



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(21) BR 102019010425-2 A2



(22) Data do Depósito: 22/05/2019

(43) Data da Publicação Nacional: 10/12/2019

(54) Título: MÉTODO PARA ENVIO DE DADOS A PARTIR DE UM DISPOSITIVO A MONTANTE PARA UM DISPOSITIVO A JUSANTE, PROGRAMA DE COMPUTADOR DESCARREGÁVEL A PARTIR DE UMA REDE DE COMUNICAÇÃO OU GRAVADO EM SUPORTE LEGÍVEL OU EXECUTÁVEL POR UM PROCESSADOR E SISTEMA DE ENVIO DE DADOS

(51) Int. Cl.: G06F 3/01.

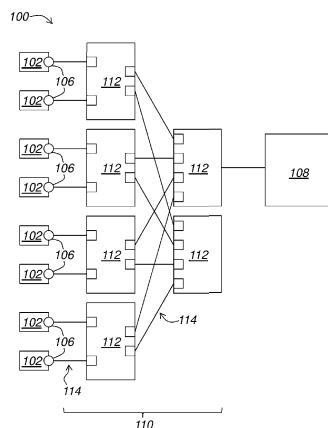
(52) CPC: G06F 3/01.

(30) Prioridade Unionista: 22/05/2018 FR 1854232.

(71) Depositante(es): BULL SAS.

(72) Inventor(es): PIERRE AXEL LAGADEC; SAÏD DERRADJI; DOMINIQUE RIGAL; LAURENT MARLIAC.

(57) Resumo: MÉTODO PARA ENVIO DE DADOS A PARTIR DE UM DISPOSITIVO A MONTANTE PARA UM DISPOSITIVO A JUSANTE, PROGRAMA DE COMPUTADOR DESCARREGÁVEL A PARTIR DE UMA REDE DE COMUNICAÇÃO OU GRAVADO EM SUPORTE LEGÍVEL OU EXECUTÁVEL POR UM PROCESSADOR E SISTEMA DE ENVIO DE DADOS. Este método de envio de dados a partir de um dispositivo a montante (202) até um dispositivo a jusante (204) compreende o envio de dados a partir de uma pluralidade de canais virtuais que dividem uma mesma memória buffer de entrada (208) do dispositivo a jusante (204), se o canal virtual utiliza um número de locais de memória da memória buffer de entrada (208) estritamente inferior a um limite máximo atual. Compreende ainda a medição da latência da comunicação entre os dispositivos a montante (202) e a jusante (204) e o cálculo do limite máximo atual a partir da latência.



"MÉTODO PARA ENVIO DE DADOS A PARTIR DE UM DISPOSITIVO A MONTANTE PARA UM DISPOSITIVO A JUSANTE, PROGRAMA DE COMPUTADOR DESCARREGÁVEL A PARTIR DE UMA REDE DE COMUNICAÇÃO OU GRAVADO EM SUPORTE LEGÍVEL OU EXECUTÁVEL POR UM PROCESSADOR E SISTEMA DE ENVIO DE DADOS"

[0001] A presente invenção refere-se a um método de envio de dados, bem como um programa de computador e um sistema associado.

[0002] A invenção aplica-se mais especificamente a um método de envio de dados de um dispositivo a montante para um dispositivo a jusante, compreendendo:

- o envio de dados de um de uma pluralidade de canais virtuais que compartilham a mesma memória buffer de entrada do dispositivo a jusante, se o canal virtual usar um número de locais de memória buffer de entrada estritamente inferior a um limite máximo atual.

[0003] Por exemplo, os dispositivos a montante e a jusante pertencem a um sistema que compreende nós e uma rede de interconexão de nós que compreende comutadores e links de comunicação que conectam os comutadores entre si e com os nós. Cada um dos dispositivos a montante e a jusante é então um dos nós ou um dos comutadores. Os comutadores e os nós possuem portas de conexão às quais os links de comunicação estão conectados.

[0004] Neste contexto, já é conhecida a prática de determinar arbitrariamente o limite máximo atual com o mesmo valor para todas as memórias buffer de entrada presentes no sistema. Esse valor é escolhido em grande escala para que cada canal virtual tenha à sua disposição locais de memória suficientes para que os dados desse canal virtual sejam

enviados sem problemas.

[0005] Um problema com essa solução é que ela requer o uso de buffers de entrada de alta capacidade.

[0006] Assim, pode ser desejável fornecer um método de enviar dados que permita superar pelo menos alguns dos problemas e restrições supramencionados.

[0007] O objeto da invenção é, portanto, um método de envio de dados de um dispositivo a montante para um dispositivo a jusante, compreendendo:

- o envio de dados de um de uma pluralidade de canais virtuais que compartilham a mesma memória buffer de entrada do dispositivo a jusante, se o canal virtual usar um número de locais de memória buffer de entrada estritamente inferior a um limite máximo atual, caracterizado por comportar, ainda:
 - a medida de uma latência de comunicação entre os dispositivos a montante e a jusante,
 - o cálculo do limite máximo atual a partir da latência.

