音数据在 VoIP 加密模块传输流程图

（2）话音在 VoIP 加密模块解密传输流程

如图 6 所示，加密话音数据通过业务网口发送到主要控制单元 CPU，CPU 通过协议解析将已加密数据解析出来发送给算法处理单元 FPGA 进行话音数据解密，随后将解密后的数据发送给主要控制单元 CPU，其通过网口发送给 VoIP 话音单元 CPU 进行语音包处理，通过音频编解码进行数字/模拟转换成话音信号，通过音频接口输出给本端。

4 系统设计

在 VoIP 语音编码中，G.711、G.721、G.723、G.729 应用较为广泛，结合带宽占用、压缩声音质量、帧丢失、分组丢失处理机制、噪声情况、话音质量等特性，最终选用 G.729 语音编码标准。

加解密算法选用使用广泛且安全性更高的 AES 对称算法，其加密机理：数据为定长 128 bit，且密钥长度支持 128 bit、192 bit、256 bit 三种。本方案选用安全性更高的 256 bit 密钥长度[12]。

VoIP 加密模块的具体工作流程如图 7 所示。首先对 IP 数据报文进行协议解析，根据数据报文头判断此报文数据需要透传、加密或者解密。

若判断为透传需求，则不需要调用算法运算单元，直接进行数据转发即可。

若判断为加密需求，首先从 VoIP 加密模块随机数芯片 WNG5 读取 128 bit 的数据作为 AES 算法运算的 IV 数据，同时根据报文解析出的组号从主要控制单元 Flash 芯片预存的多组数据中读取相应组的 256 bit 作为 AES 算法运算的 WK 数据，IV 数据和 WK 数据在算法运算单元运算产生密码流，密码流与待加密数据进行 XOR 生成密文。随后将 IV 数据以伪装方式填入数据报文特定位置，

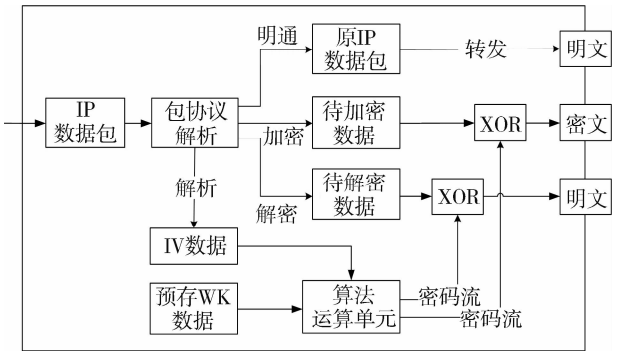

图 7　VoIP 加密模块工作流程图

通过对数据重新封装后，进行传递。

若判断为解密需求，首先对接收的数据报文进行协议解析，解析出 IV 数据和 WK 组号，同样根据 WK 组号从主要控制单元 Flash 芯片预存的多组数据中读取相应组的 256 bit 数据，同样 IV 数据和 WK 数据在算法运算单元预算产生密码流，密码流与待解密数据进行 XOR 生成明文。

5　实现与验证

本文基于 G. 729 编码格式，设计了 VoIP 加密模块硬件平台，采用 AES-256 算法，实现了 VoIP 话音加密功能。同时，搭建了测试环境，测试经过话音加密后的通话质量是否满足要求。

图 8 是用 Wireshark 软件抓取的采用 G. 729 编码的 SIP 呼叫建链流程图。

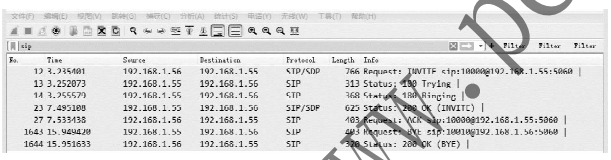

图 8　SIP 呼叫流程图

呼叫建链流程：本端（IP：192. 168. 1. 56）向对端（IP：192. 168. 1. 55）发起 INVITE 请求，对端接收 INVITE 请求并回发 Trying_100，对端开始振铃，提醒有来电，同时回发 Ringing_180；对端应答呼叫，并回发应答帧 200_OK；本端发送 ACK 到对端；至此 SIP 对话被建立起来。

SIP 呼叫成功建链后，两个 VoIP 电话之间进行 RTP 流的传输。图 9 是用 Wireshark 软件抓取的采用 G. 729 编码的含有双边未加密的 RTP 流（IP 五元组相反）的 pcap 文件。

