

技术白皮书

Open-AVD SDK 开放的会议+直播能力平台

V 3.0

杭州叁体网络科技有限公司

2017/04



目录

1. 背景	2
2. 产品线	2
2.1 产品线构成	2
2.2 OPEN-AVD SDK	3
2.2.1 概述	3
2.2.2 服务器端接口	3
2.2.3 客户端接口	4
2.2.4 SDK 分层	6
2.2.5 客户产品架构	6
2.2.5 设备与平台	7
2.2.6 服务器类型	8
2.3 幸会与爱秀	9
3. 技术特点与优势	9
3.1 全平台支持	10
3.2 领先音视频技术	10
3.3. 领先通讯技术	12
3.3.1 P2P 与 MCU	12
3.3.2 MCU 服务器	13
3.3.3 多核 CPU 利用	14
3.3.4 分布式服务器集群	15
3.3.5 混合云	16
3.3.6 旁路直播	17
3.4 其它	18
3.4.1 视频硬件编解码	18
3.4.2 视频混合编解码	19
3.4.3 视频动态码流	19
3.4.4 视频多流	19
3.4.5 Sip/H.323	19
3.4.6 RTMP	20
3.4.7 RTSP	20
3.4.8 第三方音视频流倒入与倒出	20
3.5 重要技术特性与指标	20
4 附录	22
4.1 HTML5 与 WEBRTC	22
4.2 术语表	24

1. 背景

杭州叁体网络科技有限公司，是国内最专业的视频通讯 SDK 技术公司，成立于 2014 年 12 月，总部位于杭州。成立以来，叁体已经发展了以浙江大华股份、创业软件为代表的众多安防、医疗及教育等领域众多客户，处于快速发展期。

公司创始团队核心成员，来自思科 Cisco、网迅 WebEx 等国内外知名网络视频会议及通讯系统公司，整个团队拥有累计数十人年相关领域产品开发与运营经验，具有深厚的国际化背景与广阔的技术与商业视野。

在多年相关经验基础上，结合业界前沿的 HTML5/WebRTC 等技术，叁体精心打造出了核心产品 Open-AVD SDK（花名阿梵达）实时多方会议+直播能力平台（中间件），提供文字聊天、音频、视频、直播、桌面分享、录制、白板及批注等功能，全面支持 iOS、Android、Windows、Mac、Linux、HTML5/WebRTC 浏览器、机顶盒及智能电视等各种平台与设备。

第三方团队可在 Open-AVD SDK 基础之上做开发，方便快捷地在自己的产品中引入实时音视频通讯能力，实现类似视频聊天及深度定制化的会议系统+直播等功能，可广泛应用于安防应急、远程教育、远程医疗等几十个领域。

2. 产品线

2.1 产品线构成

叁体的主要产品线如下图中绿框和蓝框所示。

产品线



作为一家专业视频 SDK 技术公司，叁体的核心产品是 Open-AVD SDK，英文全名是 Open Audio，

Video, Data, 意指开放的音频、视频与数据能力，中文花名是“阿梵达”，而“幸会”(nice2meet = Nice to Meet) 则是一款全功能的免费使用的会议产品，爱秀则是叁体用于演示直播能力的演示程序。

“幸会”和“爱秀”基于 Open-AVD SDK 开发，免费使用，中短期内，其主要用途是演示 Open-AVD SDK 的能力，其代码开源，方便第三方团队快速学习，在 Open-AVD SDK 基础上快速实现自己会议+直播功能。

2.2 Open-AVD SDK

2.2.1 概述

Open-AVD SDK 提供人与人实时沟通协作过程中需要用到所有基本能力，涵盖了网络会议系统、IM 即时通讯系统及直播系统三大类终端产品音视频通讯的主要功能。

Open-AVD SDK 由业界资深工程师精心打造，稳定可靠，第三方团队拿来就能用，不必自己去造“轮子”，从而降低了第三方团队的技术风险，减少了项目的开发投入，尤其是能大幅缩短第三方团队开发具有多方音视频+数据协作能力的 App/Web 应用的时间。

Open-AVD SDK 可用于几乎所有行业，很多业务场景中需要用到人与人实时沟通与协作的能力，而类似 QQ，微信或会议系统这种通用沟通工具又不能直接使用或不能满足功能，这种情况下，Open-AVD SDK 就是您最好的选择。市场调研表明，Open-AVD SDK 在医疗、教育、金融、能源、交通等各个领域，都有巨大的市场需求。

2.2.2 服务器端接口

- 主要功能

帐户管理，用户认证，单点登陆

协同（会议）安排，修改，启动

系统日志

统计、状态报告

- 接口形式

接口以 REST API 方式提供，相关数据为 JSON 格式

- 接口语言

服务器端接口支持如下语言：

- JavaScript

- PHP

- Python
- Java
- C++
- Ruby

2.2.3 客户端接口

● 主要功能

客户端接口主要支持以下功能：

- 文字聊天（单聊，多聊）
- 音频（两方，多方）
- 视频（多方）
- 桌面共享
- 白板
- 批注
- 文档共享
- 文件传输
- 录制
- 直播（单向音视频，2-10 秒延时）
- 各种控制功能，如禁音，踢人等等

● 接口形式

根据第三方团队不同的开发需要，Open-AVD SDK 提供三种不同级别和类型的接口，分别描述如下：

接口级别	接口类型	功能特点	适用场景
1 级	内嵌会议/IM	带 UI 的完整的会议或 IM，可直接嵌入第三方业务系统 通过配置接口，第三方应用可改变嵌入会议皮肤布局等等	适用于需要完整会议或 IM 功能并简单定制 UI 的一些业务应用 第三方开发工作量极小，最快的几十行代码几个小时就可完成
2 级	组件	带 UI 的完整功能组件，如文字聊天、视频、音频、白板、文档共享、桌面共享等等。 第三方应用系统根据需要嵌入	适用于需要对业务功能及用 UI/UE 进行深度定制的并需要和应用业务流程进行深度整合的场景。

		其用户界面中，由第三方系统决定何时以何种方式显示相关组件	比如只需要音频而不需要视频和数据的纯语音系统。 第三方开发代码量小
3 级	逻辑接口	没有 UI 的纯数据接口 除了音视频采集、编解码及传输，第三方系统完全控制音视频功能的其它所有方面，包括其 UI/UE 或一些特殊处理等等。	适用于需要完全控制用户 UI/UE 及功能的应用，同时，也为一些特殊功能开发（对将视频由彩色变黑白）提供了可能性 第三方开发代码量比前两种方式大，但提供了最大的灵活性

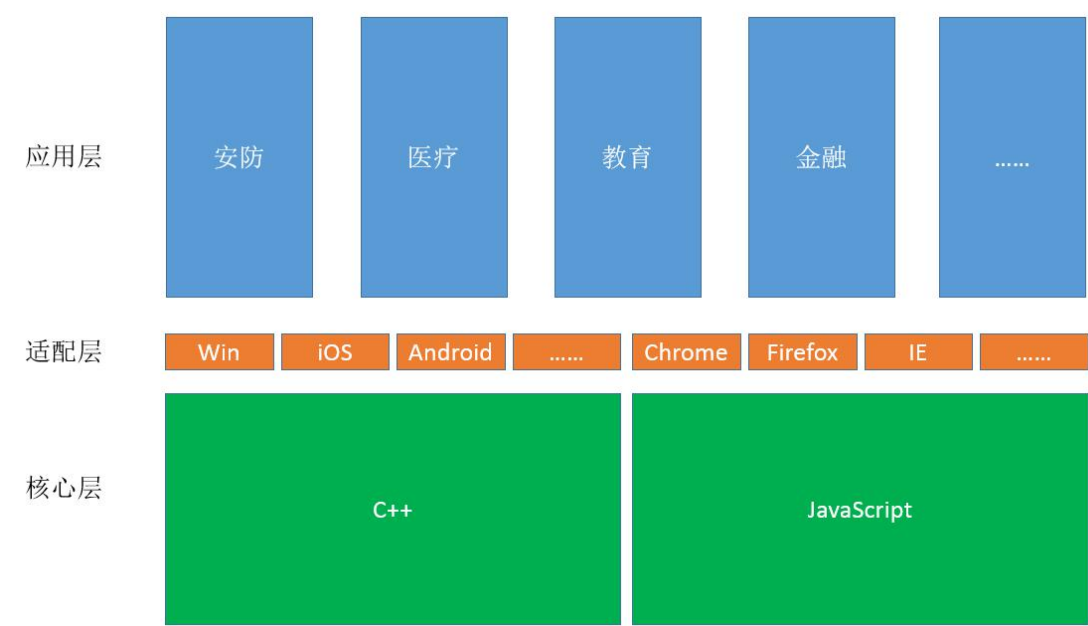
注：目前阶段，主要提供第 3 级接口，第 2 级接口正处于开发过程中。

● 接口语言

支持主流的编程环境及语言，如

- JavaScript（for Web browser）
- C++（for Windows, Linux）
- Objective C（for iOS, Mac）
- Java（for Android）
- C（for 嵌入式 Linux 或 Windows）
- C#（for Windows）