[0008] De fato, o comprimento dos links de comunicação não é uniforme entre os nós do sistema. Graças à invenção, é possível calcular um limite máximo diferente para cada par de dispositivos a montante e a jusante. O limite máximo é sempre calculado exatamente de acordo com a latência de comunicação, que depende do comprimento real do link de comunicação, o que torna possível usar apenas a quantidade necessária de memória para o canal virtual, mantendo um bom desempenho.

[0009] Opcionalmente, a medição de latência compreende a medição do tempo gasto por uma troca de mensagens entre os dispositivos a montante e a jusante.

[0010] Também de modo opcional, o método compreende ainda:

- determinar uma taxa de utilização da memória buffer de

entrada, e o limite máximo atual é calculado a partir da latência e da taxa de uso.

[0011] De modo opcional também, o cálculo do limite máximo atual compreende:

- o cálculo de vários limites máximos possíveis a partir da latência,
- a seleção, quanto ao limite máximo atual, de um dos limites máximos possíveis a partir da taxa de utilização.

[0012] Também opcionalmente, os possíveis limites máximos são, respectivamente, associados a faixas adjacentes de taxas de utilização, e o limite máximo possível selecionado é aquele em que a taxa de utilização determinada é encontrada.

[0013] Também de modo opcional, um intervalo inferior e um intervalo superior adjacentes são delimitados um do outro por um limiar de valor alto quando o limite máximo possível selecionado é aquele associado ao intervalo inferior e um valor baixo, menor que valor alto, quando o limite máximo possível selecionado é aquele associado ao intervalo superior.

[0014] Também de modo opcional, o método compreende ainda:

[0015] o cálculo, a partir da latência, de um número reservado de locais de memória para cada canal virtual,

[0016] e os dados são enviados se o canal virtual usar um número de locais de memória estritamente menor que o limite máximo e se a memória buffer de entrada tiver mais locais de memória livre do que os locais reservados livres para todos os canais virtuais.

[0017] Também de modo opcional, os dispositivos a montante e a jusante pertencem a um sistema que compreende:

- nós,

- uma rede de interconexão de nós que compreende comutadores e links de comunicação que conectam comutadores entre si e aos nós,
- cada um dos dispositivos a montante e a jusante é um nó ou um comutador.

[0018] A invenção também se refere a um programa de computador que pode ser baixado de uma rede de comunicação e/ou gravado em um meio legível por computador e/ou executável por um processador, caracterizado por compreender instruções para a execução de passos de um método de acordo com a invenção, quando o referido programa é executado em um computador.

[0019] A invenção também se refere a um sistema de envio de dados que comporta:

- um dispositivo a jusante compreendendo uma memória buffer de entrada,
- um dispositivo a montante concebido para enviar dados de um de uma pluralidade de canais virtuais que partilham a memória buffer de entrada, se este canal virtual utilizar um número de locais de memória buffer de entrada estritamente inferior a um limite máximo atual,
- meios para medir a latência de comunicação entre os dispositivos a montante e a jusante,
- meios de cálculo do limite máximo atual a partir da latência.

[0020] A invenção será melhor compreendida com o auxílio da descrição que se segue, dada apenas a título de exemplo e com referência aos desenhos anexos, nos quais:

[0021] A figura 1 representa esquematicamente a estrutura geral de um sistema, de acordo com uma forma de realização da

invenção,

[0022] A figura 2 representa esquematicamente a estrutura geral de um dispositivo a montante e um dispositivo a jusante do sistema da figura 1,

[0023] A figura 3 ilustra os passos sucessivos de um método de envio de dados do dispositivo a montante para o dispositivo a jusante da figura 2, de acordo com uma forma de realização da invenção.

[0024] Com referência à figura 1, um sistema 100 que concretiza a invenção será agora descrito.

[0025] O sistema 100 agrupa dispositivos de computador 102 independentes que parecem externamente com um único computador com poder computacional muito alto, denominado computador de alto desempenho ou supercomputador ou computador HPC (High Performance Computing, em inglês).

[0026] Cada dispositivo de computação 102 tem pelo menos uma interface de rede. Cada dispositivo de computação 102 pode ainda compreender uma unidade de processamento central ("central processing unit" em inglês), uma memória principal ("main memory" em inglês) na qual as instruções para a unidade de processamento central devem ser gravadas (tal dispositivo de computação 102 também é denominado computador) e/ou um dispositivo de armazenamento de dados (tal dispositivo de computação é também denominado entrada/saída (ou I/O, do inglês "Input/Output")), tal como uma memória de massa como um ou mais discos rígidos, para gravação e ou recuperação de dados, por exemplo, e/ou produtos para dispositivos de computação.

[0027] As interfaces de rede compõem respectivamente os nós 106 do sistema 100. Várias interfaces de rede (e portanto

vários nós 106) podem pertencer ao mesmo dispositivo de computação 102, ou, como no exemplo descrito, cada dispositivo de computação 102 pode compreender apenas uma interface de rede única formando um dos nós 106. Um nó 106 formado por uma interface de rede de um dispositivo de computação 102 é chamado de "nó de computação", enquanto um nó 106 formado por uma interface de rede de um dispositivo de computação de entrada/saída 102 é chamado de "nó de entrada/saída".

