



(12) 发明专利

(10) 授权公告号 CN 114885045 B

(45) 授权公告日 2022.10.04

(21) 申请号 202210791427.2

H04L 47/50 (2022.01)

(22) 申请日 2022.07.07

G06F 9/50 (2006.01)

(65) 同一申请的已公布的文献号

G06F 13/28 (2006.01)

申请公布号 CN 114885045 A

G06F 13/42 (2006.01)

(43) 申请公布日 2022.08.09

(56) 对比文件

CN 113810791 A, 2021.12.17

(73) 专利权人 浙江锐文科技有限公司

审查员 郝凯利

地址 311121 浙江省杭州市余杭区文一西

路1001号衢州海创园D座6楼

(72) 发明人 阎燕 唐青松

(74) 专利代理机构 北京志霖恒远知识产权代理

有限公司 11435

专利代理师 戴莉

(51) Int. Cl.

H04L 69/22 (2022.01)

H04L 69/06 (2022.01)

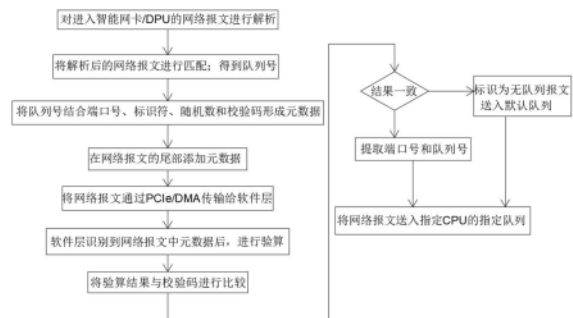
权利要求书1页 说明书6页 附图1页

(54) 发明名称

一种在高速智能网卡/DPU内节约DMA通道资源方法和装置

(57) 摘要

本发明一种在高速智能网卡/DPU内节约DMA通道资源方法和装置,对进入智能网卡/DPU的网络报文进行解析;将解析后的网络报文进行匹配;得到队列号;将队列号结合端口号、标识符、随机数和校验码形成元数据;在网络报文的尾部添加元数据;将网络报文通过PCIe/DMA传输给软件层;软件层识别到网络报文中元数据后,通过标识符、端口号、队列号和随机数进行验算,并将验算结果与校验码进行比较;若验算结果与校验码一致,提取出元数据中的端口号和队列号;若验算结果与校验码不一致,标识为无队列报文,送入默认队列;将网络报文送入指定CPU的指定队列。在高速智能网卡/DPU内用单个通道来实现对应的多个端口及多个队列;节省芯片内的逻辑资源,节省CPU资源,降低成本。



1. 一种在高速智能网卡/DPU内节约DMA通道资源方法,其特征在于,具体包括如下步骤:

S1、对进入智能网卡/DPU的网络报文进行解析;

S2、通过解析网络报文得到的五元组或用户定义的字段进行匹配,得到队列号;将队列号结合端口号、标识符、随机数和校验码形成元数据;所述端口号为1个字节,通过网络报文进入智能网卡/DPU的端口序号所决定;所述标识符为3个字节,通过用户自定义获得;

校验码的计算方式如下:

S21、系统准备256个素数;

S22、用标识符、端口号、队列号和随机数的8个字节,根据每个字节中的值,取对应的素数,得到8个素数;

S23、将所得到的8个素数进行左移,按位加运算,具体公式如下:

$R = (R \ll 1) + V$;其中R的初始值是第一个素数,V的初始值是第二个素数,每运算一次V的值变成下一个素数,共运算7次;最终结果即为校验码;

S3、在网络报文的尾部添加元数据;并将网络报文通过DMA数据搬运方式的从PCIe标准协议接口传输给软件层;

S4、软件层识别到网络报文中元数据后,通过标识符、端口号、队列号和随机数进行验算,并将验算结果与校验码进行比较;

S5、若验算结果与校验码一致,则提取出元数据中的端口号和队列号;若验算结果与校验码不一致,则标识为无队列报文,送入默认队列;

S6、将网络报文送入指定CPU的指定队列。

2. 如权利要求1所述的一种在高速智能网卡/DPU内节约DMA通道资源方法,其特征在于,步骤S2中所述随机数为2个字节,通过系统随机生成。

3. 如权利要求1所述的一种在高速智能网卡/DPU内节约DMA通道资源方法,其特征在于,步骤S4中通过标识符、端口号、队列号和随机数进行验算的具体过程如下:

