



(12) 发明专利

(10) 授权公告号 CN 110875953 B

(45) 授权公告日 2023. 03. 24

(21) 申请号 201811028228.6

H04W 28/08 (2023.01)

(22) 申请日 2018.09.04

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 101437015 A, 2009.05.20

申请公布号 CN 110875953 A

CN 101114969 A, 2008.01.30

CN 101340633 A, 2009.01.07

(43) 申请公布日 2020.03.10

审查员 王星

(73) 专利权人 中兴通讯股份有限公司

地址 518000 广东省深圳市南山区高新技

术产业园科技南路中兴通讯大厦

(72) 发明人 易飞

(74) 专利代理机构 北京市金杜律师事务所

11256

专利代理师 王茂华

(51) Int. Cl.

H04L 67/1008 (2022.01)

H04L 67/61 (2022.01)

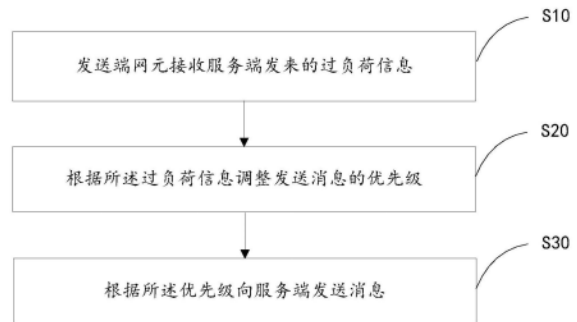
权利要求书1页 说明书7页 附图4页

(54) 发明名称

一种过负荷控制方法、装置、设备及可读存储介质

(57) 摘要

本文公开了一种过负荷控制方法、装置、设备及可读存储介质,属于通信技术领域,该方法包括:发送端网元接收服务端发来的过负荷信息;根据所述过负荷信息调整发送消息的优先级;根据所述优先级向服务端发送消息;通过在业务消息中增加网元负荷相关的参数,灵活的进行过负荷控制,并通过接收端和发送端的交互,实时调整网络中的消息优先级,通过两边协商的方式保证业务的最大可靠性。



1. 一种过负荷控制方法,包括:
发送端网元接收服务端发来的过负荷信息;
根据所述过负荷信息调整待发送消息的优先级,以使得所述待发送消息的优先级调整为高于所述服务端能处理的消息优先级;
根据所述优先级向服务端发送消息。
2. 根据权利要求1所述的一种过负荷控制方法,其特征在于,所述过负荷信息包括:过负荷等级、所述能处理的消息优先级、消息丢弃比率和过负荷持续时间。
3. 根据权利要求2所述的一种过负荷控制方法,其特征在于,所述过负荷等级根据所述服务端的CPU占用率和内存占用率来获取。
4. 根据权利要求2所述的一种过负荷控制方法,其特征在于,所述根据所述过负荷信息调整待发送消息的优先级之后还包括:
根据所述消息丢弃比率调整所述过负荷持续时间内发送的消息数量。
5. 根据权利要求4所述的一种过负荷控制方法,其特征在于,所述根据所述优先级向服务端发送消息具体为:
在所述过负荷持续时间内,向所述服务端发送的消息数量为调整后的消息数量;所述消息的优先级高于所述服务端能处理的消息优先级。
6. 根据权利要求1所述的一种过负荷控制方法,其特征在于,所述根据所述过负荷信息调整待发送消息的优先级还包括:
预设用户的消息优先级高于默认优先级;
预设时间段的消息优先级高于默认优先级;
预设业务类型的消息优先级高于默认优先级。
7. 一种过负荷控制装置,其特征在于,包括:
接收模块,用于发送端网元接收服务端发来的过负荷信息;
调整模块,用于根据所述过负荷信息调整待发送消息的优先级,以使得所述待发送消息的优先级调整为高于所述服务端能处理的消息优先级;
控制模块,用于根据所述优先级向服务端发送消息。
8. 一种电子设备,包括存储器、处理器和至少一个被存储在所述存储器中并被配置为由所述处理器执行的应用程序,其特征在于,所述应用程序被配置为用于执行权利要求1-6任一项所述的过负荷控制方法。
9. 一种可读存储介质,其特征在于,其上存储有计算机程序,该程序被处理器执行时实现如权利要求1-6任一所述的过负荷控制方法。