[0028] O sistema 100 compreende ainda um servidor de administração 108. O servidor de administração 108 é, em particular, concebido para comunicar com clientes que desejam utilizar o sistema 100, sendo este último visto pelos clientes como um único computador. O servidor de administração 108 é ainda concebido para programar, entre dispositivos informáticos de computação, tarefas recebidas de clientes.

[0029] Tal sistema 100 permite a repartição de processos complexos e/ou de cálculos paralelos em ao menos uma parte dos dispositivos informáticos de computação.

[0030] O sistema 100 comporta ainda uma rede 110 de interconexão de nós 106.

[0031] A rede 110 compreende os comutadores 112 e os links de comunicação 114 que ligam os comutadores 112 uns aos outros e aos nós 106. Cada comutador 112 e cada nó 106 tem uma pluralidade de portas de conexão (representadas por pequenos quadrados na figura 1 para os comutadores 112).) e cada link de comunicação 114 está conectado entre uma porta de conexão de um comutador 112 e uma porta de conexão de outro comutador 112 ou uma porta de conexão de um nó 106.

[0032] O sistema 100 tem uma determinada topologia que define a disposição relativa dos nós 106, dos comutadores 112 e dos links de comunicação 114. No exemplo descrito, a topologia é PGFT (do inglês "Parallel Generalized Fat Tree").

[0033] Com referência à Fig. 2, um dispositivo a montante 202, um dispositivo a jusante 204 e um link de comunicação 114 que os liga, serão agora descritos. No exemplo descrito, os dispositivos a montante 202 e a jusante 204 são dois dos comutadores 112 da Fig. 1. Alternativamente, um dos dispositivos a montante 202 e a jusante 204 poderia ser um dos nós 106.

[0034] No exemplo descrito, as portas e os links de comunicação 114 são bidirecionais. Assim, cada comutador 112 e cada nó 106 é um dispositivo a jusante e um dispositivo a montante. Assim, cada uma de suas portas é tanto uma porta de entrada quanto uma porta de saída, de modo que eles possuem todos os elementos que serão descritos, tanto a porta de saída do dispositivo a montante 202 quanto a porta de entrada do dispositivo a jusante 204.

[0035] O dispositivo a jusante 204 tem uma porta de entrada 206 à qual o link de comunicação 114 está ligado para recebimento de dados. A porta de entrada 206 suporta um número predeterminado K de canais virtuais e uma taxa de bits D de, por exemplo, 25 GB/s. A porta de entrada 206 inclui um buffer de entrada 208 tendo um número de locais de memória cada para receber uma unidade de dados, cujo número é referido como a capacidade C da memória buffer de entrada 208. Como descrito, a unidade de dados é uma unidade de controle de fluxo ou FLIT, que é composta por 32 bytes. A capacidade C do buffer de entrada 208 é, por exemplo, 4096

FLITs.

[0036] O dispositivo a jusante 204 inclui ainda uma unidade de processamento 212, tal como um microprocessador, e uma memória 214 acoplada de modo funcional à unidade de processamento 212 e tendo um programa de computador incluindo instruções para a unidade de processamento 212.

[0037] O dispositivo a montante 202 tem uma porta de saída 216 à qual o link de comunicação 114 está ligado. A porta de saída 216 suporta o mesmo número predeterminado K de canais virtuais e a mesma taxa de bits D que a porta de entrada 206. Ela compreende contadores de entrada $218_1 \dots 218_K$ para os canais virtuais K, respectivamente, bem como um contador de utilização 220 do buffer de entrada 208 para contar os locais de memória utilizados por todos os canais virtuais. Assim, o buffer de entrada 208 tem uma taxa de utilização TU igual à razão entre o valor do contador de utilização 220 e a capacidade C do buffer de entrada 208.

[0038] O dispositivo a montante 202 compreende ainda uma unidade de processamento 222, tal como um microprocessador, e uma memória 224 acoplada operacionalmente à unidade de processamento 222 e incluindo um programa de computador com instruções para a unidade de processamento 222.

[0039] Assim, no exemplo descrito, os dispositivos a montante 202 e a jusante 204 compreendem meios de software, na forma de um programa de computador descarregável a partir de uma rede de comunicação e/ou gravado num suporte legível por computador e/ou executável por um processador, compreendendo instruções para executar os passos do método que será descrito mais tarde quando o referido programa de computador for executado nos dispositivos a montante 202 e a

jusante 204. Alternativamente, todos ou parte destes meios podem ser meios de hardware microprogramados ou com microfios em circuitos integrados dedicados. Assim, como variante, os dispositivos a montante 202 e a jusante 204 poderiam ser, cada um, um dispositivo eletrônico composto apenas por circuitos digitais (sem um programa de computador) para realizar as mesmas ações.

[0040] Além disso, a capacidade C da memória buffer de entrada 208 e a taxa D são armazenadas na memória 224 para permitir a sua utilização nos cálculos que serão descritos mais tarde.

[0041] Com referência à figura 3, um método 300 para enviar dados do dispositivo a montante 202 para o dispositivo a jusante 204 será agora descrito.

[0042] Durante um passo 302, o dispositivo a montante 202 mede uma latência (denominada em seguida L) de comunicação entre os dispositivos a montante 202 e os dispositivos a jusante 204.