S41、取与S21相同的256个素数;

S42、用标识符、端口号、队列号和随机数的8个字节,根据每个字节中的值,取对应的素数,得到8个素数;

S43、将所得到的8个素数进行左移,按位加运算,具体公式如下:

$R = (R \ll 1) + V$;其中R的初始值是第一个素数,V的初始值是第二个素数,每运算一次V的值变成下一个素数,共运算7次;得到验算结果。

4. 一种在高速智能网卡/DPU内节约DMA通道资源装置,其特征在于,包括存储器和一个或多个处理器,所述存储器中存储有可执行代码,所述一个或多个处理器执行所述可执行代码时,用于实现权利要求1-3任一项所述的一种在高速智能网卡/DPU内节约DMA通道资源方法。

5. 一种计算机可读存储介质,其特征在于:其上存储有程序,该程序被处理器执行时,实现权利要求1-3任一项所述的一种在高速智能网卡/DPU内节约DMA通道资源方法。

一种在高速智能网卡/DPU内节约DMA通道资源方法和装置

技术领域

[0001] 本发明涉及智能网卡/DPU技术领域,特别涉及一种在高速智能网卡/DPU内节约DMA通道资源方法和装置。

背景技术

[0002] 随着5G及IoT等技术的应用及网络的部署,网络吞吐量及接入量大幅增加,网络报文处理的压力汇聚到了服务器端,随着带宽的增加,25GbE,到100GbE,网络需要多个CPU核参与处理,这就不得不影响CPU正在处理的业务,用智能网卡来协助CPU处理网络报文可以释放CPU资源。而DPU(数据处理单元)为智能网卡的升级版,可以在智能网卡基础上也同时卸载了网络及存储负荷,让CPU可以参与更多业务。

[0003] 现有的高速智能网卡/DPU通过PCIe/DMA的方式来与主机进行数据交互,支持1-4个端口,每个端口支持多个队列,如果在网卡芯片内实现,多个端口需要有DMA对应的多个通道来一一对应支持,同时为了支持多个队列,需要在高速智能网卡/DPU内实现多队列的转发,多个DMA通道及多队列的转发会占用网卡/DPU芯片内的逻辑资源,增加逻辑复杂性及相应成本;如果在软件层面实现,则需要软件层面来实现报文的解析及转发,消耗大量CPU资源。

发明内容

[0004] 本发明的目的在于提供一种在高速智能网卡/DPU内节约DMA通道资源方法和装置,以克服现有技术中的不足。

[0005] 为实现上述目的,本发明提供如下技术方案:

[0006] 本发明公开了一种在高速智能网卡/DPU内节约DMA通道资源方法,具体包括如下步骤:

[0007] S1、对进入智能网卡/DPU的网络报文进行解析;

[0008] S2、将解析后的网络报文进行匹配;得到队列号;将队列号结合端口号、标识符、随机数和校验码形成元数据;

[0009] S3、在网络报文的尾部添加元数据;并将网络报文通过PCIe/DMA传输给软件层;

[0010] S4、软件层识别到网络报文中元数据后,通过标识符、端口号、队列号和随机数进行验算,并将验算结果与校验码进行比较;

[0011] S5、若验算结果与校验码一致,则提取出元数据中的端口号和队列号,将网络报文送入指定CPU的指定队列;若验算结果与校验码不一致,则标识为无队列报文,送入默认队列。

[0012] 作为优选,步骤S2中端口号为1个字节,通过网络报文进入智能网卡/DPU的端口序号所决定。

[0013] 作为优选,步骤S2中标识符为3个字节,通过用户自定义获得;所述随机数为2个字节,通过系统随机生成;