一种过负荷控制方法、装置、设备及可读存储介质

技术领域

[0001] 本文涉及通信技术领域,尤其涉及一种过负荷控制方法、装置、设备及可读存储介质。

背景技术

[0002] 通信技术的不断发展,网元之间的消息种类、消息量都有了巨大的增加,在发展到5G以后,协议引入了虚拟化和服务化接口。由于消息量的增加,网元间过负荷时有发生,传统的2G、3G、4G网元过负荷控制手段是在接收端根据自定义的消息优先级进行丢弃,此种控制方案逻辑简单,但是对于发送端不够友好,不同接收端定义的消息优先级差别会导致不同的消息丢弃场景。根据3GPP的要求,在同等负荷条件下,HLR处理移动性管理(Mobility Management)的业务优先级高于短消息(Short Message Service)相关的业务。在同等条件下,无论哪些网元发送的短信消息处理优先级都会小于移动性管理的相关业务。

[0003] 随着移动网络的发展,特别是移动物联网业务的兴起,包括5G网络切片技术的引入,业务网元的功能越来越单一。比如提供物联网业务功能的网元可能只需要提供短信业务。在此应用场景下,同样的短信业务,物联网切片中的优先级要高于人网中的优先级。传统的过负荷控制方案无法区分此种业务逻辑。

发明内容

[0004] 本文在于提供一种过负荷控制方法、装置、设备及可读存储介质,通过在业务消息中增加网元负荷相关的参数,灵活的进行过负荷控制,并通过接收端和发送端的交互,实时调整网络中的消息优先级,通过两边协商的方式保证业务的最大可靠性。

[0005] 本文解决上述技术问题所采用的技术方案如下:

[0006] 根据本文的一个方面,提供的一种过负荷控制方法,包括:

[0007] 发送端网元接收服务端发来的过负荷信息;

[0008] 根据所述过负荷信息调整发送消息的优先级;

[0009] 根据所述优先级向服务端发送消息。

[0010] 可选地,所述过负荷信息包括:过负荷等级、能处理的消息优先级、消息丢弃比率和过负荷持续时间。

[0011] 可选地,所述过负荷等级根据所述服务端的CPU占用率和内存占用率来获取。

[0012] 可选地,所述根据所述过负荷信息调整发送消息的优先级具体为:

[0013] 调整发送消息的优先级高于所述服务端能处理的消息优先级。

[0014] 可选地,所述根据所述过负荷信息调整发送消息的优先级之后还包括:

[0015] 根据所述消息丢弃比率调整所述过负荷持续时间内发送的消息数量。

[0016] 可选地,所述根据所述优先级向服务端发送消息具体为:

[0017] 在所述过负荷持续时间内,向所述服务端发送的消息数量为调整后的消息数量;所述消息的优先级高于所述服务端能处理的消息优先级。

- [0018] 可选地,所述根据所述过负荷信息调整发送消息的优先级还包括:
- [0019] 预设用户的消息优先级高于默认优先级;
- [0020] 预设时间段的消息优先级高于默认优先级;
- [0021] 预设业务类型的消息优先级高于默认优先级。
- [0022] 根据本文的另一个方面,提供的一种过负荷控制装置,包括:
- [0023] 接收模块,用于发送端网元接收服务端发来的过负荷信息;
- [0024] 调整模块,用于根据所述过负荷信息调整发送消息的优先级;
- [0025] 控制模块,用于根据所述优先级向服务端发送消息。
- [0026] 根据本文的再一个方面,提供的一种电子设备,包括存储器、处理器和至少一个被存储在所述存储器中并被配置为由所述处理器执行的应用程序,所述应用程序被配置为用于执行以上所述的过负荷控制方法。
- [0027] 根据本文的再一个方面,提供的一种可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现以上所述的过负荷控制方法。
- [0028] 本发明实施例的一种过负荷控制方法、装置、设备及可读存储介质,该方法包括:发送端网元接收服务端发来的过负荷信息;根据所述过负荷信息调整发送消息的优先级;根据所述优先级向服务端发送消息;通过在业务消息中增加网元负荷相关的参数,灵活的进行过负荷控制,并通过接收端和发送端的交互,实时调整网络中的消息优先级,通过两边协商的方式保证业务的最大可靠性。

附图说明

- [0029] 图1为本发明实施例一提供的一种过负荷控制方法流程图;
- [0030] 图2为本发明实施例一提供的另一种过负荷控制方法流程图;
- [0031] 图3为本发明实施例一提供的一种过负荷控制逻辑图;
- [0032] 图4为本发明实施例一提供的另一种过负荷控制逻辑图;
- [0033] 图5为本发明实施例二提供的一种过负荷控制装置的示范性结构框图。
- [0034] 本文目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0035] 为了使本文所要解决的技术问题、技术方案及有益效果更加清楚、明白,以下结合附图和实施例,对本文进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本文,并不用于限定本文。

