



一种适用可控组播的时钟同步系统

孙高睿^{1,2}, 曹争^{1,2}

(1.东南大学计算机科学与工程学院, 南京, 211189; 2.江苏省计算机网络技术重点实验室, 南京, 211189)

摘要: 在可控组播系统中, 网络中的时钟是否同步对 MCS (组播控制服务器) 能否快速准确地发现网络异常, 并对网络进行调整控制起着关键的作用。现有的网络同步协议包括 GPS 同步协议、NTP 同步协议和 PTP 同步协议, 但这些协议并不能很好地满足可控组播系统的要求。本文在基于对 NTP 时钟同步协议改进的基础上设计了一种适用于可控组播的时钟同步系统。以为可控组播提供时钟同步服务。

关键词: 组播; 时钟同步; NTP

A clock synchronization system for Controlled multicast system

Sun Gaorui^{1,2}, Cao Zheng^{1,2}

(1.School of Computer Science and Engineering, Southeast University, Nanjing, 211189;

2.Jiangsu Provincial Key Laboratory of Computer Network Technology;Nanjing,211189)

Abstract: In a controlled multicast system, the clock synchronization plays a key role for the MCS (Multicast Control Server) to find the network anomalies quickly and accurately, and adjust the control network. Existing network synchronization protocol, including GPS synchronization protocol, NTP synchronization protocol and the PTP synchronization protocol, but these protocols can not meet the requirements of controlled multicast system. In this paper, improved the NTP clock synchronization protocols and design of a controllable multicast for clock synchronization system.It provides clock synchronization services for controllable multicast.

Key words: Multicast; Clock synchronization;NTP

组播是一种高效的通信模式, 可以以较小的传输开销提供一点到多点和多点到多点通信服务, 在下一代 IP 网络中有着巨大的发展潜力。在可控组播服务中, 需要对组成员进行管理, 并为组播用户提供可靠地服务质量保证, 因此要求组播系统能够实时地获取组成员在网络中的分布及其获取的组播服务质量, 并且在用户获取的服务质量不能满足要求的时候, 能够提供一定的调整和控制机制。这就要求在同一个组播组中的测量点成员能够处于相对一致的时间, 以使测量模块能够较为精确地反映出网络的真实质量状态。目前已有的同步协议有 GPS 同步协议、NTP 同步协

议和 PTP 同步协议。其中 GPS 同步协议能够提供高精度的时间同步支持, 但其实施复杂, 可扩展性弱, 费用高昂, 不适用于组播环境; NTP 协议虽然易部署, 可扩展性强, 但其基于网络中数据包往返路径一致的假设, 未能充分考虑到网络结构的复杂性, 只能提供毫秒级别的时钟同步效果, 在可控组播系统的实际测量中, 通过 NTP 进行同步所测得的单向时延值中甚至出现时间值为负值得情况; PTP 协议基于同 NTP 同样的数据包往返路径一致的假设, 此外 PTP 协议在传输路径中引入了透明时钟的概念, 以收集额外的信息作为时钟同步的修正信息, 提高同步精度。但在目前大多数网络中间设备都不支持 PTP 协议的情况下, 该协议同样不适用于可控组播系统。因此为了满足可控组播时钟同步的需求, 需要对已有的协议进行扩充的基础上设计出一个时钟同步系统, 为可控组播系统提供时钟同步服务。

基金项目: 国家科技支撑计划项目(2008BAH37B04)

作者简介: 孙高睿, (1987-), 男, 硕士研究生, E-mail: grsun@njnet.edu.cn; 曹争, (1958-), 男, 副教授, E-mail: zcao@njnet.edu.cn



一个时钟同步系统主要包括三个方面:1) 同步模型; 2) 同步方法; 3) 容错性考虑。下面将对这三个方面进行讨论。

1 同步模型

本文提出的时间同步系统的同步模型在 NTP 协议同步模型基础上进行了修改。采用层级式的同步模型。

模型中具有三种节点: 1) 位于时间同步顶层的主服务器, 系统中所有的实体的时间都将同步到顶点主服务器, 即在系统中做为时钟源存在; 2) 顶层以下的次级服务器, 次级服务器从其它服务器获得时间消息进行同步, 同时也为其它服务器提供时间同步服务; 3) 客户端, 将本地时间同步到服务器, 并不为其它实体提供时间同步服务。

具体模型如图 1 所示。所有实体同步到统一的时间基准。这个基准由处于层级最顶层的时钟提供。所有底层的实体在一定范围内选择位于上层的或与自身位于同一层的合适的服务器进行同步, 但不同步到下层时钟。

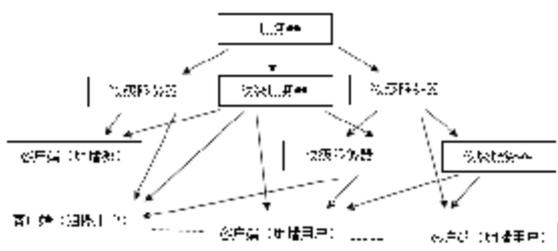


图 1

为了能够发现合适的时钟服务器采用如下方式来发现选择同步时钟。用户发送数据包到指定的组播地址。并通过在包头中限定包的跳数 (TTL) 以选择一定范围内的任播服务器。如果一个时钟同步服务器侦听到用户发送的数据包, 则使用用户的单播地址返回一个数据包。用户在收到数据包后启动对服务器的临时关联。

