



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

(11) PI 0911244-8 B1



(22) Data do Depósito: 10/03/2009

(45) Data de Concessão: 06/10/2020

(54) Título: MÉTODO PARA DISTRIBUIÇÃO DE MENSAGENS, DISPOSITIVO DE BALANCEAMENTO DE CARGA E SISTEMA DE BALANCEAMENTO DE CARGA

(51) Int.Cl.: H04L 29/06.

(30) Prioridade Unionista: 14/04/2008 CN 200810027427.5.

(73) Titular(es): HUAWEI TECHNOLOGIES CO., LTD..

(72) Inventor(es): SHENGPING CHEN; YIMIN LI.

(86) Pedido PCT: PCT CN2009070711 de 10/03/2009

(87) Publicação PCT: WO 2009/127132 de 22/10/2009

(85) Data do Início da Fase Nacional: 14/10/2010

(57) Resumo: MÉTODO PARA DISTRIBUIÇÃO DE MENSAGENS, DISPOSITIVOS DE BALANCEAMENTO DE CARGA, DISPOSITIVO DE NÓ DE AGRUPAMENTO E SISTEMA DE BALANCEAMENTO DE CARGA. A presente invenção refere-se a um método, dispositivo (61, 72, 74, 81) e sistema de distribuição de mensagens. Um método de distribuição de mensagens inclui as etapas de: criar (101) uma sessão com base em um primeiro tipo recebido de mensagem de sessão e alocar um nó de agrupamento correspondente para a sessão; e ao receber um segundo tipo de mensagem de sessão associado com a sessão, distribuir (102) o segundo tipo de mensagem de sessão para o nó de agrupamento correspondendo à sessão. Com a solução técnica de acordo com as modalidades de invenção, tanto o primeiro tipo de mensagem de sessão quanto o segundo tipo de mensagem de sessão podem ser processados no mesmo nó de agrupamento, expansão de capacidade uniforme e dinâmica pode ser alcançada para o agrupamento, o desempenho de sistema pode ser aprimorado e manutenção de sistema pode ser facilitada.

Relatório Descritivo da Patente de Invenção para **"MÉTODO PARA DISTRIBUIÇÃO DE MENSAGENS, DISPOSITIVO DE BALANCEAMENTO DE CARGA E SISTEMA DE BALANCEAMENTO DE CARGA"**.

[0001] O presente pedido reivindica prioridade para o Pedido de Patente Chinês No. 200810027427.5, intitulado "MESSAGE DISTRIBUTION METHOD, DEVICE, AND SYSTEM", depositado anteriormente no Escritório de Patentes Chinês em 14 de abril de 2008, cuja totalidade está incorporada neste documento pela referência.

Campo da Invenção

[0002] A descrição refere-se ao campo de comunicações e, mais particularmente, a um método, dispositivo e sistema para distribuição de mensagens.

Antecedentes

[0003] O Protocolo de Aplicação Sem Fio (WAP) é uma combinação de uma série de especificações similares às aplicações de Internet na rede móvel, para padronizar dispositivos de comunicação sem fio para acesso à Internet, incluindo transmissão e recepção de correio eletrônico, navegação de página WAP, etc. Atualmente, a tecnologia WAP tem se tornado um padrão global para telefones móveis com base no Sistema Global para Comunicações Móveis (GSM), Serviço Geral de Rádio por Pacotes (GPRS), Acesso Múltiplo com Divisão de Código em Banda Larga (WCDMA), ou coisa parecida, e para outros terminais para acessar serviços de informação sem fio. Protocolos WAP incluem principalmente o protocolo WAP1.X e o protocolo WAP2.0.