[0036] 实施例一

[0037] 如图1所示,在本实施例中,一种过负荷控制方法,包括:

- [0038] S10、发送端网元接收服务端发来的过负荷信息;
- [0039] S20、根据所述过负荷信息调整发送消息的优先级;
- [0040] S30、根据所述优先级向服务端发送消息。

[0041] 在本实施例中,通过在业务消息中增加网元负荷相关的参数,灵活的进行过负荷控制,并通过接收端和发送端的交互,实时调整网络中的消息优先级,通过两边协商的方式保证业务的最大可靠性。

[0042] 在本实施例中,所述过负荷信息包括:过负荷等级、能处理的消息优先级、消息丢弃比率和过负荷持续时间。

[0043] 其中,过负荷等级用于指示当时系统的过负荷状态;能处理的消息优先级用于指示当时能够保证成功处理的消息优先级,发送端的网元可以调整消息的优先级;消息丢弃比率用于指示当前的消息丢弃率,据此,发送端的网元可以调整消息数目;过负荷持续时间用于指示当时过负荷可能的周期,发送端的网元可以判断过负荷的时间,用于后续恢复正常业务。

[0044] 在本实施例中,所述过负荷等级根据所述服务端的CPU占用率和内存占用率来获取。

[0045] 在网元进行消息交互时,处理业务消息的网元定时检测当前的cpu占用率,内存占用率,通过计算函数func(cpu,mem)计算判断自身发生过负荷,并计算得到具体的负荷等级;计算逻辑参考下表1:

[0046]

CPU占用率	内存占用率	负荷等级
<50%	<40%	0
50%~60%	40%~50%	1
60%~70%	50%~60%	2
70%~80%	60%~70%	3
80%~90%	70%~80%	4
90%~100%	80%~90%	5

[0047] 表1

[0048] 基于表1中的负荷等级,服务端设置的一种自身过负荷处理机制如下表2所示,其中,消息丢弃比率按默认优先级设置,对于更低优先级的消息,丢弃比率= $\min(\text{默认丢弃率} \times (\text{默认优先级}/\text{实际优先级}), 100)$;例如:过负荷等级4时,优先级40的消息丢弃比率是50%。优先级20的消息丢弃率是100%。

[0049]

负荷等级	消息优先级	消息丢弃比率
0	10	0%
1	50	10%
2	60	15%
3	70	20%
4	80	25%
5	90	30%

[0050] 表2

[0051] 在本实施例中,所述步骤S20具体为:

[0052] 调整发送消息的优先级高于所述服务端能处理的消息优先级。

[0053] 如图2所示,在本实施例中,所述步骤S20之后还包括:

[0054] S21、根据所述消息丢弃比率调整所述过负荷持续时间内发送的消息数量。

[0055] 在本实施例中,所述步骤S30具体为:

[0056] 在所述过负荷持续时间内,向所述服务端发送的消息数量为调整后的消息数量;所述消息的优先级高于所述服务端能处理的消息优先级。

[0057] 在本实施例中,以图3所示的一个发送端网元NE_C(Net Equipment Client)向服务端NE_S(Net Equipment Server)发送消息为例,来说明本实施例中的过负荷控制方法:

[0058] NE_C发送业务请求,发送了100条消息;

[0059] NE_S处于过负荷状态,只能处理80条消息,返回相应,携带80条成功处理,20条错误处理。同时,携带过负荷相关参数:当前系统过负荷等级1,消息的丢弃比率为20%,能处理的消息优先级为60,期望的过负荷时长20s。

[0060] NE_C收到响应以后,根据过负荷参数调整自己的发送参数,减少单位周期的消息发送数量,携带消息的优先级。

[0061] NE_S收到高优先级指示的消息,正常处理业务,由于接收的消息减少,过负荷状态变为非过负荷状态。

[0062] NE_C在超过过负荷周期的时间以后,重新正常携带业务请求消息,由于NE_S恢复正常状态,能够正常处理所有业务消息。

[0063] 作为另一种实施例,以图4所示的多个发送端网元NE_C1、NE_C2向服务端NE_S发送消息为例,来说明本实施例中的过负荷控制方法:

[0064] NE_C1发送业务请求,发送了100条消息,消息优先级为50;