在发现时间同步服务器的过程中, 用户可能会与多个时间同步服务器建立关联。在种情况时, 需要对服务器进行筛选, 因为这些服务器中可能存在时间错误的服务器。因此用户必须从中选择一个合适的同步服务器集作为可用的时间同步服务使用。选择的方法基于以下假设, 即大多数的时间同步服务器提供的时间处于准确的状态,

因此要做的是对计算出的服务器时间在一定的阈值范围内聚类, 剔除时间不落在阈值范围内的服务器提供的时间。

在从多个服务器时钟中提取出合适的时钟之后为了考虑在每一次时钟同步时测得的时间的稳定性和正确性, 还希望对选择出的时钟进行进一步筛选。方法大致是计算服务器到本地距离、到顶点主服务器距离及服务器所在层级加权和, 选择和小于一定阈值的服务器。这样做是基于层级越高的服务器能够提供越接近主服务器时间的数据及数据包在路径中传输的路径越短所受到的路径不确定性因素影响几率就越小的假设。

如果在每一次时间同步过程中用户都进行上述步骤的选择将极大的浪费系统资源, 因此希望用户能够只在需要时进行上述的服务器选择工作, 其它时候则根据保留的与服务器的关联进行同步。同时为了防止服务器或用户在某些情况下离线 (例如线路故障、服务器本身运行错误、用户下线等) 则在用户和服务器之间加入心跳机制以确保可用性是合理的。

2 同步方法

系统基于数据包在网络路径中往返路径一致的假设进行时钟同步。采用交换网络实体间的时间信息进行同步。系统与 NTP 协议使用相同的数学模型。大致步骤如图 2 所示。

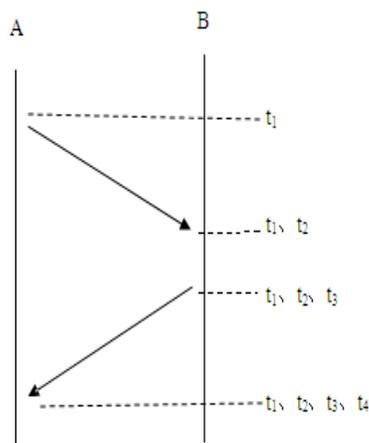


图 2

时间同步过程由服务请求方 A 发起。在图中, 第一个数据包由 A 传输, 并将数据包发送时间信



息 t_1 加入到数据包中。B 在 t_2 时刻收到，将 t_1 此时或一段时间之后在 t_3 时刻，B 发送包含 t_2 、 t_3 信息的数据包给 A。A 在 t_4 时刻收到 B 发出的数据包，根据公式一和二计算 B 关于 A 的时差和时延。

公式一：

$$\Theta = \frac{(t_2 - t_1) + (t_3 - t_4)}{2}$$

公式二：

$$\Delta = (t_4 - t_1) - (t_3 - t_2)$$

其中 Θ 为 A、B 之间的时间偏差， Δ 为数据包往返 AB 所花费的时间。

系统将使用两种方式对同步时间进行修正，以希望达到更高同步精度的目的。

第一种方式通过使用多个服务器提供冗余时间信息以进行修正。采用这种修正方式时用户将不再使用得到的消息计算一个准确的时间，而是计算出一个时间的范围，正确时间处在这个范围内的概率满足一定的阈值。将多个服务器计算出的范围进行重叠选出重叠最多的范围，最后将这个范围规约成一个时间点做为准确的时间。通过不同服务器信息计算时间进行同步基于的假设是基于每个服务器信息计算出的范围都在很大程度上包含了准确的时间。那么只有当准确时间以很大概率落于重叠范围最大部分时上述假设才能成立^[2]。

