OpenTSN 时钟同步设计文档

（版本 1.0）

OpenTSN 开源项目组

2019 年 1 月

版本历史

版本	修订时间	修订内容	修订人	文件标识
1.0	2019.1	初版编制		OpenTSN1.0

目录

1、设计目标

时间同步的总目标是实现端系统与交换机之间的亚微秒级时间同步。主从时钟通过同步报文(即 Sync，Delay_req，Delay_resq 三类报文）收集信息以实现时钟同步。同步原理如图 1-1 所示，其中t_1为 Sync 报文的发送时间戳，t_2为 Sync 报文的接收时间戳，t_3为 Delay_req 报文的发送时间戳，t_4为 Delay_req 报文的接收时间戳。

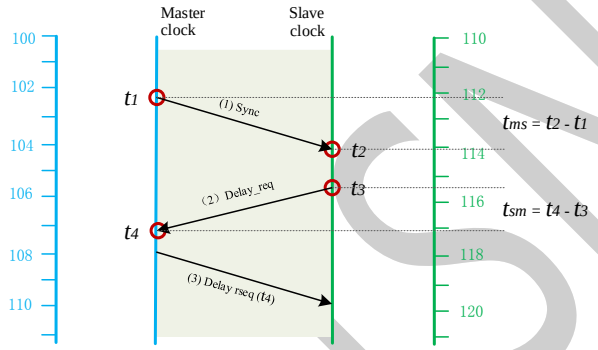


图 1-1 时钟同步原理

主从时钟偏移量计算如下公式所示：

$$t_{ms} = t_2 - t_1 = delay + offset \quad \text{公式 1}$$

$$t_{sm} = t_4 - t_3 = delay - offset \quad \text{公式 2}$$

$$offset = \frac{(t_{ms} - t_{sm})}{2} = \frac{(t_2 - t_1) - (t_4 - t_3)}{2} \quad \text{公式 3}$$

实验拓扑如图 1-2 所示， Pm，Ps，SW 皆为 Openbox-s4，SW 为交换机。Pm 扮演主时钟的角色，Ps 扮演从时钟的角色，SW 扮演从时钟的角色并可以计算非本设备 PTP 报文的透明时钟。同步过程可分为 5 步：

1）主时钟 Pm 定时发送 Sync 报文（报文中携带时间戳t_1）；

2）交换 SW 接收到 Sync 报文后获取t_1与接收时间戳t_2，构造并返回 Delay_req 报文（获取 Delay_req 的发送时间戳t_3），并转发 Sync 报文给 Ps。Sync 为广播报文，会广播到 Ps0 和 Ps1 所在端口。

3）从时钟 Ps 接收 Sync 报文后获取t_1与接收时间戳t_2，构造并返回 Delay_req

报文（获取 Delay_req 的发送时间戳t_3）；

4）主时钟 Pm 接收到 Delay_req 报文后返回 Delay_resq 时，在 Delay_resq 报文中填充 Delay_req 报文的接收时间戳t_4；

5）从时钟 Ps 和 SW 接收到 Delay_resq 后，从此报文中提取t_4，而后根据公式 3 计算主从时钟的偏移量。

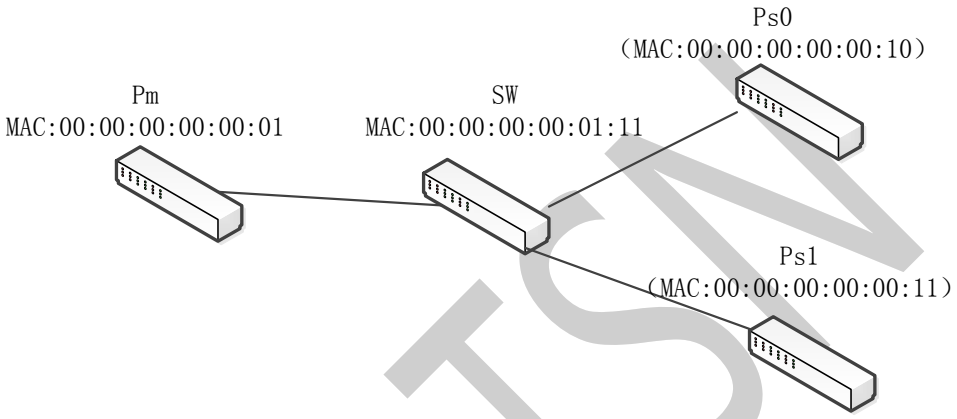


图 1-2 时钟同步实验拓扑

2、概要设计

2.1 设备内时间同步

为了实现时钟同步目标，Openbox-s4 设备内的所有接口时钟需要同步。Openbox-s4 时钟分布如 2-1 所示。其中四个输入端口的时钟分别为 Rx_clk1，Rx_clk2，Rx_clk3 与 Rx_clk4，四个输出端口的时钟分别为 Tx_clk1，Tx_clk2，Tx_clk3 与 Tx_clk4，以及 PTP 的时钟为 Core_clk。其中 Core_clk 为核心时钟，所有接口时钟统称为外围时钟。