[0043] No exemplo descrito, quando é iniciado, o sistema 100 implementa uma fase de inicialização durante a qual os comutadores 112 trocam mensagens de inicialização entre si e com os nós 106. Estas trocas são utilizadas para determinar a latência L, medindo o tempo gasto por uma troca de mensagens entre os dispositivos a montante 202 e a jusante 204 durante a fase de inicialização. Mais especificamente, o dispositivo a montante 202 envia uma primeira mensagem de inicialização ao dispositivo a jusante 204 através do link de comunicação 114. O dispositivo a jusante 204 recebe a primeira mensagem de inicialização e, em resposta, envia uma segunda mensagem de inicialização ao dispositivo a montante

202, novamente através do link de comunicação 114. O dispositivo a montante 202 recebe então esta segunda mensagem de inicialização. Esse recebimento é separado do envio da mensagem de inicialização por um certo tempo, que o dispositivo a montante 202 mede. No exemplo descrito, a latência L corresponde a esse tempo medido.

[0044] Durante um passo 304, o dispositivo a montante 202 calcula, para cada uma de uma pluralidade de intervalos adjacentes de taxas de utilização TU do buffer de entrada 208, um possível limite máximo a partir da latência L . Como será descrito posteriormente, cada um desses limites máximos possíveis deve ser selecionado para formar um limite máximo atual P representando um número máximo de locais de memória que podem ser usados por cada canal virtual.

[0045] Por exemplo, cada um destes limites máximos possíveis é calculado a partir da latência L e da taxa D da porta de saída à qual o link de comunicação 114 está conectado.

[0046] Preferencialmente, cada limite máximo possível tem um valor mínimo igual à capacidade C do buffer de entrada 208 dividido pelo número de canais virtuais K . Por exemplo, com a capacidade C igual a 4096 locais de memória e o número de canais virtuais K igual a 20, o valor mínimo dos limites máximos possíveis é de 205 locais de memória.

[0047] De preferência, os possíveis limites máximos diminuem de um intervalo para o próximo, na direção crescente da taxa de utilização.

[0048] No exemplo descrito, dois intervalos são usados. O primeiro intervalo corresponde a uma taxa de utilização TU menor ou igual a 75% e o segundo intervalo corresponde a uma

taxa de utilização TU estritamente superior a 75%. O primeiro intervalo é associado a um primeiro teto máximo possível P1 dado no exemplo descrito pela fórmula : $P1 = \max (L \times D; C/K)$, em que L é a latência, D é a saída (em número de FLITs), C é a capacidade da memória buffer de entrada 208 e K é o número de canais virtuais. Por exemplo, se a latência L medida vale 400 ns e se a saída (em octetos) vale 25 Go/s ($0,78125 \times 10^9$ FLITs/s), o primeiro limite máximo possível P1 vale 313. O segundo intervalo é associado a um segundo limite máximo possível (denominado P2) dado no exemplo descrito pela fórmula: $P2 = \max (50\% \times L \times D; C/K)$, em que L é a latência, D é a saída (em número de FLITs), C é a capacidade da memória buffer de entrada 208 e K é o número de canais virtuais. Por exemplo, se a latência L medida vale 400 ns e se a saída (em octetos) vale 25 Go/s ($0,78125 \times 10^9$ FLITs/s), $50\% \times L \times D$ vale 157, de modo que o segundo limite máximo possível P2 tem o valor mínimo de 205.

[0049] Observa-se que dois intervalos adjacentes, um (chamado intervalo inferior) estando localizados abaixo do outro (chamado intervalo superior) na direção crescente da taxa de utilização TU, são delimitados um em relação ao outro por um limiar. No exemplo descrito, o limiar que separa os dois intervalos é de 75% e é fixo. Contudo, em outras formas de realização, o limiar pode variar de acordo com um fenómeno de histerese. Especificamente, quando o possível limite máximo selecionado é aquele associado ao intervalo inferior, o limiar assume um valor alto, enquanto que quando o possível limite máximo selecionado é aquele associado à faixa mais alta, o limiar assume um valor baixo, menor que o valor alto. Por exemplo, quando o possível limite máximo P1 é

selecionado, o limiar para passar para o segundo intervalo é de 80%, enquanto que quando o possível limite máximo P2 é selecionado, o limite para ir para o intervalo inferior é de 70%. Este fenômeno de histerese evita mudanças frequentes no limite máximo atual P quando a taxa de ocupação TU flutua em torno do limiar.

[0050] Durante um passo 306, o dispositivo a montante 202 calcula um número de locais de memória reservados por canal virtual (com tal número sendo denominado N_R) a partir da latência L. No exemplo descrito, a seguinte fórmula é utilizada: $N_R = 10\% \times L \times D$, em que L é a latência e D é a saída (em número de FLITs).

[0051] Durante uma etapa 308, o dispositivo a montante 202 seleciona o primeiro limite máximo possível P1 e o limite máximo atual P. No exemplo descrito, essa etapa é realizada inicializando cada contador de entrada $218_1 \dots 218_K$ com o primeiro número P1.

[0052] Durante uma etapa 310, o dispositivo a montante 202 inicializa o contador de utilização 220 a zero.