[0065] NE_C2发送业务请求,发送了100条消息,消息优先级为40;

[0066] NE_S处于过负荷状态,只能处理部分消息,返回响应,

[0067] 对于NE_C1携带80条成功处理,20条错误处理。同时,携带过负荷相关参数:当前系统过负荷等级1,消息的丢弃比率为20%,能处理的消息优先级为60,期望的过负荷时长20s。

[0068] 对于NE_C2携带40条成功处理,60条错误处理。同时,携带过负荷相关参数:当前系统过负荷等级1,消息的丢弃比率为20%,能处理的消息优先级为60,期望的过负荷时长20s。

[0069] NE_C1收到响应以后,根据过负荷参数调整自己的发送参数,减少单位周期的消息发送数量,携带消息的优先级。单位周期内发送80条消息,消息优先级调整为80;NE_S正常处理80条消息,返回全部成功。

[0070] NE_C2收到响应以后,根据过负荷参数调整自己的发送参数,减少单位周期的消息发送数量,携带消息的优先级。单位周期内发送40条消息,消息优先级调整为100;NE_S正常处理40条消息,返回全部成功。

[0071] NE_S由于接收的消息数目减少,从过负荷状态变为非负荷状态。

[0072] NE_C1在过负荷时间周期以后正常发送消息,消息数100,优先级50,NE_S正常处理业务消息。

[0073] NE_C2在过负荷时间周期以后正常发送消息,消息数100,优先级50,NE_S正常处理业务消息。

[0074] 在本实施例中,所述步骤S20还包括:

[0075] 预设用户的消息优先级高于默认优先级;

[0076] 预设时间段的消息优先级高于默认优先级;

[0077] 预设业务类型的消息优先级高于默认优先级。

[0078] 例如:维护人员可以人为设置5G网络中不同切片的优先级,来保证不同切片的业

务优先级。比如,某一个网络切片只在晚上搜集业务信息(上报每天的工业数据),则可以在晚上将这个切片的优先级调高,保证高优先级,其他时间默认优先级。

[0079] 通过服务端和客户端过负荷设置和优先级交互,实时调整网络中的消息优先级,通过两边协商的方式保证业务的最大可靠性。本发明提供了一种更加灵活、更加智能的网元间过负荷控制方案,可以基于消息交互实时调整业务消息优先级,提高网元间业务消息交互的可靠性。

[0080] 实施例二

[0081] 如图5所示,在本实施例中,一种过负荷控制装置,包括:

[0082] 接收模块10,用于发送端网元接收服务端发来的过负荷信息;

[0083] 调整模块20,用于根据所述过负荷信息调整发送消息的优先级;

[0084] 控制模块30,用于根据所述优先级向服务端发送消息。

[0085] 在本实施例中,通过在业务消息中增加网元负荷相关的参数,灵活的进行过负荷控制,并通过接收端和发送端的交互,实时调整网络中的消息优先级,通过两边协商的方式保证业务的最大可靠性。

[0086] 在本实施例中,所述过负荷信息包括:过负荷等级、能处理的消息优先级、消息丢弃比率和过负荷持续时间。

[0087] 其中,过负荷等级用于指示当时系统的过负荷状态;能处理的消息优先级用于指示当时能够保证成功处理的消息优先级,发送端的网元可以调整消息的优先级;消息丢弃比率用于指示当前的消息丢弃率,据此,发送端的网元可以调整消息数目;过负荷持续时间用于指示当时过负荷可能的周期,发送端的网元可以判断过负荷的时间,用于后续恢复正常业务。

[0088] 在本实施例中,所述过负荷等级根据所述服务端的CPU占用率和内存占用率来获取。

[0089] 在网元进行消息交互时,处理业务消息的网元定时检测当前的cpu占用率,内存占用率,通过计算函数 $\text{func}(\text{cpu}, \text{mem})$ 计算判断自身发生过负荷,并计算得到具体的负荷等级;计算逻辑参考下表1:

[0090]

CPU占用率	内存占用率	负荷等级
<50%	<40%	0
50%~60%	40%~50%	1
60%~70%	50%~60%	2
70%~80%	60%~70%	3
80%~90%	70%~80%	4
90%~100%	80%~90%	5

[0091] 表1

[0092] 基于表1中的负荷等级,服务端设置的一种自身过负荷处理机制如下表2所示,其中,消息丢弃比率按默认优先级设置,对于更低优先级的消息,丢弃比率= $\text{min}(\text{默认丢弃率} \times (\text{默认优先级}/\text{实际优先级}), 100)$;例如:过负荷等级4时,优先级40的消息丢弃比率是50%。优先级20的消息丢弃率是100%。