图 10 为随机抓取的一包 RTP 流，其中 RTP 包中 Synchronization Source identifier 字段为 0x00000000，Payload 前 12 个 Byte 为 0x000000000000000000000000，组合起来为 128 bit 的 IV，用于 AES 算法运算。

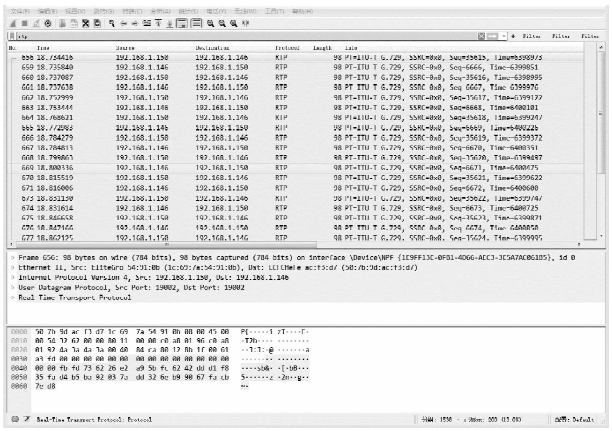

图 9　Wireshark 抓取的未加密的 RTP 流

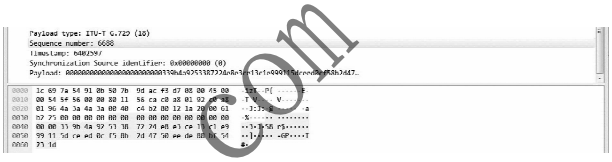

图 10　RTP 流同步码、负载部分解析

图 11 是用 Wireshark 软件抓取的采用 G. 729 编码的含有双边加密的 RTP 流（IP 五元组相反）的 pcap 文件。

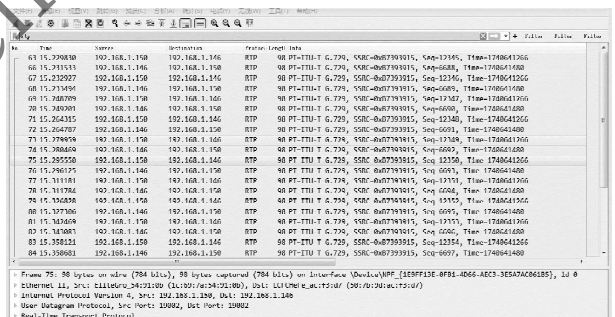

图 11　Wireshark 抓取的加密的 RTP 流

其中根据 RTP 包中的 Synchronization Source identifier 字段和 Payload 前 12 个 Byte，知本包数据流选取 IV 为 0xb7393915d75d607963017459c01833da，根据 Payload 知 VoIP 加密话音后，密文字段为 0xf751a985c6d44f28a618b046d25e32370586b6ec4cc1ae4d35546e6840cb188c。具体如图 12 所示。

图 13 呈现了通过 Wireshark 软件 RTP 流分析功能解码还原的双方通话音频波形图（源自加密 RTP 流）。为了评估话音的整体感知质量，招募了 20 名听力正常的被试者，在安静的实验室环境中使用专业耳机进行听音测试。测试材料包括 20 条由本方案处理的语音样本。被试根据