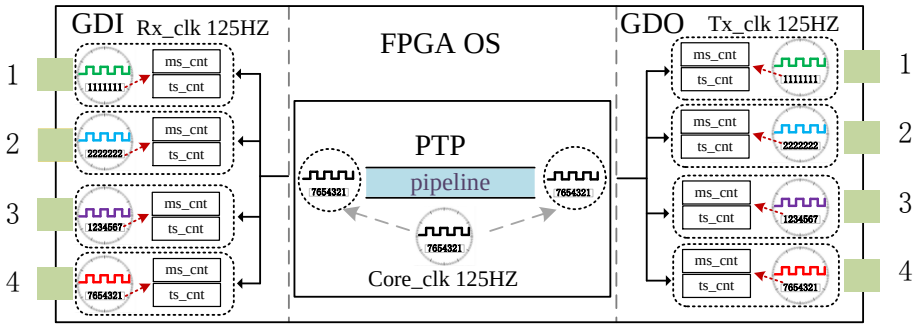


图 2-1 Openbox-S4 时钟分布

Openbox-s4 设备核心时钟与外围时钟同步设计如图 2-2 所示。

1）将 PTP 的核心时钟 Core_clk 设置为主时钟。核心时钟频率为 125MHZ，时钟周期为 8ns，并且维持一个 31 位的计数器 master_ms_cnt 与一个 17 位的计数器 master_ts_cnt。其中，master_ms_cnt 以每毫秒为时间单位，master_ts_cnt 以 8ns 为时间单位；

2) 将所有接口的外围时钟设置为从时钟。外围时钟频率统一为 125MHZ，时钟周期为 8ns，并且每个外围时钟也都维持一个 31 位的计数器 rxN_ms_cnt(txN_ms_cnt，N 为设备端口号)与一个 17 位的计数器 rxN_ts_cnt(txN_ts_cnt)。其中，rxN_ms_cnt(txN_ms_cnt)以每毫秒为时间单位，rxN_ts_cnt(txN_ts_cnt)以 8ns 为时间单位；

3）核心时钟 Core_clk 与所有外围时钟通过 temp_cnt 同步时间。

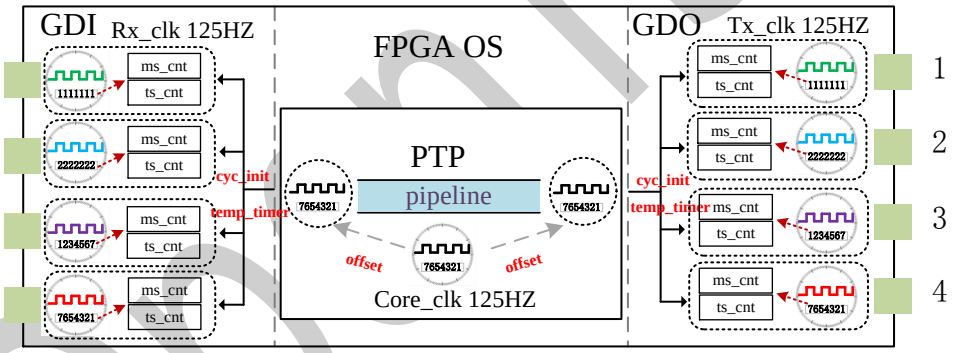


图 2-2 Openbox-S4 设备内时钟同步示意

设备内核心时钟与外围时钟同步流程如图 2-2 所示。

1) 核心时钟 Core_clk 以 1ms 为同步时间，即核心时钟的 master_ts_cnt 计数器值达到 124999 时(同步周期可以修改范围为大于等于 100us 小于 1s）。将 temp_cnt 寄存器的低 48 位赋值给{master_ms_cnt, master_ts_cnt}。同时，将 cyc_init 位置 1，送给所有的接口，并将 temp_cnt、cyc_init 保持三个时钟周期以确保数据稳定传输。

2) 外围时钟检测到 cyc_init 位为 1 后(通过采样上升沿异步处理方式实现)，将

rxN_ts_cnt(txN_ts_cnt)置为 master_ts_cnt，并将 rxN_ms_cnt(txN_ms_cnt)置为 master_ms_cnt。