[0093]

负荷等级	消息优先级	消息丢弃比率
------	-------	--------

0	10	0%
1	50	10%
2	60	15%
3	70	20%
4	80	25%
5	90	30%

[0094] 表2

[0095] 在本实施例中,以图3所示的一个发送端网元NE_C(Net Equipment Client)向服务端NE_S(Net Equipment Server)发送消息为例,来说明本实施例中的过负荷控制方法:

[0096] NE_C发送业务请求,发送了100条消息;

[0097] NE_S处于过负荷状态,只能处理80条消息,返回相应,携带80条成功处理,20条错误处理。同时,携带过负荷相关参数:当前系统过负荷等级1,消息的丢弃比率为20%,能处理的消息优先级为60,期望的过负荷时长20s。

[0098] NE_C收到响应以后,根据过负荷参数调整自己的发送参数,减少单位周期的消息发送数量,携带消息的优先级。

[0099] NE_S收到高优先级指示的消息,正常处理业务,由于接收的消息减少,过负荷状态变为非过负荷状态。

[0100] NE_C在超过过负荷周期的时间以后,重新正常携带业务请求消息,由于NE_S恢复正常状态,能够正常处理所有业务消息。

[0101] 作为另一种实施例,以图4所示的多个发送端网元NE_C1、NE_C2向服务端NE_S发送消息为例,来说明本实施例中的过负荷控制方法:

[0102] NE_C1发送业务请求,发送了100条消息,消息优先级为50;

[0103] NE_C2发送业务请求,发送了100条消息,消息优先级为40;

[0104] NE_S处于过负荷状态,只能处理部分消息,返回响应,

[0105] 对于NE_C1携带80条成功处理,20条错误处理。同时,携带过负荷相关参数:当前系统过负荷等级1,消息的丢弃比率为20%,能处理的消息优先级为60,期望的过负荷时长20s。

[0106] 对于NE_C2携带40条成功处理,60条错误处理。同时,携带过负荷相关参数:当前系统过负荷等级1,消息的丢弃比率为20%,能处理的消息优先级为60,期望的过负荷时长20s。

[0107] NE_C1收到响应以后,根据过负荷参数调整自己的发送参数,减少单位周期的消息发送数量,携带消息的优先级。单位周期内发送80条消息,消息优先级调整为80;NE_S正常处理80条消息,返回全部成功。

[0108] NE_C2收到响应以后,根据过负荷参数调整自己的发送参数,减少单位周期的消息发送数量,携带消息的优先级。单位周期内发送40条消息,消息优先级调整为100;NE_S正常处理40条消息,返回全部成功。

[0109] NE_S由于接收的消息数目减少,从过负荷状态变为非负荷状态。

[0110] NE_C1在过负荷时间周期以后正常发送消息,消息数100,优先级50,NE_S正常处理业务消息。

[0111] NE_C2在过负荷时间周期以后正常发送消息,消息数100,优先级50,NE_S正常处理业务消息。

[0112] 实施例三

[0113] 在本实施例中,一种电子设备,包括存储器、处理器和至少一个被存储在所述存储器中并被配置为由所述处理器执行的应用程序,所述应用程序被配置为用于执行实施例一所述的过负荷控制方法。

[0114] 实施例四

[0115] 本发明实施例提供一种可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现如上述过负荷控制方法实施例中任一所述的方法实施例。

[0116] 需要说明的是,上述装置、设备实和可读存储介质实施例与方法实施例属于同一构思,其具体实现过程详见方法实施例,且方法实施例中的技术特征在装置实施例中均对应适用,这里不再赘述。

[0117] 本发明实施例的一种过负荷控制方法、装置、设备及可读存储介质,该方法包括:发送端网元接收服务端发来的过负荷信息;根据所述过负荷信息调整发送消息的优先级;根据所述优先级向服务端发送消息;通过在业务消息中增加网元负荷相关的参数,灵活的进行过负荷控制,并通过接收端和发送端的交互,实时调整网络中的消息优先级,通过两边协商的方式保证业务的最大可靠性。

[0118] 本领域普通技术人员可以理解,上文中所公开方法中的全部或某些步骤、系统、装置中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。