[0053] Subsequentemente, o primeiro ciclo de passos seguinte é implementado cada vez que o dispositivo a montante 202 procura enviar um FLIT, cujo canal virtual é denominado k_1 , para o dispositivo a jusante 204 através do link de comunicação 114.

[0054] Ao longo de uma etapa 312, o dispositivo a montante 202 determina a taxa de utilização TU do buffer de entrada 208, por exemplo, a partir da capacidade C do buffer de entrada 208 e do contador de utilização 220.

[0055] Durante uma etapa 314, o dispositivo a montante 202 seleciona, como limite máximo atual P, o número associado ao

intervalo em que a taxa de utilização determinada TU é encontrada.

[0056] Durante uma etapa 316, se o limite máximo P foi alterado após a etapa 314, o dispositivo a montante 202 calcula a diferença entre o limite máximo antigo P e o novo limite máximo P e aplica essa diferença a cada contador de entrada $218_1 \dots 218_K$.

[0057] Por exemplo, se o limite máximo atual P era, antes da etapa 314, o limite máximo possível P_2 , e se o limite máximo atual P é, no final da etapa 314, o limite máximo possível P_1 , a diferença será $P_1 - P_2$, ou seja, $313 - 205 = 108$, recuperando o exemplo descrito anteriormente. Assim, 108 entradas são adicionadas a cada contador de entradas $218_1 \dots 218_K$.

[0058] Inversamente, se o limite máximo atual P era, antes da etapa 314, o possível limite máximo P_1 e se o limite máximo atual P é, no final da etapa 314, o possível limite máximo P_2 , então a diferença é $P_2 - P$, ou -108 , tomando novamente o exemplo desenvolvido anteriormente. Assim, 108 entradas são deduzidas de cada contador de entrada $218_1 \dots 218_K$. Isso pode levar um contador de entrada a um valor negativo. Isto significa que o canal virtual correspondente excede o limite máximo atual P e, portanto, usa muitos locais de memória no buffer de entrada 208.

[0059] Durante um passo 318, o dispositivo a montante 202 verifica se o canal virtual do FLIT considerado usa um número de posições de memória estritamente menor ao limite máximo atual P .

[0060] No exemplo descrito, este passo é implementado verificando se o contador de entrada correspondente 218_K é

estritamente maior que zero.

[0061] Durante uma etapa 320, o dispositivo a montante 202 verifica se o buffer de entrada 208 tem mais locais de memória livre do que os locais reservados livres para todos os canais virtuais.

[0062] No exemplo descrito, essa etapa é implementada calculando, para cada canal virtual, o número de locais de memória utilizados por tal canal virtual (com este número sendo denominado $N(k)$, em que k é o canal virtual), utilizando por exemplo a fórmula: $N_U(k) = P - C(k)$, em que $C(k)$ é o valor do contador de entradas 218_k. Em seguida, o número de locais reservados livres deste canal virtual (denominado $N_{RL}(k)$) é calculado subtraindo o número de locais de memória utilizados pelo canal virtual do número de locais de memória reservados N_R e limitando o resultado a zero. Por exemplo, a seguinte fórmula: $N_{RL}(k) = \max(N_R - N_U(k); 0)$ é utilizada. O número total de locais reservados livres (ou seja, para todos os canais virtuais), denominado N_{RL} , é então calculado adicionando, a todos os canais virtuais, o número de locais reservados livres de cada canal virtual. Por exemplo, a fórmula seguinte: $N_{RL} = \sum_k N_{RL}(k)$ é utilizada. O número de locais de memória livres na memória buffer de entrada 208 é, por exemplo, dado pela seguinte fórmula: $N_L = C - E_U$, em que E_U é o número de locais de memória utilizados, dado pelo contador de utilização 220. O número total de locais reservados livres é então comparado ao número de locais de memória livres na memória buffer de entrada 208.

[0063] Durante uma etapa 322, se o canal virtual k_1 utiliza um número de locais de memória estritamente inferior ao limite máximo atual P e se a memória buffer de entrada 208

apresenta mais locais de memória livres do que os locais reservados livres para todos os canais virtuais, o dispositivo a montante 202 envia o FLIT F1, reduz o contador de entrada C_k e incrementa o contador de utilização 220. Caso contrário, ele não envia o FLIT F1 e volta à etapa 312 para outro FLIT.

[0064] Durante um passo 324, o dispositivo a jusante 204 recebe o FLIT F1 e guarda-o no seu buffer de entrada 208.

[0065] Em paralelo, o segundo ciclo de passos seguinte é implementado cada vez que o dispositivo a jusante 204 retira do seu buffer de entrada 208 um FLIT F2 (cujo canal virtual é denominado k_2), por exemplo, para o enviar para outro dispositivo.

[0066] Durante uma etapa 326, o dispositivo a jusante 204 remove o FLIT F2 do seu buffer de entrada 208.

[0067] Durante uma etapa 328, o dispositivo a jusante 204 envia uma entrada pelo canal virtual k_2 para o dispositivo a montante 202, através do link de comunicação 114.

[0068] Durante uma etapa 330, o dispositivo a montante 202 recebe o a entrada pelo canal virtual k_2 , incrementa o contador de entradas 218_{k_2} correspondente a este canal virtual k_2 e diminui o contador de utilização 220.