2.2 设备间时间同步

当多台设备进行同步时，从设备中的 PTP 模块计算出设备间的时钟偏移量 offset 后，从设备根据 offset 值调整核心时钟 Core_clk 的两个计数器的值（master_ms_cnt 与 master_ts_cnt），以达到设备间的时钟同步。而后，从设备中的核心时钟 Core_clk 需要根据调整后的计数器 master_ms_cnt 值与 master_ts_cnt 同步从设备外围时钟的 rxN_ms_cnt(txN_ms_cnt)与 rxN_ts_cnt(txN_ts_cnt)值。具体方法如图 2-3 所示。

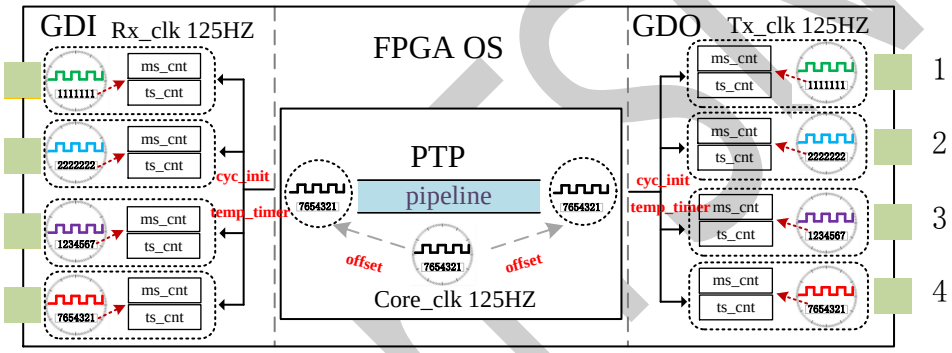


图 2-3 计数器同步示意

当核心时钟 Core_clk 中的计数器 master_ms_cnt 与 master_ts_cnt 根据偏移量 offset 同步时，将两个计数器值{master_ms_cnt, master_ts_cnt }与 offset 的差值存储在 48 位寄存器 temp_cn 中，而后，将 temp_cnt 寄存器的 48 位赋值给{master_ms_cnt, master_ts_cnt}。同时，将 cyc_init 位置 1,并保持三个时钟周期，以确保将 temp_cnt 的 48 位数据有效的赋给{rxN_ms_cnt, rx_ts_cnt}。若同步过程中，offset 的值大于 1ms，则返回同步失败。

2.3 设备间主从时间同步

如图 3-1 所示， Pm 为主时钟，Ps 为从时钟，SW 为交换机作为从时钟的同时用于传递透明时钟。如图 3-1 所示，主从时钟的整体流程如下。

1) 主时钟构造并发送 Sync 报文（携带时间戳$t1$）给从时钟；

2) 从时钟接收到 Sync 报文，获取$t1'$，$t2$与 T1($T_1 = T_{1.1} + T_{1.2} + T_{1.3}$)，并根据$t1 = t1' + T_1$更新 $t1$；
3) 从时钟发送 Delay_req 报文，并获取该报文的发送时间戳$t3$；
4) 主时钟接收到 Delay_req 报文，获取$t4'$与 T2($T_2 = T_{2.1} + T_{2.2} + T_{2.3}$)，并根据$t4 = t4' - T_2$更新 $t4$；
5) 主时钟构造并发送 Delay_resq 报文（携带更新后的$t4$）给从时钟；
6) 从时钟接收到 Delay_resq 报文，获取$t4$；
7) 从时钟根据公式 3 获取 offset，并根据该 offset 值修正从时钟片内所有时钟计数器。

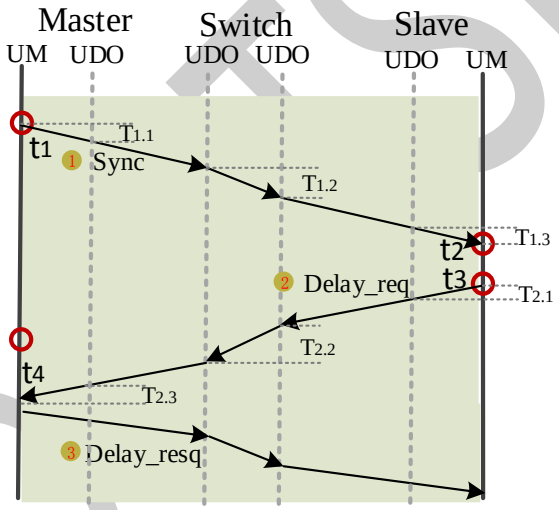


图 3-1 设备间主从时钟同步流程

各设备的工作内容如下：

主时钟（Pm）

- Pm 定时构造并发送 Sync 报文（携带$t1$）给 Ps；
- Pm 解析接收报文，若为 Delay_req 报文，获取$t4'$与 T2($T_2 = T_{2.1} + T_{2.2} + T_{2.3}$)，并根据$t4 = t4' - T_2$更新 $t4$后，给相应的对端返回 Delay_resq 报文(含更新后的$t4$)；
- 若接收的时间同步报文 DMAC 地址不为本设备 MAC 地址的则丢弃；若