[0004] Um mecanismo de agrupamento de balanceamento de carga é aplicado no sistema WAP corrente, onde um agrupamento de Base de Dados de Informação de Usuário (UIDB) e um agrupamento de Porta de Comunicação WAP (WAP GW) são implementados separadamente. Para uma mensagem de controle de sessão tal como uma mensagem

de solicitação de Serviços de Autenticação Remota de Usuário em Chamada Recebida (Radius) (por exemplo, mensagem de solicitação Radius), um Nó de Suporte de Porta de Comunicação GPRS (GGSN) transmite a mensagem de solicitação Radius para um certo nó de agrupamento do agrupamento UIDB por meio de um dispositivo de balanceamento de carga (Balanceamento de Carga, LB) do agrupamento UIDB. Para uma mensagem de tráfego de sessão tal como uma solicitação WAP, o GGSN transmite a solicitação WAP para um certo nó de agrupamento do agrupamento WAP GW por meio de um LB do agrupamento WAP GW. O agrupamento UIDB é independente do agrupamento WAP GW. Os dois agrupamentos têm diretrizes de balanceamento de carga separadas e não relacionadas. Quando o usuário de um aparelho de telefone tenta acessar a rede, um nó de agrupamento A do agrupamento WAP GW pesquisa um nó de agrupamento B do agrupamento UIDB para a informação em linha e fora de linha de aparelho de telefone com base no endereço de Protocolo de Internet (IP) do aparelho de telefone. De acordo com os recursos técnicos fornecidos na técnica anterior, o nó de agrupamento A determina um nó de agrupamento B no qual a informação em linha e fora de linha de aparelho de telefone está armazenada de acordo com o restante ao dividir os últimos 4 bits do endereço IP do aparelho de telefone pelo número de nós de agrupamento no agrupamento UIDB (este recurso técnico também é usado quando a informação em linha e fora de linha de aparelho de telefone está sendo armazenada).

[0005] O mecanismo na técnica anterior onde o agrupamento UIDB e o agrupamento WAP GW são implementados separadamente tem defeitos como se segue.

[0006] Com os recursos técnicos indicados acima fornecidos na técnica anterior, a informação de controle de sessão tal como a informação em linha e fora de linha de aparelho de telefone pode ser

pesquisada ou a informação de tráfego de sessão pode ser pesquisada. Entretanto, quando o número de nós de agrupamento no agrupamento UIDB ou no agrupamento WAP GW aumenta ou diminui, é difícil determinar onde a informação de controle de sessão ou informação de tráfego de sessão está armazenada. Assim, expansão de capacidade uniforme e dinâmica é impossível para cada agrupamento, o desempenho de sistema é baixo e manutenção não é conveniente.

Sumário

[0007] A invenção fornece um método, dispositivo e sistema para distribuição de mensagens em várias modalidades. A mensagem de controle de sessão e a mensagem de tráfego de sessão podem ser transmitidas para o mesmo nó de agrupamento. Armazenamento e recuperação da informação de controle de sessão e da informação de tráfego de sessão podem ser alcançados no nó de agrupamento. Expansão de capacidade uniforme e dinâmica pode ser alcançada para o agrupamento, o desempenho de sistema pode ser aprimorado e manutenção de sistema pode ser facilitada.

[0008] Uma modalidade da invenção fornece um método para distribuição de mensagens, incluindo:

criar uma sessão com base em uma mensagem de sessão tipo 1 recebida, e alocar um nó de agrupamento correspondente para a sessão; e

ao receber uma mensagem de sessão tipo 2 associada com a sessão, distribuir a mensagem de sessão tipo 2 para o nó de agrupamento correspondendo à sessão.

[0009] Desta maneira, uma modalidade da invenção fornece um dispositivo de balanceamento de carga, incluindo:

uma unidade de pré-processamento configurada para criar uma sessão com base em uma mensagem de sessão tipo 1 recebida e alocar um nó de agrupamento correspondente para a sessão; e

uma unidade de distribuição configurada para: ao receber uma mensagem de sessão tipo 2 associada com a sessão, distribuir a mensagem de sessão tipo 2 para o nó de agrupamento correspondendo à sessão.

[00010] Desta maneira, uma modalidade da invenção fornece um dispositivo de nó de agrupamento, incluindo uma Porta de Comunicação (GW) de Protocolo de Aplicação Sem Fio (WAP) e uma base de dados de informação de usuário (UIDB) correspondente.