[0119] 在硬件实施方式中,在以上描述中提及的功能模块/单元之间的划分不一定对应于物理组件的划分;例如,一个物理组件可以具有多个功能,或者一个功能或步骤可以由若干物理组件合作执行。某些物理组件或所有物理组件可以被实施为由处理器,如中央处理器、数字信号处理器或微处理器执行的软件,或者被实施为硬件,或者被实施为集成电路,如专用集成电路。这样的软件可以分布在计算机可读介质上,计算机可读介质可以包括计算机存储介质(或非暂时性介质)和通信介质(或暂时性介质)。如本领域普通技术人员公知的,术语计算机存储介质包括在用于存储信息(诸如计算机可读指令、数据结构、程序模块或其他数据)的任何方法或技术中实施的易失性和非易失性、可移除和不可移除介质。计算机存储介质包括但不限于RAM、ROM、EEPROM、闪存或其他存储器技术、CD-ROM、数字多功能盘(DVD)或其他光盘存储、磁盒、磁带、磁盘存储或其他磁存储装置、或者可以用于存储期望的信息并且可以被计算机访问的任何其他的介质。此外,本领域普通技术人员公知的是,通信介质通常包含计算机可读指令、数据结构、程序模块或者诸如载波或其他传输机制之类的调制数据信号中的其他数据,并且可包括任何信息递送介质。

[0120] 以上参照附图说明了本发明的优选实施例,并非因此局限本发明的权利范围。本领域技术人员不脱离本发明的范围和实质内所作的任何修改、等同替换和改进,均应在本发明的权利范围之内。