2.2.4 SDK 分层



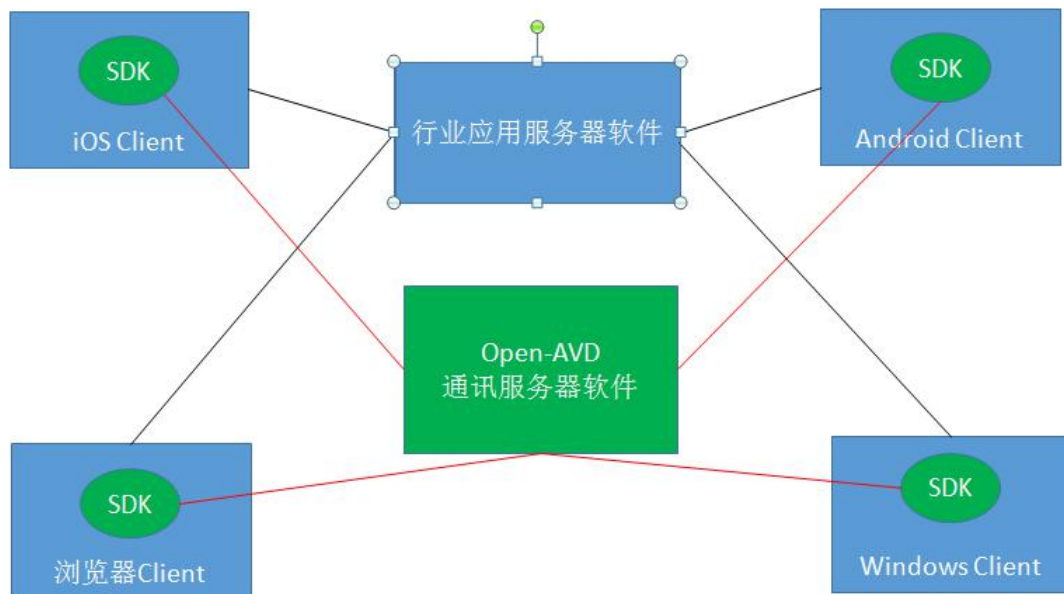
如上图所示，SDK 本身分成两层，核心层+适配层，在 SDK 之上，是客户开发的应用层。

所有和音视频通讯相关的核心代码，都封装在核心层的绿色方块代表的 C++库和 JavaScript 库中，在适配层是非常薄（因而工作量小）的一层代码，做了少量开发，以在需要的地方对不同的操作系统或浏览器做适配。

客户开发团队开发的应用层代码则根据产品需要，调用不同平台的适配层接口完成相应的功能，完全不用理会功能细节及一些技术细节，如线程模型等等。

2.2.5 客户产品架构

使用了 Open-AVD SDK 的第三方业务系统，其基本架构如下图所示：



如上图所示, 一个使用了 SDK 的第三方行业应用, 主要由两部分组成, 分别是 SDK 相关功能组件 (绿色) 及行业应用自身特定功能组件 (蓝色)。

SDK 本身, 也是由两部分组成, 一是嵌入在各种业务应用客户端中的 SDK 代码, 二是 SDK 通讯服务器。SDK 代码与服务器代码通过网络 (局域网、互联网、移动互联网) 连接, 交换音视频及其它数据, 从而完成相关功能。

行业应用的开发人员在需要使用音视频通讯能力的时候, 简单地调用 SDK 中提供的各种接口即可, 不必再去搞清楚音视频通讯过程需要用到的非常复杂的技术细节, 如音视频采集、压缩、传输及渲染等等。

行业应用自身与音视频通讯无关的功能, 则完全由行业应用开发工程师根据需要自己开发, 相关数据传输、存储也完全由行业应用业务服务器自己来处理, 与 SDK 没有关系。

需要说明的是, SDK 通讯服务器中安装了所谓的 USB Key (俗称硬件加密狗) 来控制整个系统能使用的最大音视频并发数, 从而控制 Box/私有云模式中软件 license 的数量。

2.2.5 设备与平台

Open-AVD SDK 支持现在流行的所有硬件设备及软件平台, 可以运行在各种移动终端、桌面系统乃至机顶盒和智能电视上。

2015 年 3 月, 在 HTML5/WebRTC 标准通过仅仅 4 个月, Open-AVD SDK 就在国内第一个宣布对它的全面兼容和支持, 使得叁体成为业界第一家支持 Htm15/WebRTC 的公司, 能在浏览器中直接启动多方音视频通讯及数据协同功能, 免除了用户在浏览器中下载客户端的各种麻烦,

开创了极致的用户体验。

以下为支持的主要设备与平台：

表 1、设备支持表

支持的设备	支持方式	接口语言
iPhone, iPad	原生 iOS 本地客户端	Objective -C
Android Phone, Android Pad	原生 Android 本地客户端	Java
	Webrtc 兼容浏览器，如 chrome/firefox	Javascript
Windows PC	原生 Windows 本地客户端	C++
	Webrtc 兼容浏览器	Javascript
Mac PC	Webrtc 兼容浏览器，如 Chrome/firefox	Javascript
Linux PC	Webrtc 兼容浏览器，如 Chrome/firefox	Javascript
机顶盒	Android 本地客户端	Java 目前主流的机顶盒都基于 Android 开发
嵌入式 Linux 设备，如摄像头等等	原生 C 本地程序	C
智能硬件，如机器人	原生 Android 本地客户端	Java

表 2、平台（操作系统）支持表

支持平台	备注
iOS	手机、平板
Android	手机、平板 智能硬件、机器人 机顶盒、智能电视
Windows	Windows 桌面 专用工控机
Mac	
桌面 Linux	
嵌入式 Linux	一些硬件设备，如摄像头
Webrtc 兼容浏览器	

2.2.6 服务器类型

根据可靠性及复杂度不同，Open-AVD SDK 的通讯服务器有不同的类型，现简要描述如下：

服务器类型	描述	特点
-------	----	----

单机	服务器软件只有单个实例	可靠性低，如果出现服务器软件故障，音视频通讯服务将不可用 实现简单，运维简单
集群	在同一个数据中心中部署很多个物理服务器运行多份 SDK 通讯服务器软件实例，根据房间内并发人数大小，系统选择多个物理服务器一起工作来提供服务 覆盖了双机热备的所有能力	高可靠性 高延展性，可支持数千人在同一个房间中
分布式集群	在不同地方的数据中心分别部署服务器集群，所有这些集群中的服务器可以为单个房间的客户端提供音视频服务 覆盖了集群的所有能力	高可靠性，可实现异地灾备 客户端可就近智能选择数据中心连接，可实现跨境的大规模实时多方音视频通讯

注：叁体服务器支持 Linux 及 Windows 操作系统软件。

2.3 幸会与爱秀

“幸会”，是基于 Open-AVD SDK 中间件打造的一个云会议产品，提供所有会议系统的标准功能，如文字聊天、多方实时音视频、旁路直播、屏幕共享、音视频录制等功能。