[00011] A UIDB é configurada para receber uma mensagem de sessão de controle e armazenar a informação de sessão de controle na mensagem de sessão de controle.

[00012] A WAP GW é configurada para receber uma mensagem de sessão de tráfego, procurar pela informação de sessão de controle na UIDB com base na mensagem de sessão de tráfego, e processar a mensagem de sessão de tráfego com base na informação de controle de sessão.

[00013] Desta maneira, uma modalidade da invenção fornece um sistema de balanceamento de carga, incluindo um dispositivo de balanceamento de carga e um dispositivo de nó de agrupamento.

[00014] O dispositivo de balanceamento de carga é configurado para criar uma sessão com base em uma mensagem de sessão tipo 1 recebida, alocar um nó de agrupamento correspondente para a sessão e, ao receber uma mensagem de sessão tipo 2 associada com a sessão, distribuir a mensagem de sessão tipo 2 para o dispositivo de nó de agrupamento correspondendo à sessão.

[00015] O dispositivo de nó de agrupamento é configurado para receber uma mensagem de sessão tipo 1 e uma mensagem de sessão tipo 2, a mensagem de sessão tipo 1 sendo uma mensagem de controle de sessão e a mensagem de sessão tipo 2 sendo uma mensagem de tráfego de sessão, ou a mensagem de sessão tipo 1 sendo uma

mensagem de tráfego de sessão e a mensagem de sessão tipo 2 sendo uma mensagem de controle de sessão; armazenar a informação de controle de sessão na mensagem de controle de sessão; e processar a mensagem de tráfego de sessão com base na informação de controle de sessão.

[00016] Nas modalidades da invenção, uma sessão é criada com base em um primeiro tipo recebido de mensagem de sessão e um nó de agrupamento correspondente é alocado para a sessão; quando um segundo tipo de mensagem de sessão associado com a sessão é recebido, o segundo tipo de mensagem de sessão é distribuído para o nó de agrupamento correspondendo à sessão. Portanto, a mensagem de controle de sessão e a mensagem de tráfego de sessão podem ser transmitidas para o mesmo nó de agrupamento. Armazenamento e recuperação da informação de controle de sessão e da informação de tráfego de sessão podem ser alcançados no nó de agrupamento. Expansão de capacidade uniforme e dinâmica pode ser alcançada para o agrupamento, o desempenho de sistema pode ser aprimorado e manutenção de sistema pode ser facilitada.

Descrição Resumida dos Desenhos

[00017] A figura 1 é um fluxograma mostrando um método para distribuição de mensagens de acordo com uma modalidade da invenção;

A figura 2 é um fluxograma específico mostrando a etapa 102 do método para distribuição de mensagens mostrado na figura 1 de acordo com uma modalidade da invenção;

A figura 3 é um diagrama de blocos mostrando um dispositivo de balanceamento de carga de acordo com uma modalidade da invenção;

A figura 4 é um outro diagrama de blocos mostrando um dispositivo de balanceamento de carga de acordo com uma modalidade

da invenção;

A figura 5 é um diagrama de blocos mostrando um dispositivo de nó de agrupamento de acordo com uma modalidade da invenção;

A figura 6 é um diagrama de blocos mostrando um sistema de balanceamento de carga de acordo com uma modalidade da invenção;

A figura 7 é um diagrama de blocos mostrando um sistema de balanceamento de carga no modo proxy de acordo com uma modalidade da invenção; e

A figura 8 é um diagrama de blocos mostrando um sistema de balanceamento de carga no modo servidor de acordo com uma modalidade da invenção.

Descrição Detalhada

[00018] A descrição fornece um método para distribuição de mensagens, um método para recuperação de informação e o dispositivo e sistema correspondentes. Dois tipos diferentes de mensagens (tais como a mensagem de controle de sessão e a mensagem de tráfego de sessão) podem ser transmitidos para o mesmo nó de agrupamento. Armazenamento e recuperação da informação de controle de sessão e da informação de tráfego de sessão podem ser alcançados no mesmo nó de agrupamento. Expansão de capacidade uniforme e dinâmica pode ser alcançada para o agrupamento, o desempenho de sistema pode ser aprimorado e manutenção de sistema pode ser facilitada.