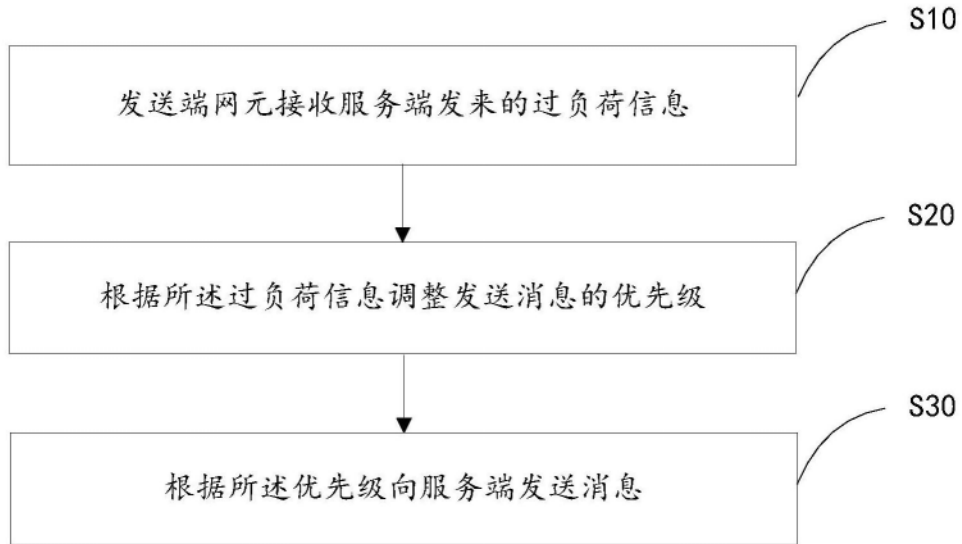


图1

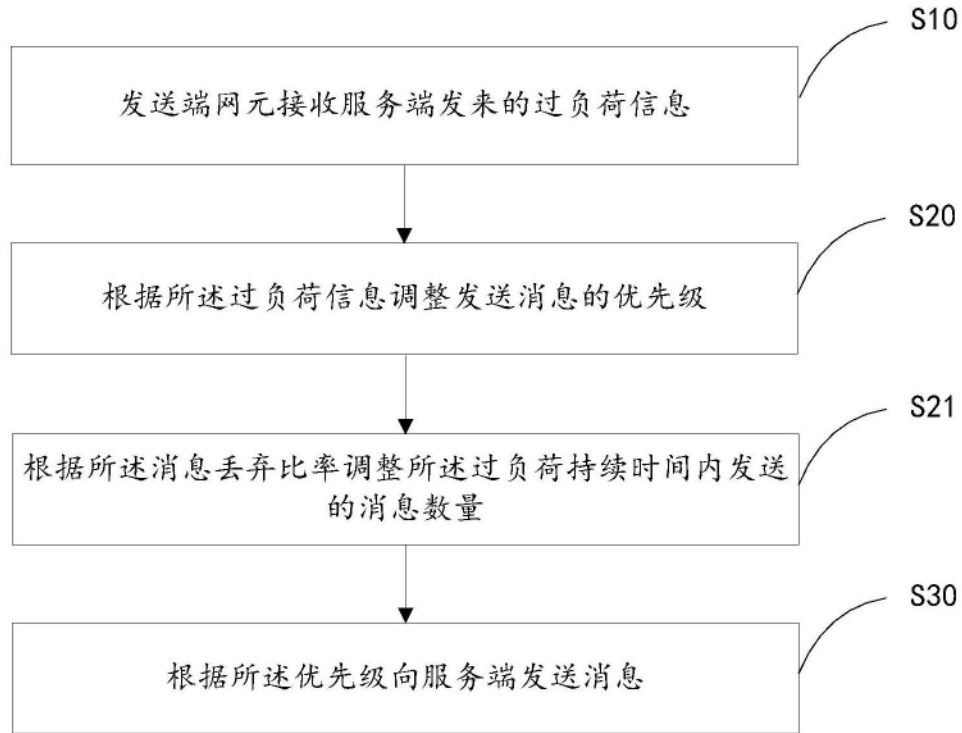


图2

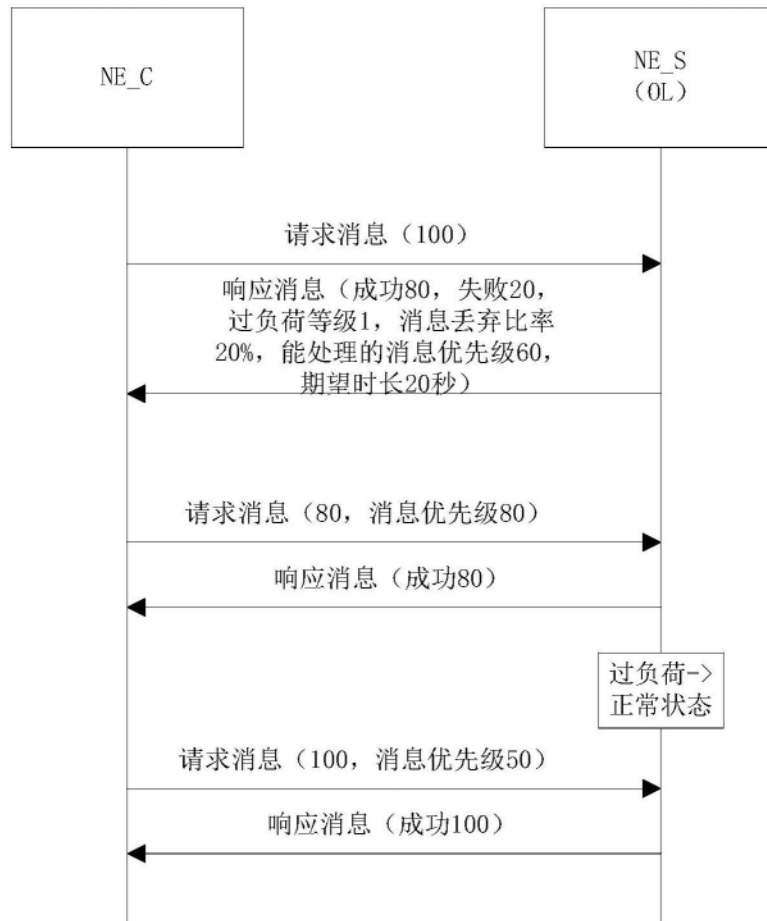