“幸会”支持所有的平台，尤其是支持 HTML5/WebRTC 兼容浏览器，能直接从浏览器里启动、加入会议，不用下载任何插件。

“幸会”基于 Open-AVD SDK 开发，目前免费使用，代码开源，主要目的是演示 Open-AVD SDK 的能力，同时为第三方团队使用 Open-AVD SDK 的 API/SDK 提供样例代码。

“爱秀”则是叁体的直播能力演示程序。

3. 技术特点与优势

叁体团队核心成员来自于 WebEx 和 Cisco 这种做网络视频会议和通讯系统的公司，在音视频及通讯产品的技术方面有多年的积累，在音视频处理、音视频通讯、系统稳定性、安全性、大并发、高可用等方面，有丰富经验。

同时，作为一个新锐的技术型创业公司，叁体有能力和机会紧跟技术发展的最新趋势，兼容业界近几年发展出来的最新的成熟技术，打造出全球领先的技术平台，引领技术潮流。

现对叁体的主要产品 SDK 的技术特点与优势分别描述如下：

3.1 全平台支持

2014 年 10 月 28 日，在历经 8 年多的艰苦努力后，万维网联盟 W3C 终于宣布 HTML5 标准获得通过，这标志着新一代 Web 应用标准化的努力获得了决定性的进展。

除了其它很多革命性的创新功能外，HTML5 中最引人入胜的特性就是 WebRTC（Web Real Time Communication – Web 实时通讯），它第一次使用一种标准的方法，在 WebRTC 兼容浏览器中原生引入了音视频和双向通讯能力，无需下载任何附加插件或本地 Client，就能在浏览器里实现音视频聊天、会议等各种功能，开创了极致的用户体验。

叁体团队核心技术成员，一直站在音视频通讯技术的最前沿，通过和 HTML5/WebRTC 标准化组织专家组重要成员单位的合作，4 年前 WebRTC 刚刚诞生之初，就开始研究跟踪 HTML5/WebRTC 相关技术，在兼容 HTML5/WebRTC 技术开发音视频及通讯产品方面，在业界遥遥领先。

2014 年 12 月，在 HTML5 标准公布后仅 2 个月，叁体团队就注册成立了“杭州叁体网络科技有限公司”，推出了兼容 HTML5/WebRTC 的革命性产品，Open-AVD SDK 音视频通讯+数据实时协同中间件和“幸会”云会议系统，成为业界首家能无缝支持所有平台的音视频技术服务公司。

关于叁体产品支持的设备与平台详细情况，请参考 2.2.5 设备与平台。

关于 HTML5/WebRTC 更多信息，详见附录 1。

3.2 领先音视频技术

针对互联网尤其是移动互联网的特点，叁体设计和使用专门的技术方案并做了专门的优化与调校，使得系统即使在网络情况不佳的情况下，仍能保持音视频的清晰、流畅与稳定。

尤其是，在唇音同步、音频回声消除与降噪、视频动态码流、视频混合编解码、视频硬件编解码、视频多流等方面，叁体远远领先于市场。

现把叁体 Open-AVD SDK 音视频特点总结如下：

表 2.Open-AVD SDK 音频特点

项目	指标、值	说明
编解码	iSAC, iLBC, opus Speex, AAC	iSAC, iLBC 专门为互联网环境设计，在 3G/4G 环境下表现优异，质量远优于传统技术 可扩展支持 Speex, AAC 等编解码

回声消除	优秀	直接使用一般的笔记本外放话筒也都不会出现回声或啸叫等异常
语音动态码流	支持	支持语音动态码流, 根据网络的情况对码流进行自适应调节, 因而可以在不同的网络环境中保持出色的语音质量, 即使在 2G 网络环境中也能通话
语音处理	优秀	具有出色的降噪、自动增益控制技术, 保持语音清晰响亮
网络处理	优秀	具有先进的抖动缓冲和丢包补偿技术, 保持语音稳定、低延时
音质	电话音质	使用了最优秀的语音处理与语音编解码技术
混音	支持	

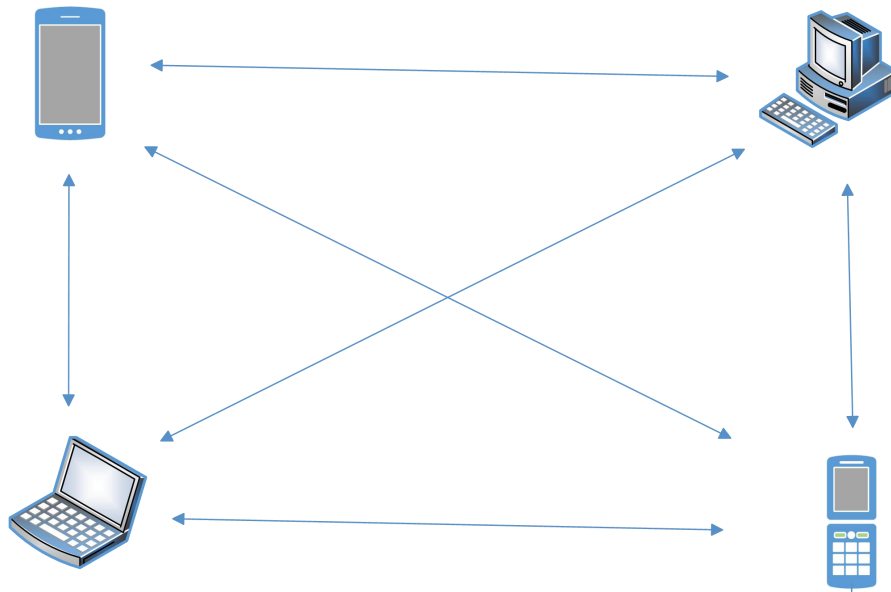
表 3.Open-AVD SDK 视频特点

项目	指标、值	说明
编解码	VP8/VP9 H. 264	VP8/VP9 专门为互联网环境设计, 在 3G/4G 环境下表现优异 可扩展支持 H. 264, 并可支持硬件 H. 264 编解码 可在单个会议类场景中支持 VP8 与 H. 264 混合编码
CPU 消耗	低	
码流	低	根据网络情况智能调整分辨率与帧率等核心指标, 从而动态调整码流
视频规格	全面	支持从 176*144 到 1920*1080 的各种规格视频
图像质量	优秀	出色的图像处理技术, 保持图像清晰稳定
流畅度	优秀	具有先进的网络处理技术, 保持图像的流畅度
视频多流	支持	编码端将视频编码成高、中、低三路不同分辨率的码流, 发送到服务器, 服务器根据接收端的不同类型, 发送相应的码流。比如, 手机终端接收低分辨率码流, 笔记本接收中等分辨率码流, 大屏幕终端接收高分辨率码流。
支持路数	理论上不限	根据带宽和 CPU 情况确定, 建议 PC 端不超过 8 路, 移动端不超过 3 路

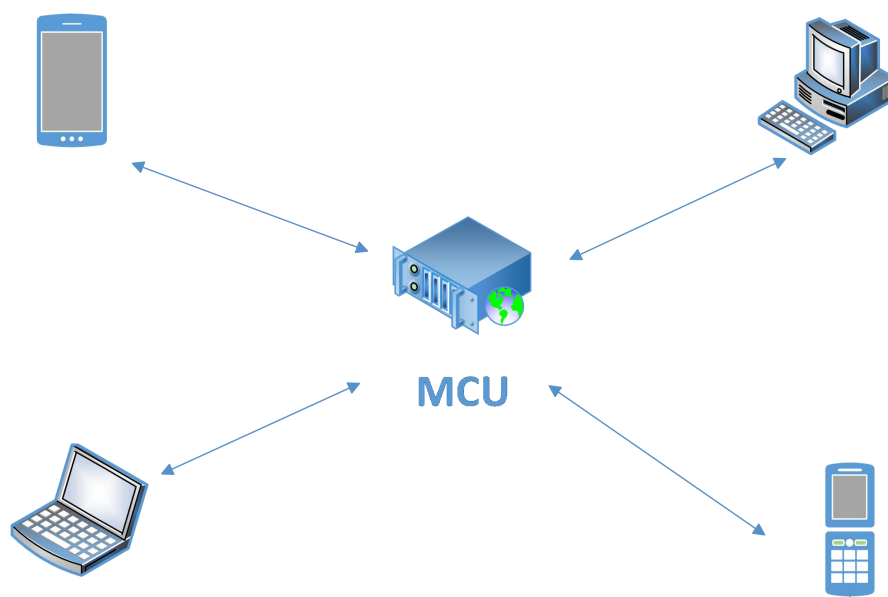
3.3. 领先通讯技术

3.3.1 P2P 与 MCU

从架构上讲, 我们有两种模式来搭建整个音视频通讯系统, 分别是 P2P(Peer-to-Peer)和 MCU (Multipoint Control Unit)。