报文不为时间同步报文，则转发至下一功能模块。

从时钟（Ps）

- Ps 解析接收报文，若报文为 Sync 报文，则获取$t1'$，$t2$与 T1($T_1 = T_{1.1} + T_{1.2} + T_{1.3}$)，并根据$t1 = t1' + T_1$更新 $t1$。并且，构造 Delay_req 报文返回给 Pm，期间获取 Delay_req 报文的发送时间戳$t3$；
- 若报文为 Pm 返回的 Delay_resq 报文，则从该报文中提取 Delay_req 接收时间戳$t4$，并如公式 3 计算时间差值以进行时间同步；
- 若接收的时间同步报文 DMAC 地址不为本设备 MAC 地址的则丢弃；若报文不为时间同步报文，则转发至下一功能模块。

交换机（SW）

- 交换机实现从时间的角色功能的同时在接收到非本设备的单播时间同步报文、广播时间同步报文以及非时间同步的报文不会丢弃而是转发给下一级模块。对于非本设备时间同步报文交换机则加上经过交换机所花费的时间 T（T=输出时间-输入时间）后转发报文。

附录 A　metadata 格式以及 PTP 同步报文格式

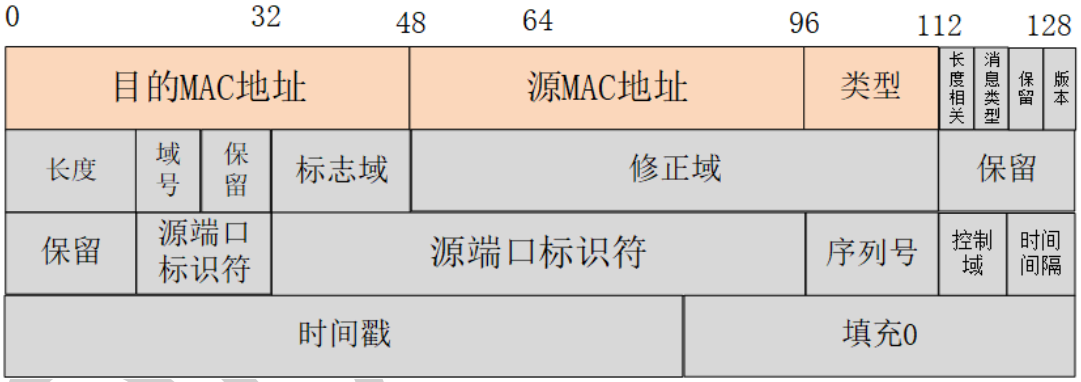
1、Metadata 格式

表 5-1　Metadata 格式

Metadata 0			
[127]	1	pktsrc	分组的来源，0 为网络接口输入，1 为 CPU 输入
[126]	1	pktdst	分组目的，0 为网络接口输出，1 为送 CPU
[125:120]	6	inport	分组的输入端口号
[119:118]	2	outType	00:单播；01：组播；10：泛洪；11：从输入接口输出
[117:112]	6	outPort	单播：分组输出端口 ID，组播/泛洪：组播/泛洪表地址索引
[111:109]	3	priority	分组优先级

[108]	1	discard	丢弃位
[107:96]	12	len	包含 MetaData 字段的分组长度（用于状态管理）
[95:88]	8	smid	最近一次处理分组的模块 ID
[87:80]	8	dmid	下一个处理分组的模块 ID
[79:72]	8	pst	标准协议类型（增加 HTTP GET/POST，SMTP，FTP 定义）
[71:64]	8	seq	分组接收序列号
[63:50]	14	flowid	流 ID/IDS 规则标识，即 ruleID
[49:32]	18	reserve	保留
[31:0]	32	ts	报文时间戳信息

2、Sync，Delay_req，Delay_resq 与 test 报文格式



类型为：16’h88F7；

消息类型：sync 为 4’d1，delay_req 为 4’d3，delay_resq 为 4’d4，delay_test 为 4’d5；

长度为：16’d64 字节；

修正域:透明时钟，起始时，该域为 0；

时间戳：为时间戳（其他无需关系的 PTP 字段填充 0）