[0069] Observa-se que em outras formas de realização, o limite máximo atual pode ser diferente de um canal virtual para outro. Por exemplo, o limite máximo atual de cada canal virtual poderia ser calculado a partir de um histórico de uso desse canal virtual, a fim de reduzir o limite máximo de canais virtuais pouco utilizados no passado e, por outro lado, aumentar o limite máximo para canais virtuais muito usados.

[0070] Fica evidente que um método como o descrito acima torna possível reduzir a quantidade de memória buffer de entrada enquanto se mantém um bom desempenho.

[0071] Note-se também que a invenção não está limitada às formas de realização descritas acima. Será evidente para os especialistas na técnica que podem ser feitas várias modificações às formas de realização descritas acima, à luz do ensinamento que acaba de ser divulgado. Na apresentação detalhada da invenção que foi feita acima, os termos utilizados não devem ser interpretados como limitativos da invenção às formas de realização apresentadas nesta especificação, devendo ser interpretados como que incluindo todos os equivalentes cuja previsão está dentro do escopo dos especialistas, aplicando seus conhecimentos gerais à implementação do ensino que acaba de ser divulgado.

REIVINDICAÇÕES

1. Método para envio de dados a partir de um dispositivo a montante para um dispositivo a jusante, compreendendo:

- enviar (320) dados de um de uma pluralidade de canais virtuais partilhando o mesmo buffer de entrada (208) do dispositivo a jusante (204), se esse canal virtual utilizar um certo número de locais de memória buffer de entrada (208) estritamente menor que um limite máximo atual, caracterizado pelo fato de compreender ainda:

- medir (302) uma latência de comunicação entre os dispositivos a montante (202) e a jusante (204),
- calcular (304, 308) o limite máximo atual a partir da latência.

2. Método, de acordo com a reivindicação 1, caracterizado pelo fato de a medição (302) da latência incluir a medição do tempo levado por uma troca de mensagens entre os dispositivos a montante (202) e a jusante (204).

3. Método, de acordo com qualquer uma das reivindicações 1 ou 2, caracterizado pelo fato de compreender ainda:

- determinar uma taxa de utilização da memória buffer de entrada (208), e

sendo que o limite máximo atual é calculado a partir da latência e da taxa de utilização.

4. Método, de acordo com a reivindicação 3, caracterizado pelo fato de o cálculo do limite máximo compreender:

- calcular (304) vários limites máximos possíveis a partir da latência,
- selecionar (308), como limite máximo atual, um dos limites máximos possíveis a partir da taxa de utilização.

5. Método, de acordo com a reivindicação 4, caracterizado

pelo fato de os limites máximos possíveis serem, respectivamente, associados a uma relação de intervalos adjacentes de taxa de utilização, e sendo que o limite máximo possível selecionado é aquele em que se encontra a taxa de utilização determinada.

6. Método, de acordo com a reivindicação 5, caracterizado pelo fato de um intervalo inferior e um intervalo superior adjacentes serem delimitados um em relação ao outro por um limiar tendo um valor elevado quando o limite máximo possível selecionado é aquele associado ao intervalo inferior, e tendo um valor baixo, menor que o valor alto, quando o limite máximo possível selecionado é aquele associado ao intervalo superior.

7. Método, de acordo com qualquer uma das reivindicações de 1 a 6, caracterizado pelo fato de compreender ainda:

- calcular (306), a partir da latência, um número reservado de locais de memória para cada canal virtual, e sendo que os dados são enviados se o canal virtual utiliza um número de locais de memória estritamente inferior ao limite máximo e se a memória buffer de entrada (208) apresenta mais locais de memória livres do que locais reservados livres para todos os canais virtuais.

8. Método, de acordo com qualquer uma das reivindicações de 1 a 7, caracterizado pelo fato de os dispositivos a montante (202) e a jusante (204) pertencerem a um sistema (100) que compreende:

- nós (106),
- uma rede (110) de interconexão de nós (106) compreendendo comutadores (112) e links de comunicação (114) conectando comutadores (112) entre si e aos nós (106), com cada um dos

dispositivos a montante (202) e a jusante (204) sendo nós (106) ou comutadores (112).

9. Programa de computador descarregável a partir de uma rede de comunicação ou gravado em suporte legível ou executável por um processador, caracterizado pelo fato de compreender instruções para executar as etapas de um método (300), conforme definido em qualquer uma das reivindicações de 1 a 8, quando o referido programa for executado num computador.

10. Sistema de envio de dados, caracterizado pelo fato de compreender:

- um dispositivo a jusante (204) compreendendo uma memória buffer de entrada (208),
- um dispositivo a montante (202) concebido para enviar dados a partir de uma pluralidade de canais virtuais compartilhando a memória buffer de entrada (208), se esse canal virtual utiliza um número de locais de memória da memória buffer de entrada (208) estritamente inferior ao limite máximo atual,
- meios de medir latência de comunicação entre dispositivos a montante (202) e a jusante (204),
- meios de cálculo do limite máximo atual a partir da latência.