- [0014] 作为优选,步骤S2中校验码的计算方式如下:
- [0015] S21、系统准备256个素数;
- [0016] S22、用标识符、端口号、队列号和随机数的8个字节,根据每个字节中的值,取对应的素数,得到8个素数;
- [0017] S23、将所得到的8个素数进行左移,按位加运算,具体公式如下:
- [0018] $R=(R\ll 1)+V$;其中R的初始值是第一个素数,V的初始值是第二个素数,每运算一次V的值变成下一个素数,共运算7次;最终结果即为校验码。
- [0019] 作为优选,步骤S2中将解析后的网络报文进行匹配,得到队列号;具体操作如下:通过解析网络报文得到的五元组或用户定义的字段进行匹配,得到对应的队列号。
- [0020] 作为优选,步骤S4中通过标识符、端口号、队列号和随机数进行验算的具体过程如下:
- [0021] S41、取与S21相同的256个素数;
- [0022] S42、用标识符、端口号、队列号和随机数的8个字节,根据每个字节中的值,取对应的素数,得到8个素数;
- [0023] S43、将所得到的8个素数进行左移,按位加运算,具体公式如下:
- [0024] $R=(R\ll 1)+V$;其中R的初始值是第一个素数,V的初始值是第二个素数,每运算一次V的值变成下一个素数,共运算7次;得到验算结果。
- [0025] 本发明还公开了一种在高速智能网卡/DPU内节约DMA通道资源装置,包括存储器和一个或多个处理器,所述存储器中存储有可执行代码,所述一个或多个处理器执行所述可执行代码时,用于上述的一种在高速智能网卡/DPU内节约DMA通道资源方法。
- [0026] 本发明还公开了一种计算机可读存储介质,其上存储有程序,该程序被处理器执行时,实现上述的一种在高速智能网卡/DPU内节约DMA通道资源方法。
- [0027] 本发明的有益效果:
- [0028] 1、高速智能网卡/DPU的PCIe/DMA,在支持多个端口及队列时,需要多个DMA通道支持数据的分流,或者在高速智能网卡/DPU内实现多队列的转发,消耗了多个通道及转发的逻辑资源;如果在软件层面实现,则需要软件层面来实现报文的解析及转发,消耗大量CPU资源;本发明在高速智能网卡/DPU内用单个通道来实现对应的多个端口及多个队列。
- [0029] 2、节省芯片内的逻辑资源,节省CPU资源,降低成本。
- [0030] 本发明的特征及优点将通过实施例结合附图进行详细说明。

附图说明

- [0031] 图1是本发明一种在高速智能网卡/DPU内节约DMA通道资源方法的流程图;
- [0032] 图2是本发明一种在高速智能网卡/DPU内节约DMA通道资源装置的结构示意图。

具体实施方式

[0033] 为使本发明的目的、技术方案和优点更加清楚明了,下面通过附图及实施例,对本发明进行进一步详细说明。但是应该理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限制本发明的范围。此外,在以下说明中,省略了对公知结构和技术的描述,以避免不必要地混淆本发明的概念。

[0034] 参阅图1,本发明实施例提供一种在高速智能网卡/DPU内节约DMA通道资源方法,具体包括如下步骤:

[0035] S1、对进入智能网卡/DPU的网络报文进行解析;

[0036] S2、将解析后的网络报文进行匹配;得到队列号;将队列号结合端口号、标识符、随机数和校验码形成元数据;

[0037] S3、在网络报文的尾部添加元数据;并将网络报文通过PCIe/DMA传输给软件层;

[0038] S4、软件层识别到网络报文中元数据后,通过标识符、端口号、队列号和随机数进行验算,并将验算结果与校验码进行比较;

[0039] S5、若验算结果与校验码一致,则提取出元数据中的端口号和队列号,将网络报文送入指定CPU的指定队列;若验算结果与校验码不一致,则标识为无队列报文,送入默认队列。

[0040] 在一种可行的实施例中,步骤S2中端口号为1个字节,通过网络报文进入智能网卡/DPU的端口序号所决定。

[0041] 在一种可行的实施例中,步骤S2中标识符为3个字节,通过用户自定义获得;所述随机数为2个字节,通过系统随机生成;

[0042] 在一种可行的实施例中,步骤S2中校验码的计算方式如下:

[0043] S21、系统准备256个素数;

[0044] S22、用标识符、端口号、队列号和随机数的8个字节,根据每个字节中的值,取对应的素数,得到8个素数;

[0045] S23、将所得到的8个素数进行左移,按位加运算,具体公式如下:

[0046] $R=(R\ll 1)+V$;其中R的初始值是第一个素数,V的初始值是第二个素数,每运算一次V的值变成下一个素数,共运算7次;最终结果即为校验码。

[0047] 在一种可行的实施例中,步骤S2中将解析后的网络报文进行匹配,得到队列号;具体操作如下:通过解析网络报文得到的五元组或用户定义的字段进行匹配,得到对应的队列号。

[0048] 在一种可行的实施例中,步骤S4中通过标识符、端口号、队列号和随机数进行验算的具体过程如下:

[0049] S41、取与S21相同的256个素数;

[0050] S42、用标识符、端口号、队列号和随机数的8个字节,根据每个字节中的值,取对应的素数,得到8个素数;