如上图所示, P2P 最大特点是, 所有音视频的信息直接在通讯的两个客户端之间传递, 无需经过服务器的中转, 在参与通讯的客户端较多时, 每个客户端都需要和其它所有客户端建立通讯链路。



另一种同样被广泛使用的架构是 MCU，其最大特点是，两个通讯的客户端之间并不直接相连，所有的音视频信息传递都通过一个处于中心位置的服务器（集群）来转发。

两种架构都有广泛的应用，分别有各自的优势与局限，现简单对比如下：

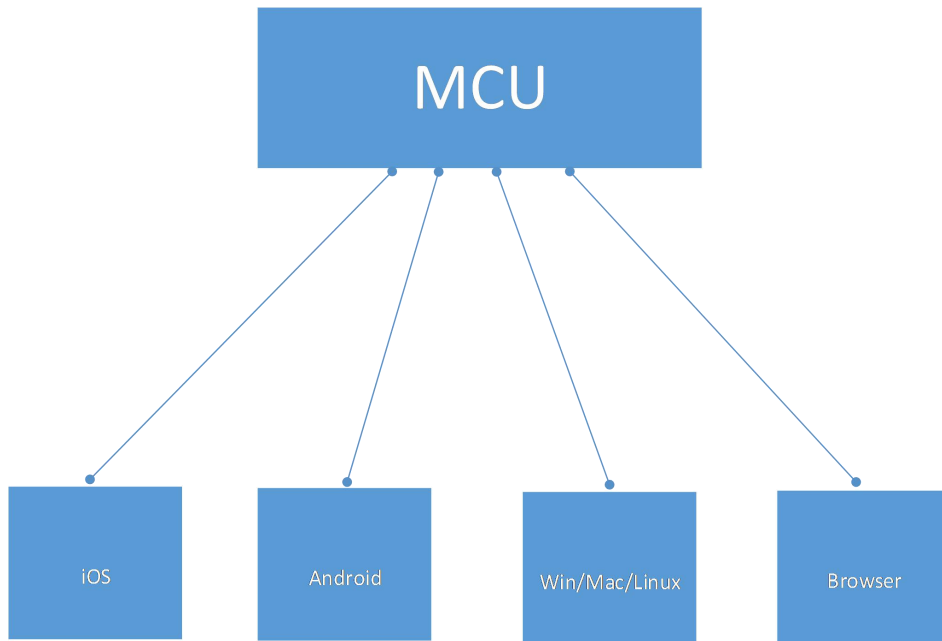
表 4. P2P vs. MCU

架构 比较项目	P2P	MCU
功能特点	所有音视频数据直接在客户端之间传递	所有音视频数据都经由服务器中转
优点	大量小规模会议时，节省带宽	通过服务器集群能支持很大规模的会议，如单个房间 5000 人可支持一些中心化的功能，如服务器端录制
缺点	一个房间只能支持 3-5 人对客户端网络质量要求高	大量小规模房间时，服务器带宽很浪费
典型产品	Skype	WebEx

依托在网络视频会议及通讯领域的多年经验积累，叁体的工程师革命性地设计和实现了同时支持 P2P 和 MCU 两种模式的技术架构，使得系统可根据人员数量做智能动态切换成 P2P 或 MCU 模式，既节约了服务器带宽，又能支持超大并发应用场景，鱼与熊掌兼得。

3.3.2 MCU 服务器

MCU 服务器的最重要设计目标，是要保证所有非浏览器客户端（如 iOS, Android, Windows/Mac/Linux）能和浏览器内包含的 WebRTC 客户端（JavaScript）互通，为此，叁体的资深工程师破解了 webrtc 的核心实现，采用了和浏览器实现类似的技术，完成了各种不同类型客户端通过 MCU 的互通。



尤其值得一提的是核心协议栈的实现。

● 协议栈

MCU 服务器实现了 HTML5/WebRTC 所有核心协议栈，同时，在非浏览器的本地客户端中，实现了同样的核心协议栈，保证所有不同的设备能和浏览器中 WebRTC 客户端通过相同的机制进行通讯并互操作。

除了的一些已有的核心协议栈实现外，考虑到功能实现及性能优化等众多因素，Open-AVD SDK 的 MCU 服务器，自己实现了包含以下所有协议在内的所有核心协议栈：

- stun
- ice
- rtp
- rtcp
- http
- websocket
-

3.3.3 多核 CPU 利用

随着近年来 CPU 制造技术的突飞猛进，16 核以上的 CPU 越来越多，而能否充分使用所有内核的能力，对于大规模 PaaS 平台的运营十分关键。

和一些传统的音视频厂商不同的是，叁体汲取了多年音视频及通讯系统设计中的经验教训，

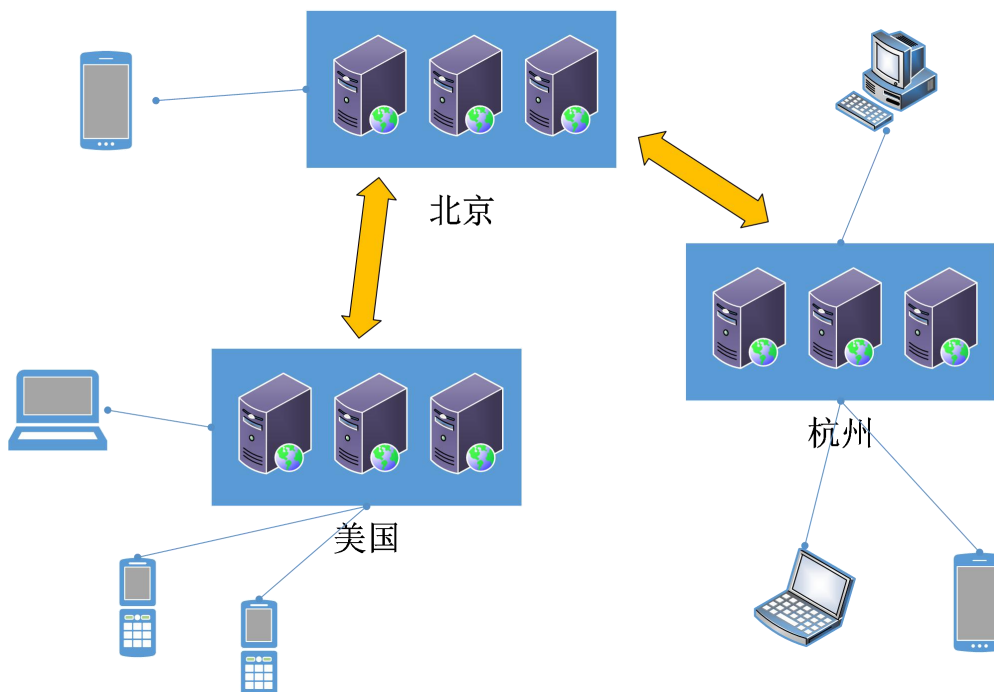
在成立之初就把充分利用多核 CPU 能力设定为最重要的技术目标之一, 在服务器技术选择、架构设计和技术实现方面, 都做了精心处理, 使得系统在运行的时候, 能充分利用所有 CPU 内核的能力, 避免了计算资源的浪费, 从而在节能降耗及 IDC 运营成本方面, 获得明显的优势。