图3

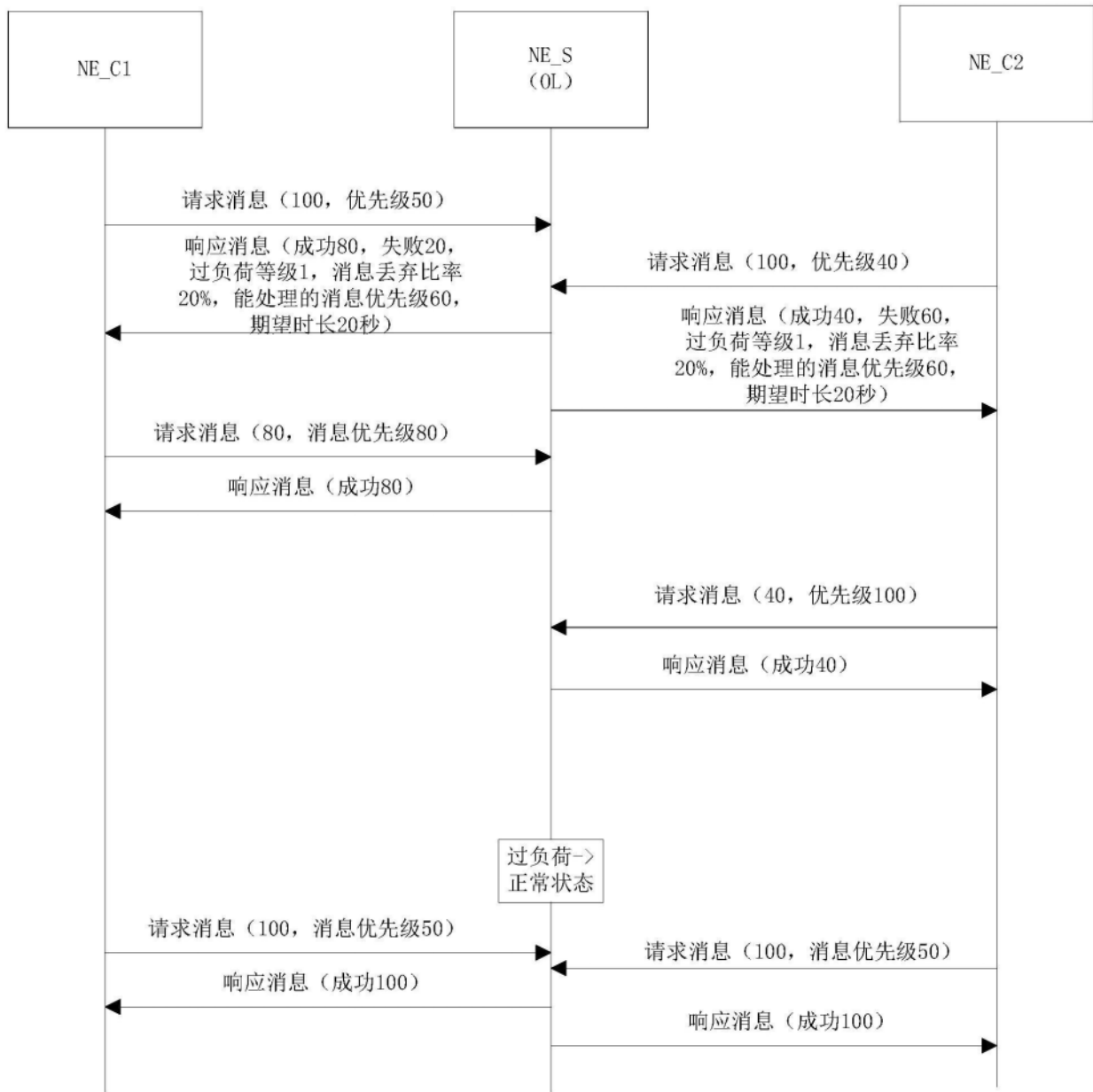


图4

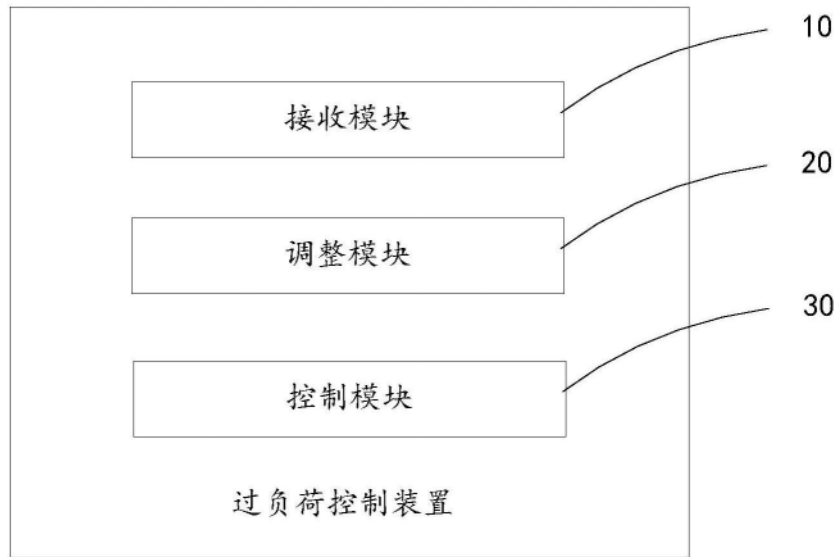


图5