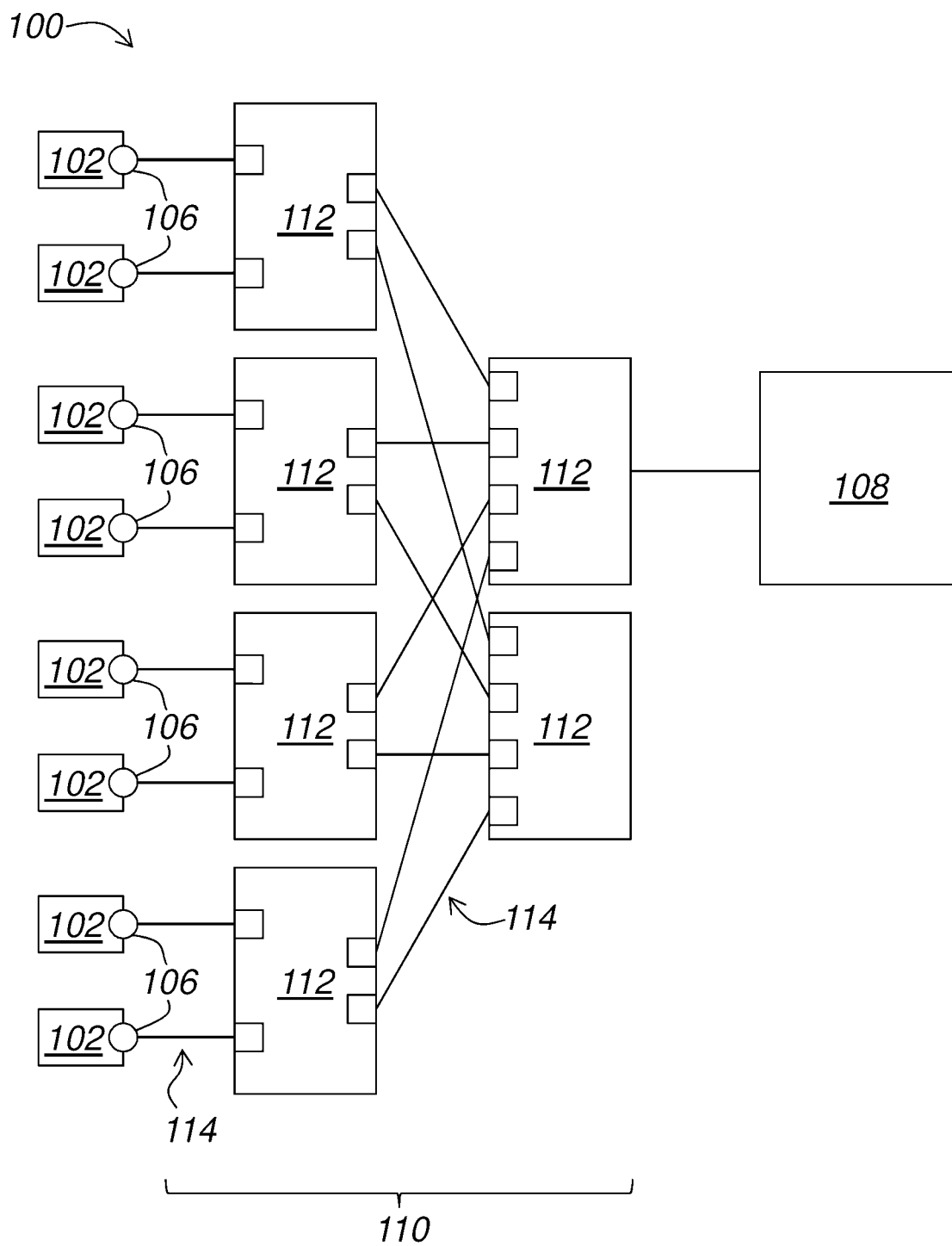


FIG.1

2/2

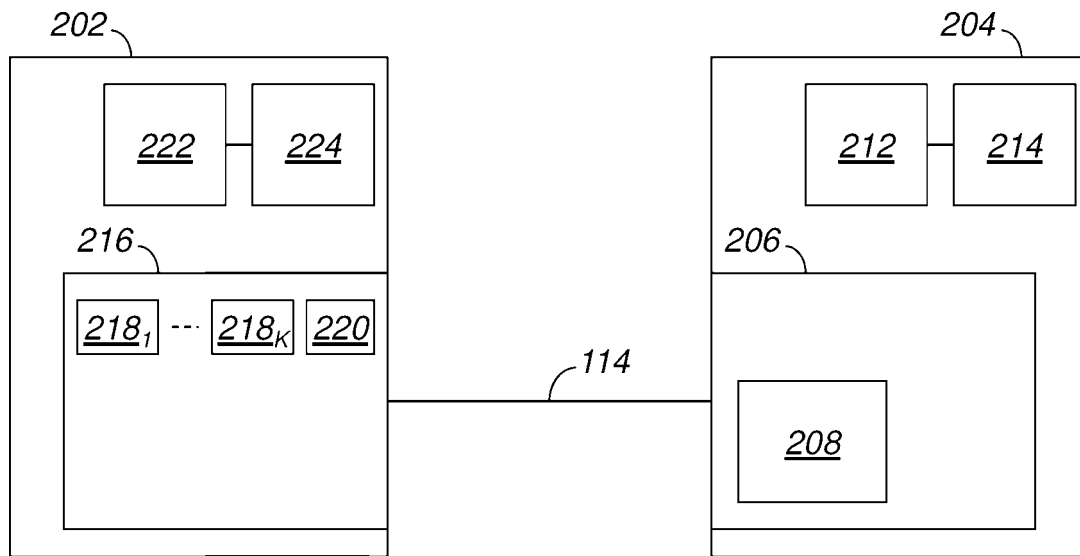


FIG. 2

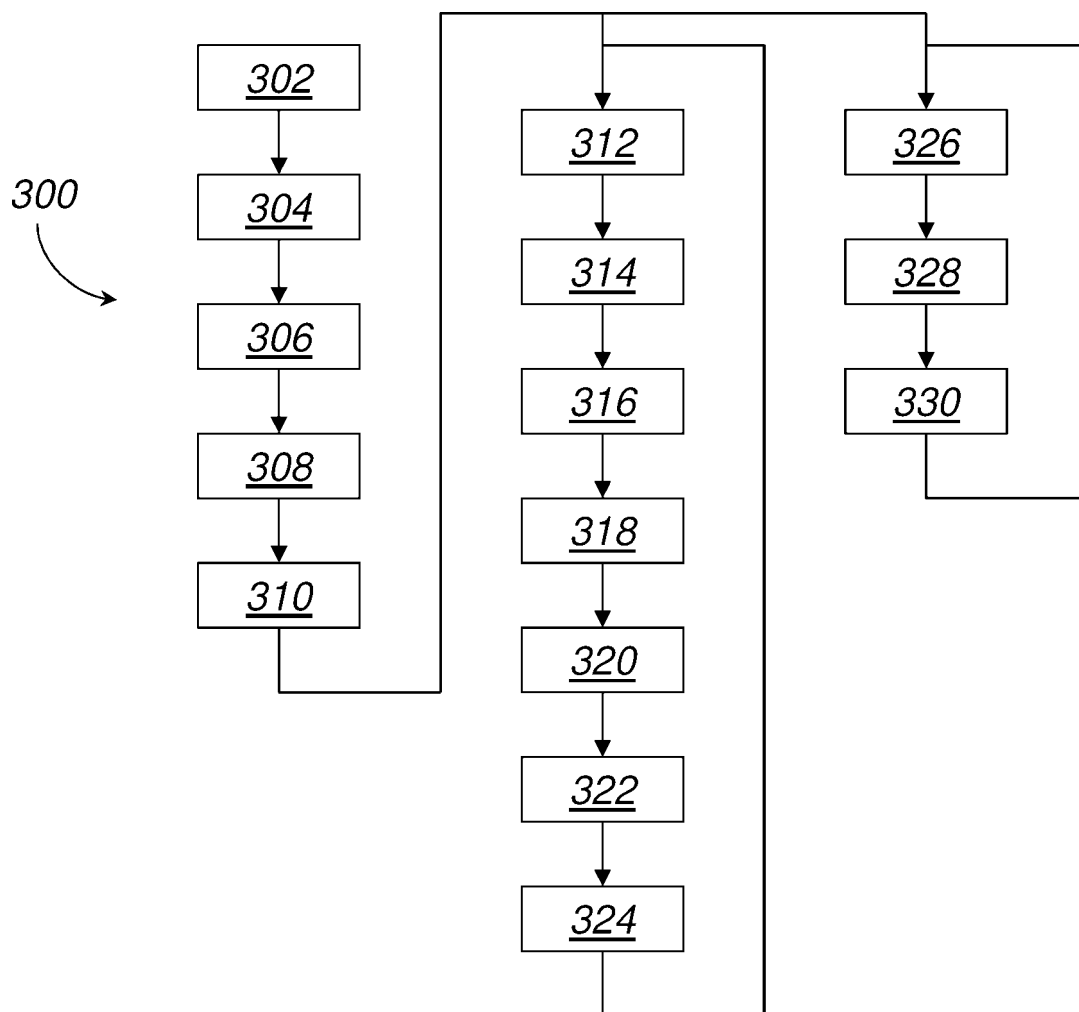


FIG. 3

RESUMO

"MÉTODO PARA ENVIO DE DADOS A PARTIR DE UM DISPOSITIVO A MONTANTE PARA UM DISPOSITIVO A JUSANTE, PROGRAMA DE COMPUTADOR DESCARREGÁVEL A PARTIR DE UMA REDE DE COMUNICAÇÃO OU GRAVADO EM SUPORTE LEGÍVEL OU EXECUTÁVEL POR UM PROCESSADOR E SISTEMA DE ENVIO DE DADOS"

Este método de envio de dados a partir de um dispositivo a montante (202) até um dispositivo a jusante (204) compreende o envio de dados a partir de uma pluralidade de canais virtuais que dividem uma mesma memória buffer de entrada (208) do dispositivo a jusante (204), se o canal virtual utiliza um número de locais de memória da memória buffer de entrada (208) estritamente inferior a um limite máximo atual. Compreende ainda a medição da latência da comunicação entre os dispositivos a montante (202) e a jusante (204) e o cálculo do limite máximo atual a partir da latência.