3.3.4 分布式服务器集群

由于历史原因, 国内互联网的网络质量存在着地区差异, 不同互联网 ICP 之间, 不同地域之间, 都存在一定的联通性问题, 同时, 跨境的互联网访问, 也存在巨大的限制, 这对于运营一个全国性乃至全球性的音视频+数据实时协同 PaaS 平台, 是一个非常大的挑战。

叁体对于这个问题的解决方案是, 设计开发了一个分布式的服务器集群架构(如下图所示), 在全国乃至全球多地部署服务器集群。“本地”的 client 根据一定智能算法, 选择最佳(通常是离自己最近的)服务器集群连接, 而服务器集群之间则由相关 ICP 通过特殊的手段来保障其连接, 这样, 所有 client 都能获得很好的网络连通性从而保证整个协同沟通的质量。

展望未来, 通过和 ICP 的合作, 叁体将开发和部署一套类似于 CDN(Content Delivery Network)的分布式系统, 为所有合作伙伴提供良好的实时音视频和数据共享能力服务, 我们叫它 RCDN(Realtime Content Delivery Network)。



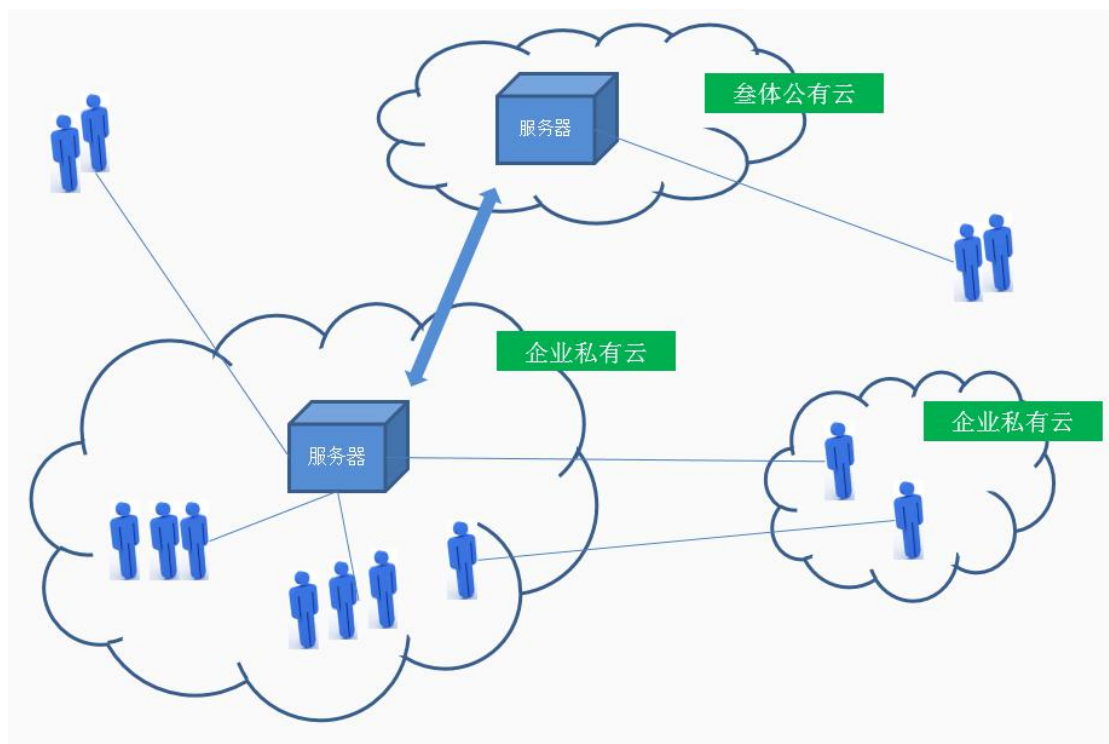
3.3.5 混合云

从部署架构上讲, Open-AVD SDK 支持“公有云”和“私有云”, 同时还支持“混合云”部署方式, 混合云的相关部署架构如下图所示。

在这个部署架构中, 最具创造性的是, 使用叁体研发的专有技术, 使得公网上的 client 能和内网上的 server 互通, 使得一个内网上的 client 可以和另一个内网上的 server 互通。

这里的要点是, 叁体公有云里部署一些 server, 同时, 在有能力的大客户自己的一个或数个私有云中部署另一些 server, 这些 server 按照上述的分布式集群机制连接在一起, 共同为内网或外网上的 client 提供音视频和数据共享能力。

使用这种部署架构, 在为一个拥有数百人或数千人的客户提供 Open-AVD SDK 服务的时候, 可以把所有内网上的 client 和大部分外网上的 client 都连在客户自己的私有云, 客户私有云和叁体公有云之间的流量非常小, 其通讯质量由 IDC 服务商来保证。



这样的部署架构, 带来以下四大优点

- 提升了用户体验

大量 Client 通过局域网连 MCU 服务器, 网络速度和质量都非常有保障, 用户体验变得超级好。

- 降低了公有云带宽

所有私有云内的 client 都通过局域网连接私有云中的 MCU, 一部分外网上的 client 也连在了

私有云中的 MCU，直接连叁体公有云的 Client 数量变得非常少（可能只有总参会人数的几十分之一），这就大幅降低了对叁体公有云带宽的要求，从而可节约大量带宽租用成本。

- 降低了私有云出口（上行）带宽

同上节原因一样，私有云内 client 都通过局域网连接私有云中的 MCU，不必连叁体公有云，这样，私有云连公有云上行流量也会大幅下降，从而降低了对私有云出口带宽的要求。

- 网络管理简单、安全

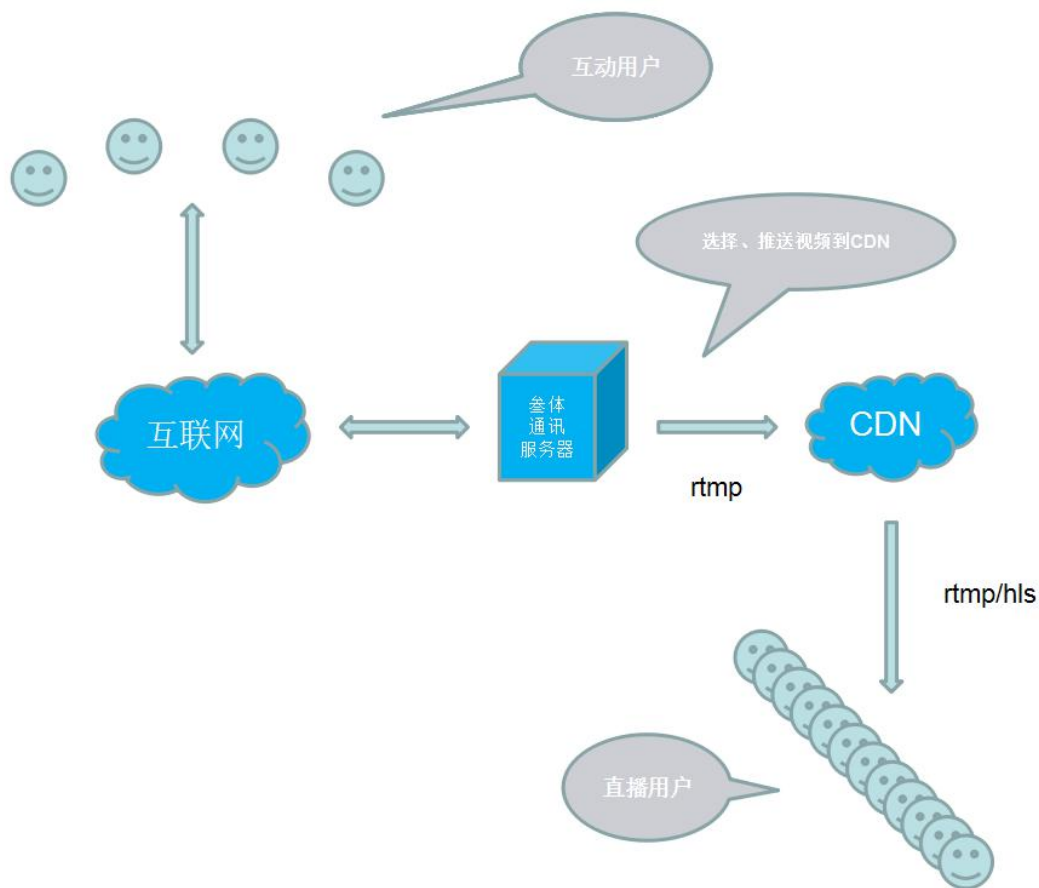
部署信息全部存储于叁体公有云的服务器中，在客户处只部署媒体服务器的一个标准实例，非常简单易行。由于不必将服务器暴露在公网上分配固定的公网 IP，既节约了客户公网 IP 资源，又降低了遭受网络攻击的概率，同时，网络本身的管理还变得简单。