[0051] S43、将所得到的8个素数进行左移,按位加运算,具体公式如下:

[0052] $R=(R\ll 1)+V$;其中R的初始值是第一个素数,V的初始值是第二个素数,每运算一次V的值变成下一个素数,共运算7次;得到验算结果。

[0053] 实施例:

[0054] 智能网卡/DPU对从光口进入网卡内的报文进行解析;例如,对以太网报文进行解析,解析出五元组或用户定义的字段。

[0055] 通过端口及五元组或其他用户定义的规则进行匹配;例如:对于端口1进来的数据标注为端口号1,对于端口2进来的报文标注为端口号2,同时对于用户定义的源IP地址进行匹配,192.168.0.*标注为队列2,对于用户定义的源IP地址进行匹配,192.168.1.*标注为

队列3;

[0056] 在匹配到的报文尾部添加端口及队列号的元数据,

[0057] 为了避免端口及队列号与数据报文的偶然性重复,定义端口及队列号的元数据格式如下:

字节	0	1	2	3	4	5	6	7
内容 (16进制)	aa	bb	cc	端口号	队列号	随机数		
	校验码							

[0058] 其中aabbcc为标识符;

[0060] 端口号为1个字节,通过网络报文进入智能网卡/DPU的端口序号所决定。

[0061] 标识符为3个字节,通过用户自定义获得;所述随机数为2个字节,通过系统随机生成;

[0062] 校验码计算方式如下:

[0063] 系统准备256个素数,编号0-255

[0064] 用标识符,端口号,队列号,随机数的8个字节,每个字节中的值,去取对应编号的素数,例如第一个字节aa十进制为170,则取第170个素数为265aa7bb,将所得的8个素数进行左移,按位加运算,公式1如下:

[0065] $R = (R \ll 1) + V$

[0066] R的初始值是第一个素数,V的初始值是第二个素数,每运算一次V的值变成下一个素数,总共运算7次;