[00019] Descrições detalhadas serão feitas a seguir para as modalidades da invenção com referência aos desenhos anexos.

[00020] A figura 1 é um fluxograma mostrando um método para distribuição de mensagens de acordo com uma modalidade da invenção. Referindo-se à figura 1, o método inclui etapas como se segue.

[00021] Na etapa 101 uma sessão é criada com base em um primeiro

tipo recebido de mensagem de sessão e um nó de agrupamento correspondente é alocado para a sessão.

[00022] Quando uma sessão está sendo criada, um primeiro ID pode ser extraído como uma senha do primeiro tipo de mensagem de sessão, a sessão é criada ao usar a senha e a sessão é retida. A senha pode ser o endereço IP ou o número de um terminal transmitindo ou recebendo o primeiro tipo de mensagem de sessão. Por exemplo, quando o primeiro tipo de mensagem de sessão é recebido do GGSN, o endereço IP ou o número do terminal transmitindo o primeiro tipo de mensagem de sessão pode ser usado como a senha. Quando o primeiro tipo de mensagem de sessão é recebido de um nó de agrupamento, o endereço IP ou o número do terminal recebendo o primeiro tipo de mensagem de sessão pode ser usado como a senha.

[00023] Alocação de um nó de agrupamento para a sessão pode depender da carga, do status de operação, ou de coisa parecida do nó de agrupamento. Se o primeiro tipo de mensagem de sessão for recebido de um nó de agrupamento, o nó de agrupamento pode ser alocado para a sessão.

[00024] Na etapa 102, quando um segundo tipo de mensagem de sessão associado com a sessão é recebido, o segundo tipo de mensagem de sessão é distribuído para o nó de agrupamento correspondendo à sessão.

[00025] Durante a determinação para verificar se o segundo tipo de mensagem de sessão está associado com a sessão, um segundo ID pode ser extraído do segundo tipo de mensagem de sessão para determinar se o segundo ID é igual à senha para criar a sessão. Se o segundo ID for o mesmo que a senha, o segundo tipo de mensagem de sessão está associado com a sessão. O segundo ID pode ser o endereço IP ou o número de um terminal transmitindo ou recebendo o segundo tipo de mensagem de sessão. Quando a senha para criar a

sessão (isto é, o primeiro ID) é o número do terminal e o segundo ID é o endereço IP do terminal, a etapa 101 pode envolver adicionalmente as operações seguintes. Um terceiro ID é extraído do primeiro tipo de mensagem de sessão, e uma relação de correspondência é estabelecida entre o terceiro ID e o primeiro ID, o terceiro ID sendo o endereço IP do terminal. Durante a determinação para verificar se o segundo tipo de mensagem de sessão está associado com a sessão, a relação é primeiro pesquisada com base no segundo ID, para descobrir o número de terminal correspondendo ao segundo ID. Então, comparação é executada com base no número de terminal descoberto e na senha para criar a sessão (isto é, o número de terminal). Se eles forem os mesmos, o segundo tipo de mensagem de sessão está associado com a sessão.

[00026] Na modalidade exposta anteriormente, as mensagens de sessão (tais como uma mensagem IP) podem ser de tipos diferentes, por exemplo, podem ser divididas na mensagem de controle de sessão e na mensagem de tráfego de sessão. O primeiro tipo de mensagem de sessão é uma mensagem de controle de sessão e o segundo tipo de mensagem de sessão é uma mensagem de tráfego de sessão, ou o primeiro tipo de mensagem de sessão é uma mensagem de tráfego de sessão e o segundo tipo de mensagem de sessão é uma mensagem de controle de sessão. Por exemplo, a mensagem de controle de sessão pode ser uma mensagem de Serviços de Autenticação Remota de Usuário em Chamada Recebida (Radius) entre as mensagens de Protocolo de Datagramas de Usuário (UDP), e a mensagem de tráfego de sessão pode ser uma mensagem não Radius entre mensagem UDP ou uma mensagem de Protocolo de Controle de Transferência (TCP).