展望未来，混合云的一种可能走向是，叁体可生产一种标准化的服务器硬件，放置在客户的私有云中（内网），这些服务器和叁体公有云连接，一起为内外网上的所有 Client 提供服务。

3.3.6 旁路直播

Open-AVD SDK 除了能支持实时互动音视频，还能选择某路客户端的音视频，通过第三方成熟的 CDN 网络推送给其它各种客户端，可支持实时互动音视频与直播功能的混合使用场景。

其基本架构如下图所示：



直播的基本特性如下：

特性	指标	备注
并发	数十万	
延时	2-10 秒	RTMP: 2-5 秒 HLS: 10 秒
客户端平台	微信客户端 Android/ios 本地客户端 浏览器/Flash	
分辨率	高至 1920*1080	
与实时互动混合	支持	

注：需要向第三方 CDN 供应商租用 CDN 网络

3.4 其它

3.4.1 视频硬件编解码

按目前的技术水平，在采用软件编解码技术的情况下，一般的移动终端能只能支持 320*240

规格的视频, 640*480 只能勉强支持, 1280*720 规格以上的高清视频基本上不可能。

为解决这个问题, 叁体的视频引擎针对市场上的主流 H.264 视频处理芯片进行了优化, 覆盖了 90%以上移动终端, 从而为移动终端上显示高清视频提供了可能性, 目前该特性支持 Android 4.2 和 iOS7 以上系统。

同时, 对于一些比较少见的芯片或深度定制过的 Android 系统, 叁体的视频引擎提供了插件接口, 第三方可开发专门插件来实现 H.264 硬编解码。

3.4.2 视频混合编解码

目前像 Chrome 这种 webrtc 兼容浏览器只支持 VP8, 在和只支持 H.264 的硬件终端共处一个会议场景的时候, 会存在互通问题。

为解决这个问题, 叁体创造性地开发了混合编解码技术, 即在可能的情况下, 在视频采集端同时使用两种编码技术分别编码, 发出两种不同的视频数据流, 由服务器根据连上来的客户端的能力, 选择不同的视频数据发出。

3.4.3 视频动态码流

互联网尤其是移动互联网的网络情况比较复杂, 容易出现各种波动, 造成视频卡顿等问题。叁体的系统在出现网络波动的时候, 会动态调整视频的分辨率及帧率等核心指标, 在网络情况不好的时候降低码流, 从而在牺牲一定图像清晰度的情况下, 保障视频的流畅度。

3.4.4 视频多流

受限于 CPU 的能力和带宽和成本, 可以预见在相当长时间内, 并不是所有移动终端都会或能处理 1280*720 规格以上的视频。

为保证参与音视频通讯的所有客户端都能参与到有效的沟通过程中, 叁体支持“多流”技术, 即对某一路视频源, 可传输几种不同规格的视频数据, 服务器根据一定的策略(如终端的 CPU 能力、网络的状况)智能选择合适的视频发给相应的客户端。

3.4.5 Sip/H.323

叁体提供对 Sip/H.323 协议的支持, 可方便集成 Sip/H.323 兼容的硬件设备, 从而保证客户的大量投资不会被浪费。

3.4.6 RTMP

叁体提供对 RTMP 协议的支持，可方便集成市场上已经存在的大量支持 RTMP 协议的音视频系统，如教学环境中使用的录播系统。

3.4.7 RTSP

叁体提供对 RTSP 协议的支持，可方便集成市场上已经存在的大量支持 RTSP 协议的音视频设备和系统，如海康和大华等监控厂商生产的 IPC（网络摄像头）。

3.4.8 第三方音视频流倒入与倒出

针对一些特殊设备（如智能眼镜），叁体提供了第三方音视频倒入与倒出功能，从而满足一些特殊设备的特殊音视频处理与传输需求。

3.5 重要技术特性与指标

类别	项	值	备注
音视频	唇音同步时差	<100 毫秒	唇音完全同步，完美用户体验
	音频优先	支持	在网络不好情况下优先保障声音的质量
视频	视频路数	1-16	取决于视频规格、客户端运算能力及网络环境，建议移动端不超过 3 路，桌面端不超过 8 路
	视频规格	176*144 320*240 352*288 640*480 720*576 1280*720 1920*1080	
	视频编解码	VP8/VP9 H.264/H.265	
	硬件编解码	支持	H.264 硬编解码
	动态分辨率调整	支持	可根据网络情况和 CPU 能力，动态调整视频发送端的分辨率，从而保证在网络情况不佳或 CPU 性能不好的时候，保持

			视频的流畅度
	多层编码	支持	可根据不同的客户端，选择不同规格的视频发送
音频	音频路数	无限制	
	音频编解码	iSac iLBC	
	回声消除	支持	优秀，笔记本外放无回声
	噪声消除	支持	优秀，电话音质
	自动增益控制	支持	
服务器	P2P 与 MCU	支持	
	MCU 级联	支持	
	负载均衡	支持	
	水平延展	支持	
	分布式集群	支持	客户端可就近连接最快最稳定的服务器集群，保证极致的用户体验，在跨境协同沟通应用如远程会诊中，意义重大。
安全	传输安全协议	DTLS	
	数据加密算法	AES	
网络	防火墙穿越	支持	自动穿透所有防火墙及代理服务器
	接入方式	有线网络 Wifi 3G/4G	
	端口开放要求	80/443	
	抗丢包率	30%	在 30%网络丢包情况下，仍能保持音视频质量
其它	支持平台	所有浏览器 iOS/Android Windows/Mac/Linux	HTML5/WebRTC 兼容浏览器，不用下载，其它浏览器可下载插件及本地客户端来支持
	支持设备	PC/Mac 桌面系统 iPhone/iPad Android 设备 机顶盒，智能电视 各种嵌入式 Linux 设备	
	H.323	支持	
	SIP	支持	
	RTMP	支持	
	RTSP	支持	

4 附录

4.1 HTML5 与 WebRTC

众所周知, HTML (Hyper-Text Markup Language, 超文本标记语言) 是 Web 页面的基础语言, 由 W3C (万维网联盟) 负责制定相关标准。HTML4.0 在 1999 年正式发布, 为整个互联网的发展起到了决定性的作用。

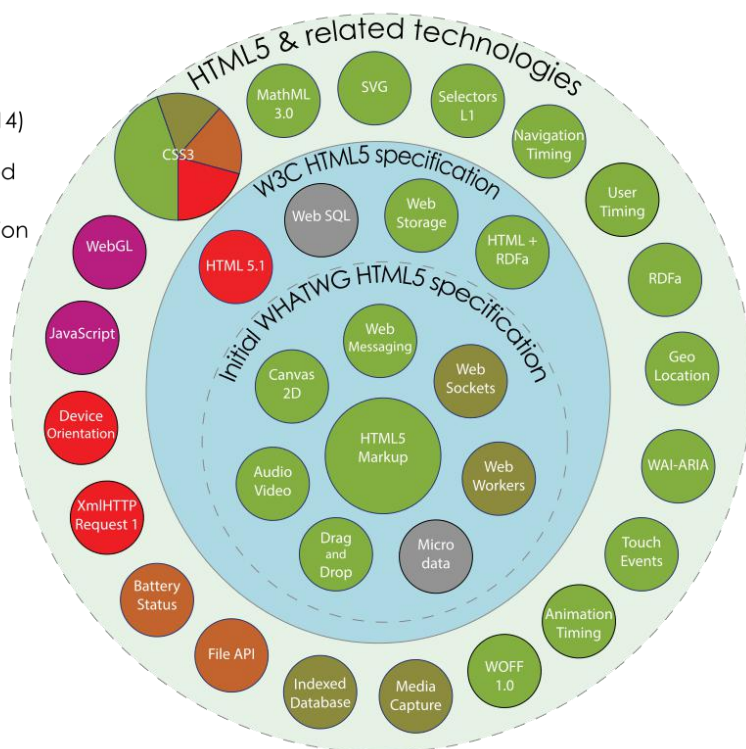
随着 10 几年来互联网的快速发展, 人们对 Web 应用的需求变得日益丰富, 从简单的文字浏览提升到实时音视频等高级特性的追求, 为顺应市场的需求, W3C 在 2006 年启动了制定第 5 代 HTML 标准的努力, 也就是大家熟知的 HTML5, 简称 H5。