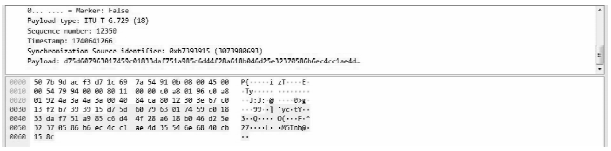

图 12　加密 RTP 流解析

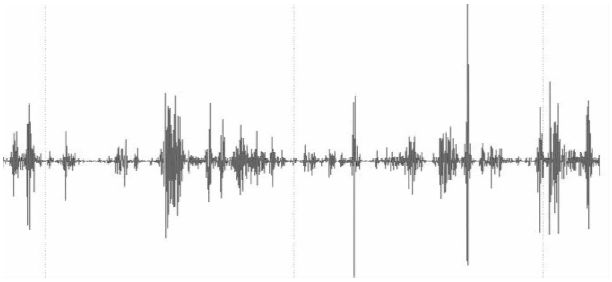

图 13　加密 RTP 流的输出音频

标准的 5 级 MOS 量表（1 = 劣，5 = 优）对每条样本的整体质量进行评分。最终主观平均意见得分（MOS）达到 3.7 分，试听评估确认，还原后的语音质量清晰[13]，通话内容可懂度高，满足预期要求。通过 Wireshark 对视频会议系统主音频流（SSRC：0x10819bc6）进行抓包分析，观测到丢包率（Lost）为 0%，符合 ITU-T G.113 对 VoIP 业务的可接受阈值（<2%）；未检测到序列号错误（Seq Errs = 0）；平均抖动（Mean Jitter）为 3.17 ms，满足 WebRTC 建议的实时语音抖动上限（30 ms）；检测到持续负向时钟漂移（Clock Drift = 0 ms），根据 RFC 3550 建议，该数值未超出 RTP/RTCP 的常规补偿能力（±5 ms/s）。本次分析仅反映客户端侧观测结果，未涵盖服务端内部处理延迟，后续研究可通过双向捕获完善评估。

6　结论

本文系统分析了 VoIP 技术面临的数据安全风险，提出了一种硬件加速的混合架构安全方案。该方案创新性地采用 CPU + FPGA 协同设计，通过 FPGA 并行化处理结合物理噪声源芯片，实现了符合 AES-256 标准的实时语音流加密，在保证算法强度的同时，解决了传统软件加密导致的处理延迟问题。

搭建了 VoIP 专用语音测试环境，采用主观与客观相结合的评价方法，对所设计的加密模块的性能和通话质量影响进行了全面评估。测试结果表明，本系统在稳定工作的情况下，可完成数据的安全传输，通话质量良好。综上所述，本方案提出的基于 CPU + FPGA 协同架构的 AES加密方案，能够在不显著影响 VoIP 通话质量的前提下，有效实现话音数据的实时安全加密传输，为解决 VoIP 应用中的数据安全问题提供了一种可行的硬件加速方案。

参考文献

[1] 赵怡松. 面向数据光网络的高精度时频同步技术研究 [D]. 北京：北京邮电大学，2025.

[2] 杨洁，邓浩江，李松斌. AbS-LPC 低速率压缩语音信息隐藏技术综述 [J]. 网络新媒体技术，2020，9（2）：9 - 15，59.

[3] 李晶，邓伟，杨勇，等. 基于多通道卷积神经网络的语音隐写分析方法 [J]. 电子器件，2022，45（5）：1105 - 1109.

[4] 耿磊. 无线网络中具有网络监管功能的端到端安全通信方案 [D]. 西安：西安电子科技大学，2017.

[5] 吴瑞东. 资源受限条件下卷积神经网络硬件加速优化方法研究 [D]. 哈尔滨：哈尔滨工业大学，2023.

[6] 杨宏浩，隋欣宇，王一然，等. 基于 FPGA 的高效近似乘法器设计 [J/OL]. 电子测量与仪器学报，1 - 9 [2025 - 07 - 22]. https://link.cnki.net/urlid/11.2488.tn.20250704.0926.002.

[7] 郑青青，田才艳. 基于 VOIP 的空管语音交换系统 [J]. 长江信息通信，2024，37（6）：149 - 151.

[8] 周詹平，高羽丰，马浩，等. 基于 VoIP 技术的通信系统在海外油区的构建 [J]. 录井工程，2022，33（3）：94 - 97.

[9] 余婷婷. 基于 nDPI 的 G.711 语音解码算法的实现与验证 [C]//第三十四届中国（天津）2020 IT、网络、信息技术、电子、仪器仪表创新学术会议论文集，2020：230 - 233.

[10] 胡坚，李福秧，丁茂，等. 基于 SIP 消息的三步四维法提升 4G/5G 语音用户感知 [J]. 电信工程技术与标准化，2025，38（5）：54 - 61.

[11] 敦科翔，戎烁，王云磊. 基于 VoIP 技术的电力调度交换网放号设备的设计与实现 [J]. 计算机与网络，2024，50（5）：386 - 390.

[12] 肖祯. 基于机器学习的电子信息系统数据加密研究 [J]. 电脑编程技巧与维护，2025（5）：88 - 90.

[13] 周柯汝，金伟. 基于 VITS 的高性能歌声转换模型 [J]. 现代信息科技，2025，9（12）：129 - 133，140.

（收稿日期：2025 - 07 - 22）

作者简介：

李斌（1991 - ），男，硕士，工程师，主要研究方向：信息安全、安全系统设计。

杨欢（1992 - ），男，本科，工程师，主要研究方向：信息安全、嵌入式软件设计。

李德阳（1978 - ），男，本科，助理研究员，主要研究方向：信息安全。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部