[0067] 例1:端口号01,队列号0002,随机数0305,

[0068] 添加报文的前8个字节为 aabbcc01 00020305

[0069] 8个素数V(16进制)为

[0070] aa(170)对应素数V0: 4bc5d4f3

[0071] bb(187)对应素数V1: 42dff19f

[0072] cc(204)对应素数V2: 1ee35bb

[0073] 01(1)对应素数V3: 3a907251

[0074] 00(0)对应素数 V4: 1fb0dfc9

[0075] 02(51)对应素数V5: 7b720269

[0076] 03(03)对应素数V6: 6850364b

[0077] 05(05)对应素数V7: 02b3b673

[0078] 按照公式 $R = (R \ll 1) + V$

[0079] $R_0 = 265aa7bb$

[0080] $R_1 = (R_0 \ll 1) + V_1 = (4bc5d4f3 \ll 1) + 42dff19f = da6b9b85$

[0081] $R_2 = (R_1 \ll 1) + V_2 = (da6b9b85 \ll 1) + 1ee35bb = d3c36cc5$

[0082] $R_3 = (R_2 \ll 1) + V_3 = (d3c36cc5 \ll 1) + 3a907251 = e2174bdb$

[0083] $R_4 = (R_3 \ll 1) + V_4 = (e2174bdb \ll 1) + 1fb0dfc9 = e3df777f$

[0084] $R_5 = (R_4 \ll 1) + V_5 = (e3df777f \ll 1) + 7b720269 = 4330f167$

[0085] $R_6 = (R_5 \ll 1) + V_6 = (4330f167 \ll 1) + 6850364b = eeb21919$

[0086] $R_7 = (R_6 \ll 1) + V_7 = (eeb21919 \ll 1) + 02b3b673 = e017e8a5$

- [0087] 校验码为e017e8a5
- [0088] 元数据为aabbcc0100020305e017e8a5;
- [0089] 例2:端口号02,队列号0008,随机数0305,
- [0090] 添加报文的前16个字节为 aabbcc0200080305
- [0091] 8个素数V(16进制)为
- [0092] aa(170)对应素数V0: 4bc5d4f3
- [0093] bb(187)对应素数V1: 42dff19f
- [0094] cc(204)对应素数V2: 1eec35bb
- [0095] 02(2)对应素数V3: 56cb4033
- [0096] 00(0)对应素数V4: 1fb0dfc9
- [0097] 08(8)对应素数V5: 45039065
- [0098] 03(3)对应素数V6: 6850364b
- [0099] 05(5)对应素数V7: 02b3b673
- [0100] 按照公式 $R=(R \ll 1)+V$
- [0101] $R0=4bc5d4f3$
- [0102] $R1=(R0 \ll 1)+V1=(4bc5d4f3 \ll 1)+42dff19f=da6b9b85$
- [0103] $R2=(R1 \ll 1)+V2=(da6b9b85 \ll 1)+1eec35bb=d3c36cc5$
- [0104] $R3=(R2 \ll 1)+V3=(d3c36cc5 \ll 1)+56cb4033=fe5219bd$
- [0105] $R4=(R3 \ll 1)+V4=(fe5219bd \ll 1)+1fb0dfc9=1c551343$
- [0106] $R5=(R4 \ll 1)+V5=(1c551343 \ll 1)+45039065=7dad6eb$
- [0107] $R6=(R5 \ll 1)+V6=(7dad6eb \ll 1)+6850364b=63aba421$
- [0108] $R7=(R6 \ll 1)+V7=(63aba421 \ll 1)+02b3b673=ca0afeb5$
- [0109] 校验码为ca0afeb5
- [0110] 元数据为aabbcc0200080305ca0afeb5;
- [0111] 把报文通过PCIe/DMA传输给软件层;
- [0112] 软件层识别到标识符后,会用相同的对应的素数及报文中携带的端口号,队列号及随机数使用公式1进行计算,方式如例1和例2
- [0113] 用计算得到的值和校验码做比较,如果结果一致,则认为这段数据是写入端口号及队列号的数据,从而提取出端口号和队列号。如果结果不一致,则是无队列报文,送入默认队列。
- [0114] 驱动把相应的报文送入指定CPU指定队列。
- [0115] 通过上述描述的方法,以Xilinx VU3P FPGA实现相应的PCIe/DMA逻辑,使用本发明中单通道DMA的逻辑资源占用及传统方式多通道多队列的方式的DMA逻辑资源占用的对比如表1所示。

[0116]

表 1: 使用发明中单通道 DMA 的逻辑资源占用 VS 双通道 32 队列 DMA 逻辑资源占用			
	CLB LUTs	CLB registers	Block RAM
单通道 DMA 的逻辑资源占用	43348	44373	73
双通道 32 队列 DMA 逻辑资源占用	76648	68175	159

[0117] 从表1中,两种方式对比可以看出逻辑资源CLB LUTs 节省43%,CLB registers节省34.9%,Block RAM节省41.5%。

[0118] 本发明一种在高速智能网卡/DPU内节约DMA通道资源装置的实施例可以应用在任意具备数据处理能力的设备上,该任意具备数据处理能力的设备可以为诸如计算机等设备或装置。装置实施例可以通过软件实现,也可以通过硬件或者软硬件结合的方式实现。以软件实现为例,作为一个逻辑意义上的装置,是通过其所在任意具备数据处理能力的设备的处理器将非易失性存储器中对应的计算机程序指令读取到内存中运行形成的。从硬件层面而言,如图2所示,为本发明一种在高速智能网卡/DPU内节约DMA通道资源装置所在任意具备数据处理能力的设备的一种硬件结构图,除了图2所示的处理器、内存、网络接口、以及非易失性存储器之外,实施例中装置所在的任意具备数据处理能力的设备通常根据该任意具备数据处理能力的设备的实际功能,还可以包括其他硬件,对此不再赘述。上述装置中各个单元的功能和作用的实现过程具体详见上述方法中对应步骤的实现过程,在此不再赘述。

[0119] 对于装置实施例而言,由于其基本对应于方法实施例,所以相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本发明方案的目的。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0120] 本发明实施例还提供一种计算机可读存储介质,其上存储有程序,该程序被处理器执行时,实现上述实施例中的一种在高速智能网卡/DPU内节约DMA通道资源方法。

[0121] 所述计算机可读存储介质可以是前述任一实施例所述的任意具备数据处理能力的设备的内部存储单元,例如硬盘或内存。所述计算机可读存储介质也可以是任意具备数据处理能力的设备的外部存储设备,例如所述设备上配备的插接式硬盘、智能存储卡(Smart Media Card, SMC)、SD卡、闪存卡(Flash Card)等。进一步的,所述计算机可读存储介质还可以既包括任意具备数据处理能力的设备的内部存储单元也包括外部存储设备。所述计算机可读存储介质用于存储所述计算机程序以及所述任意具备数据处理能力的设备所需的其他程序和数据,还可以用于暂时地存储已经输出或者将要输出的数据。