2014 年 10 月 28 日, 在经过长达 8 年的艰苦努力后, W3C 终于宣布 HTML5 标准规范制定完成, 开始正式推荐工业界使用。

HTML5

Taxonomy & Status (October 2014)

- Recommendation/Proposed
- Candidate Recommendation
- Last Call
- Working Draft
- Non-W3C Specifications
- Deprecated or inactive



如上图所示, HTML5 技术标准包含内容非常多, 主要涵盖以下几方面:

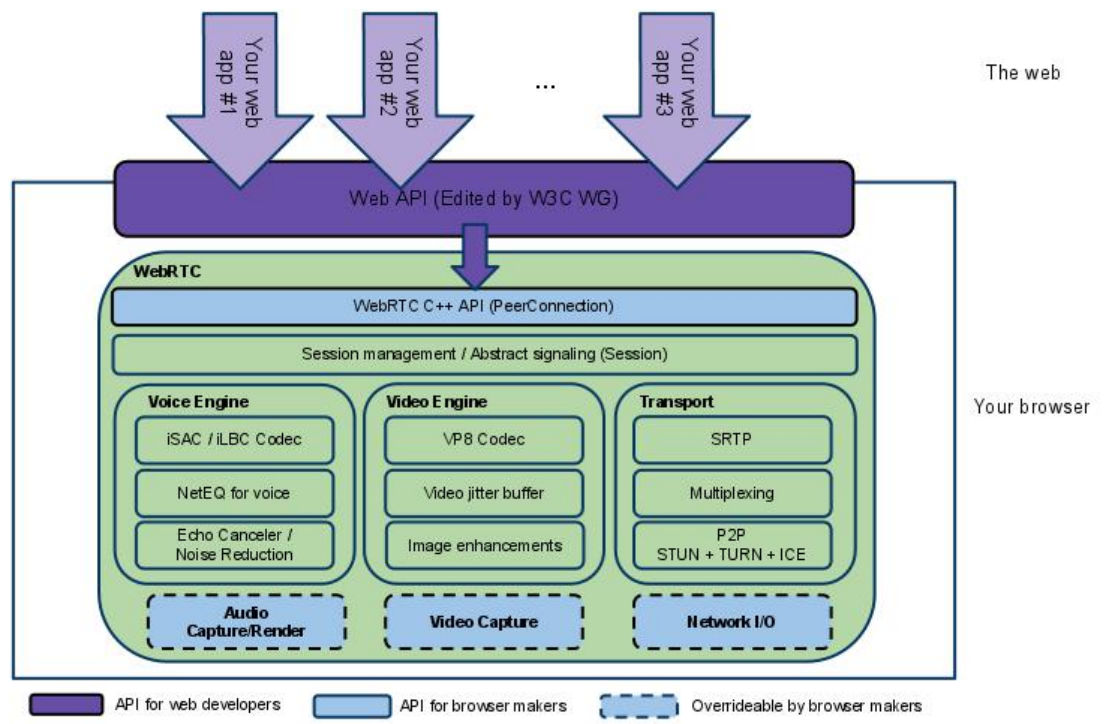
- 实时音视频
- 双工数据通讯
- 2 维绘图 API Canvas
- 浏览器内部数据存储
- 浏览器拖放能力

- 语言的规范与增强
- 浏览器本地存储
- 工作者线程
-

其中，和实时音视频相关的部分叫 WebRTC (Web RealTime Communication, Web 实时通讯)。

WebRTC 的目的是让 Web 开发者能够基于浏览器（如 Chrome、FireFox 等）轻易快捷开发出丰富的实时音视频应用而无需下载安装任何插件，开发者也无需关注实时音视频的数字信号处理（如音视频编解码）和传输的复杂技术，只需编写简单的 JavaScript 程序即可。

以下为 WebRTC 的一个简单的模块图



从概念上讲，WebRTC 底层是三个库，分别是音频引擎，视频引擎和通讯库，在这三个库之上，实现了 WebRTC 的核心概念 Session 和 PeerConnection，所有浏览器厂商再在 Session 和 PeerConnection 层之上，实现了一套标准的 Web API(Javascript)，这样，所有 WebRTC 兼容的浏览器里就能直接运行音视频应用并能互操作。

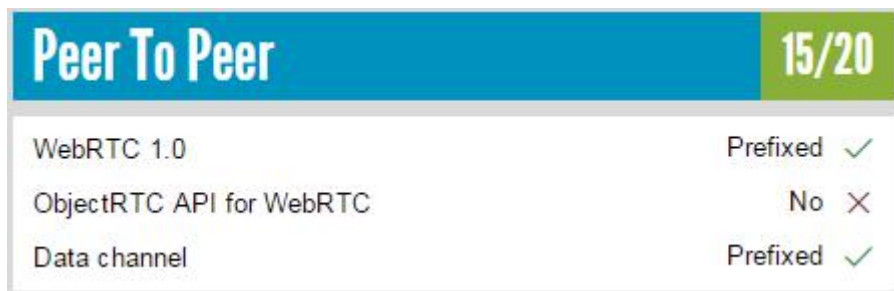
WebRTC 的名字中有 Web 两个字，缘于它的原始目的是为 Web 增加实时音视频能力，但其实它的应用并不限于 Web 应用中，而是可以用到 PC, Android/iOS 等各种平台的原生应用程序中，甚至还可以应用到服务器的开发与实现中，这样，基于 WebRTC 技术，就可以实现包括浏览器在内各种平台下的 client 应用互通音视频。

目前，并不是 HTML5 的所有特性都被所有浏览器支持，尤其是和 WebRTC 相关的部分，获

得支持的程度仍然不是非常高。但由于 HTML5 标准已经在 2014 年 10 月 28 日通过，从长远来看，相信随着时间的推移，所有浏览器都会逐渐对包括 WebRTC 在内的所有 HTML5 特性提供全方位支持，比如，近期微软已经宣布要在 IE 中提供对 WebRTC 的支持。

可以访问 <http://www.HTML5test.com> 来对浏览器 HTML5 的兼容性进行测试。测试的满分是 555 分，目前对 HTML5 所有特性支持得最好的 Google 浏览器得分 506 分。

如果要使用 WebRTC 来实现音视频能力，“Peer to Peer”类别中的“WebRTC 1.0”和“Data Channel”两项，一定要是“勾”的状态，如下图所示：



自 2014 年 10 月 28 日 HTML5 标准推出以来，得到越来越广泛的支持，其中包括微软、Apple 等业界巨头。

在 2014 年 10 月 29 日，HTML5 标准刚公布的第二天，微软就宣布支持 webrtc，详见以下报道：

<http://www.iefans.net/webrtc-bangzhuie-jicheng-skype-quanbutiyan-bingzhichi-shishi-tonghua/>

。同时，近期微软的 Win10 自带的最新浏览器 Edge 已经宣布支持 webrtc。

另有消息称，爱立信（WebRTC 标准的最有力推动者之一）正联合 AT&T 和 Apple 谈判，希望在 At&T 定制的 iPhone 手机中，让 Safari 浏览器原生支持 WebRTC。

目前包括 Chrome/Firefox 在内的主流浏览器都已经能很好的支持 html5/webrtc，随着 IE、Safari 向 html5/webrtc 标准靠拢，可以预期的是，在几个月之内，市场上所有主流浏览器都会支持 html5/webrtc，相关音视频应用会呈爆炸性增长态势。