第二种方式希望通过发送冗余的包含时间信息的数据包以得到额外的信息，来对同步时间进行修正。具体方式如图 3 所示。

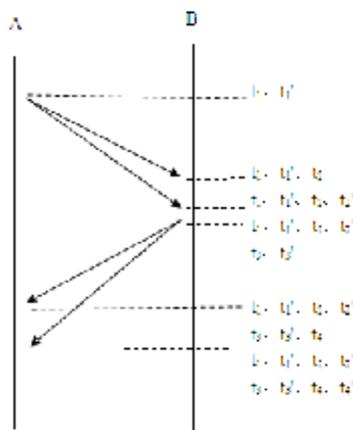


图 3

如图所示用户 A 和服务器 B 在交换信息时，发送了额外的数据包。A 在 t_1 时刻发送包含时间信息的数据包 a，此外在发送完 a 之后立刻发送一个背靠背的数据包 a'，a' 记录 a 的发送时间 t_1' ，并且要求与数据包 a 的长度不同。由于网络设备在转发数据包时的处理速度等于数据包长度除以吞吐率，每个网络设备的吞吐率相同，因此数据包 a 与 a' 在传输过程中将会错开，即以不同的时间到达 B，B 记录收到两个数据包的时间 t_2 和 t_2' 。B 按公式三计算两个数据包在网络中传输时间的差值 δ_1 ，在一段时间之后 B 在 t_3 和 t_3' 时刻按与 A 相同的方式发送数据包 b 和 b' 给 A，同时将 δ_1 也发送给 A。其中数据包 b 的长度等于数据包 a 的长度，数据包 b' 的长度等于数据包 a' 的长度。A 在 t_4 和 t_4' 时刻分别收到数据包 b 和数据包 b'。同样的，A 按公式 2 计算两个数据包在网络中传输时间的差值 δ_2 。对比 δ_1 和 δ_2 ，如果两个值不一样则说明在第一对数据包传输时的路径与第二对数据包传输时的路径必然存在不对称。对比这种不对称，对同步信息进行修正。

公式三：

$$d_1 = (t_2' - t_1') - (t_2 - t_1)$$

$$d_2 = (t_4' - t_3') - (t_4 - t_3)$$

在时间同步完成之后，由于不同的实体具有不同的时钟频率，在时钟运行过一段时间之后，原本同步的时间将变得不同步，实体间需要再次进行时间同步，这时将再次发起同步过程。

3 容错性考虑

由于 IP 网络采用尽可能传输的方式传输数据，包含时间信息的数据包可能发生丢失、重复报文、错误报文（这里指用户或服务器认为已经丢失的报文又忽然出现的情况）、失序等情况。为了使系统更加的健壮必须对这些问题进行考虑。

对于报文丢失的情况不会使系统产生错误。考虑如下情况，假设 A 进行时间同步的数据包或对同步进行辅助的数据包丢失，则 B 收不到完整的请求信息，因此不对 A 的请求作出回应，A 等待超时发出新的请求；同样的，B 发送给 A 的数



据包丢失, 则 A 在时限内收不到完整的时间同步消息, A 认为请求超时, 发起新的同步请求。因此系统只需丢弃该次同步的结果即可, 无需对丢失报文进行重传。

对于时间同步数据包与辅助信息数据包发生失序的情况两端的系统只要检查数据包即可作出判断。

对于重复报文、错误报文的情况可以在同步时使用变量存储时间值以实现鉴别错误的目的。如图 4 所示。

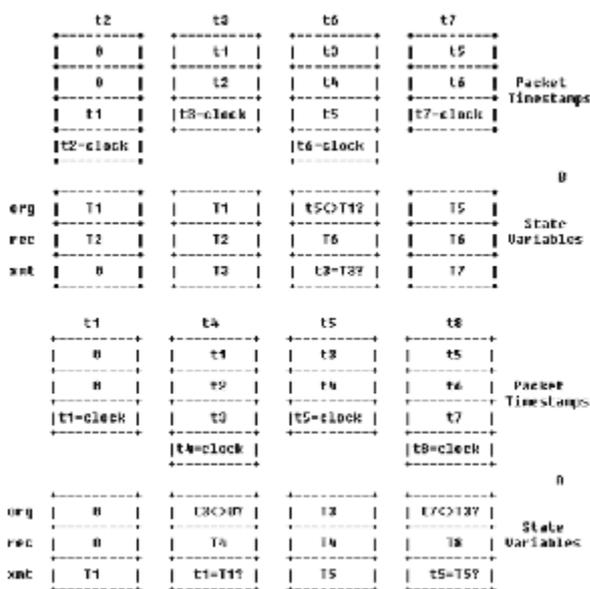


图 4

如图所示对于成对的数据包(指一对包含往返时间信息的数据包)使用 3 个状态变量 org、rec 和 xmt 分别记录 t_1 、 t_2 、 t_3 的值。状态变量将在数据包一到或一离开时被赋值。在 xmt 和 org 状态变量更新前, 进行下述检验。如果数据包中的 t_3 与 org 中的 T_3 相同则说明报文为重复报文, 如果报文中 t_1 与 xmt 总的 T_1 则说明报文是错误报文 [2]。

4 结束语

本时钟同步系统在基于 NTP 时间协议的基础上, 对 NTP 协议进行了扩充。为了在不影响可扩展性的前提下达到更高的时间精度, 系统提供了冗余信息以对同步的时间进行修正。为可控组播组的时间同步提供了更合理的方案。

参考文献

[1]IEEE1588-2008,IEEE Standard for a Precision Clock Synchronization Protocol For Networked Measurement and Control System[S].
 [2]RFC 5905,Network Time Protocol Version 4: Protocol and Algorithms Specification[S].
 [3]RFC 2330,Framework for IP Performance Metrics[S].
 [4]李明国, 宋海娜.计算机时钟同步技术研究[J]系统仿真学报,第 14 卷第 4 期:1004-731x(2002)04-0477-04