[0122] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换或改进等,均应包含在本发明的保护范围之内。

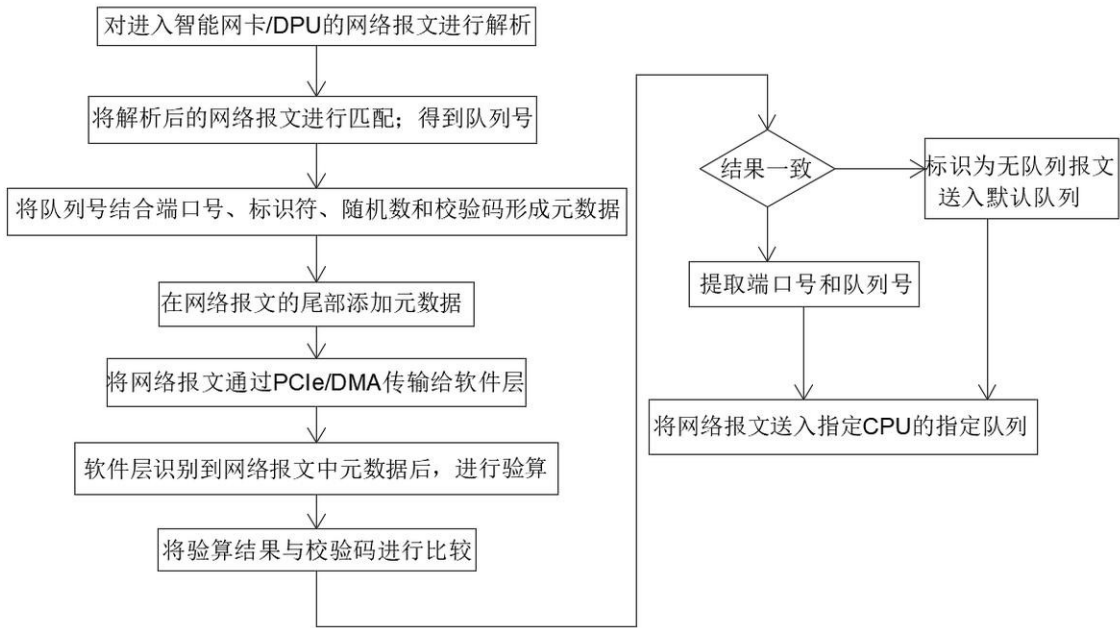


图1

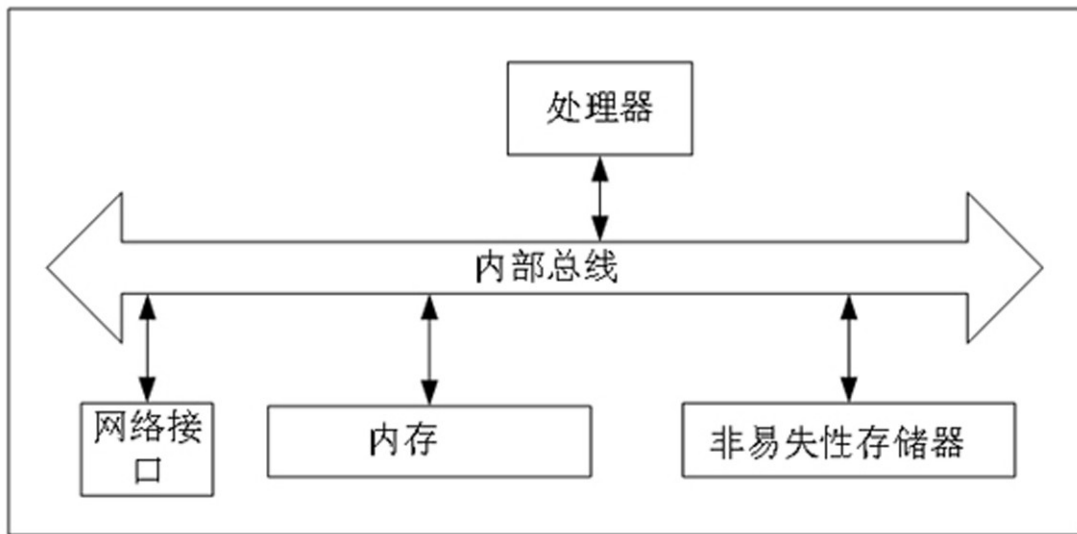


图2