4.2 术语表

术语	术语解释	备注
SDK	英文 Software Development Kit 的缩写，其含义是软件开发工具包，通常是一个软件公司提供给其它开发公司的工具集合，其它公司可以在 SDK 基础上做开发。	SDK 的典型例子有科大讯飞的语音识别 SDK，也有叁体的视频通讯 SDK。

音频优先	实时音视频通讯过程中, 会有大量声音与视频数据需要传输, 而视频的数据量远大于音频, 而从沟通效果来说, 保证音频数据的传输远比视频重要。所以, 在网络状况不良 (比如带宽受限) 的时候, 需要采取特殊的技术, 可以对视频数据进行丢弃处理, 以优先保证声音数据的传输, 这就是所谓音频优先。	音频优先会带来唇音同步问题需要解决
唇音同步	由于实时音视频传输过程中, 声音与视频数据是由不同的通道来传输的, 需要采取专门的技术来保持声音与视频之间的同步性, 从效果上来说, 需要保持说话时的嘴唇动作和声音的一致性。	叁体的唇音同步时差小于 100 毫秒, 远远优于市场上一些做网络会议的公司, 是赢得类似大华这种大客户的关键
回声消除	声音在播放过程中, 由于场地环境等因素, 扬声器传出的声音, 会从麦克风传回系统, 从声音听起来有类似山谷里回声现象, 影响收听效果, 严重的时候, 会形成 “啸叫”, 完全听不清楚。 回声消除就是采取各种手段降低乃至避免回声的技术, 是声音处理的关键技术, 直接关乎音频通讯的质量。	叁体的回声消除技术业内遥遥领先, 笔记本外放, 完全无回声。
iLBC	Internet Low Bitrate Codec, 互联网低码率编解码, 是一种开源的窄带宽话音编解码技术, 由 webrtc 原生支持的音频技术, 专门为互联网环境设计, 配合相关音频引擎, 在较差的网络条件下仍能保持较好的音质。	
iSAC	Internet Speech Audio Codec, 互联网话音编解码, 是一种开源音频编码技术, 由 webrtc 原生支持, 专为互联网环境设计。	
降噪	声音在传输处理过程中, 会有各种电子噪音混入, 从而影声音质量。降噪就是采取各种技术手段, 过滤噪音。	叁体的降噪技术业内遥遥领先, 音质和电话音质不相上下。
H.264/H.265	H.264 是一种主流的视频编码技术, 在安防等行业大量使用 H.265 是脱胎于 H.264 的一种新的视频编码技术, 预期 2016 年开始会逐渐获得工业界应用	作为叁体强大技术能力的一个标志, 叁体同时支持这两个系列的编码技术, 并能在同一个通讯场景 (如多方会议) 中, 混合使用这两种编码。

VP8/VP9	VP8 相对于 H.264, 是一种相对新的视频编码技术, 专门针对互联网通讯环境进行了优化, 在互联网行业应用广泛, 最典型的应用是在 html5 的音视频规范 webrtc 中。 VP9 是脱胎于 VP8 的新编码技术, 逐渐开始流行, 并获是了主流芯片厂商的支持。	
硬件编解码	视频编码与解码, 其中一种方式, 是使用手机或 PC 的 CPU 来做大量运算, 这个过程叫软编解码。 如果 CPU 本身不够强劲, 运算过程不够快, 在处理比较高清的视频(如 720P 或 1080P)就会造成各种视频问题, 如卡断、延时等等。 解决这个问题的一种办法是使用手机或 PC 上已经自带的硬件芯片来做编解码, 从而避免上述各种问题。	叁体 Open-AVD SDK 针对主流芯片提供了硬编解码能力, 从而保证大多数 Android、iOS 移动终端能处理高清视频。
多流	指摄像头在采集视频数据的时候, 采集不同规格(分辨率、帧率)的多份视频数据传输到通讯服务器, 服务器根据连接上来的不同客户端的能力(如是否是运算能力比较弱的移动终端)选择不同的视频进行传输, 从而保证参与视频通讯的所有客户端都能看到流畅的视频。	叁体 Open-AVD SDK 支持多流, 极大提升了用户体验
动态码流	指在视频信号处理传输过程中, 如果网络出现波动导致传输不畅的时候, 采取降低视频分辨率及帧率等手段降低传输码流, 从而使得在网络状况不好的时候, 仍能够保持视频的流畅(质量会有所下降), 待网络情况变好的时候, 系统自动恢复到更高的分辨率及帧率。	叁体有非常优异的动态码流技术
P2P	Peer to Peer, 意指点对点通讯, 这是音视频通讯的典型架构, 参与通讯的 client 直接通过网络传输音视频数据, 不必通过服务器中转。 这种通讯架构的好处是, 大量小规模场景(如远程会诊), 可以节约大量服务器(数据中心)网络带宽。	Skype 是典型的 P2P 架构, WebEx 是典型的 MCU 架构 叁体同时支持这两种架构
MCU	Multi-Control Unit, 多点控制单元, 是	

	硬件视频会议中的一个设备, 用于在硬件视频会议终端间传输数据并提供控制功能。 目前这个词的使用已经超出了它的原始含义, 在软视频会议系统中, 主要指用于转发音视频数据并提供控制功能的媒体通讯服务器。 采用了 MCU 的音视频通讯架构, 是和 P2P 相对应的一种通讯架构, 在业界也大量使用, 其典型优点是能支持大量并发用户, 如超过 5000 人的会议。	
分布式服务器集群	指服务器集群分布在物理位置不同的城市和国家, 使得 client 端可以采取就近远程最佳服务器来连接, 从而保证所有客户端最好的用户体验, 如很低的延时、优质的音视频质量等等。	
HLS	Http Live Streaming, Http 活动流, 一种 Apple 公司提出的流媒体技术, 和 RTMP 技术一样, 是目前市场上主流的直播服务支撑技术。	叁体 Open-AVD SDK 已经完成基于 HLS 技术的直播功能, 可在微信里直接看音视频直播。
RTMP	Real Time Messaging Protocol, 实时消息传输协议, 是一种设计用来进行实时数据通信的网络协议, 主要用来在 Flash/AIR 平台和支持 RTMP 协议的流媒体/交互服务器之间进行音视频和数据通信。 目前市场上很多设备和系统, 如大华海康的摄像头, 支持 RTMP 协议。	叁体 Open-AVD SDK 可以和支持 RTMP 协议的设备/系统集成。
RTSP	Real Time Streaming Protocol, 实时流传输协议, 该协议定义了一对多应用程序如何有效地通过 IP 网络传送多媒体数据。 目前市场上很多设备和系统, 如教育领域里的录播系统, 支持 RTSP 协议	叁体 Open-AVD SDK 可以和支持 RTSP 协议的设备/系统集成。
SIP	Session Initial Protocol, 会话初始协议, 在一个广泛应用的信令控制协议支持 SIP 的各种通讯 client, 如 sip 电话, 软电话终端可以互相通讯	叁体 Open-AVD SDK 可以和支持 SIP 协议的设备/系统集成
H.323	国际电信联盟 ITU-T 于 1996 年公布了 H.323 协议, 它是局域网和企业网使用的多媒体通信标准。H.323 是 Internet	叁体 Open-AVD SDK 可以和支持 H.323 协议的设备/系统集成。

	上端与端之间进行实时声音和视频会议的规程和协议，包括系统和构件的描述、呼叫模型的描述、呼叫信令过程、控制报文、服用、语音编码解码器、视像解码器及数据协议。 H.323 已经有 20 年历史，围绕这个技术协议系统，业界已经生产了大量的硬件设备和系统。	
Html5	HTML 是 Hyper-text Markup Language，也就是超媒体标记语言，是 Web 应用的通用性基础语言。HTML5 是该语言的第 5 代规范，于 2014/10/28 获是标准化组织的批准。	
webrtc	Web real time communication，是 HTML5 标准重要组成部分，专注于解决在 web 浏览器中提供原生的音视频通讯能力的问题	