[00027] Na modalidade exposta anteriormente, uma sessão é retida, um nó de agrupamento correspondente é alocado para a sessão, e tipos diferentes de mensagens de sessão podem ser distribuídos para o nó

de agrupamento correspondendo à sessão com base na associação entre os tipos diferentes de mensagens de sessão e a sessão. Armazenamento e recuperação de tipos diferentes de informação de sessão podem ser assim alcançados no mesmo nó de agrupamento. Tipos diferentes de informação de sessão podem ser obtidos mais facilmente, e tipos diferentes de informação de sessão podem ser distribuídos em um modo balanceado. Expansão de capacidade uniforme e dinâmica pode ser alcançada para o agrupamento, o desempenho de sistema pode ser aprimorado e manutenção de sistema pode ser facilitada. Além disso, o primeiro tipo de mensagem de sessão e o segundo tipo de mensagem do mesmo usuário podem ser distribuídos para o mesmo nó de agrupamento. Assim, balanceamento de carga baseado em usuário pode ser alcançado e, conseqüentemente, controle de tráfego baseado em usuário e gerenciamento de faturamento podem ser alcançados.

[00028] Na descrição exemplar a seguir para a etapa 102, uma mensagem UDP (o número de protocolo da mensagem IP sendo 17 indica uma mensagem UDP) é usada e o endereço IP da Estação Móvel (MS) é usado como a senha da sessão. Tal como mostrado na figura 2, as etapas seguintes estão incluídas.

[00029] 201: O LB recebe uma mensagem de sessão, a mensagem de sessão sendo uma mensagem UDP.

[00030] 202: O LB determina se existe uma sessão associada com a mensagem UDP. Se não existir uma sessão como esta, a etapa 203 é executada; se existir uma sessão como esta, a etapa 204 é executada.

[00031] Primeiro, o LB analisa se o número de porta de servidor da mensagem UDP é de uma porta de servidor Radius (1813 ou 1812), em outras palavras, determina se a mensagem UDP recebida é uma mensagem Radius (um tipo de mensagem de controle de sessão, incluindo informação de controle de sessão, por exemplo, a informação

em linha e fora de linha MS ou coisa parecida). Dois casos podem ser incluídos.

[00032] Caso A: Se ela for uma porta de servidor Radius, a mensagem UDP recebida é uma mensagem Radius, isto é, uma mensagem de controle de sessão.

[00033] O LB extrai um Endereço IP de Quadro (isto é, o IP da MS) da mensagem Radius. Uma vez que o IP de fonte da mensagem Radius não é o IP da MS, mas o IP da Autenticação, Autorização e Contabilidade (AAA) ou do GGSN, portanto, o ID extraído é o Endereço IP de Quadro. O Endereço IP de Quadro na mensagem Radius é o IP da MS (o IP de fonte da mensagem não Radius entre mensagens UDP ou a mensagem TCP mencionada abaixo é o IP da MS).

[00034] Case B: Se ela não for uma porta de servidor Radius, a mensagem UDP recebida é uma mensagem não Radius entre mensagens UDP.

[00035] O LB extrai um IP de fonte da mensagem UDP. A mensagem UDP é transmitida pela MS, e assim o IP de fonte da mensagem UDP é o IP da MS.

[00036] O LB determina se existe uma sessão tendo uma senha igual ao Endereço IP de Quadro ou IP de fonte (isto é, o IP da MS) extraído. Se existir uma sessão como esta, a mensagem UDP está associada com a sessão. Se não existir uma sessão como esta, não existe sessão associada com a mensagem UDP.

[00037] 203: O endereço IP extraído da MS é usado como uma senha para criar uma sessão, e um nó de agrupamento correspondente é alocado para a sessão criada. O processo de criar a sessão também pode incluir dois casos.

[00038] Se o número de porta de servidor da mensagem UDP for de uma porta de servidor Radius, o LB usa o Endereço IP de Quadro extraído no caso A diretamente como uma senha para criar uma sessão

e retém a sessão. Um período de tempo limite pode ser estabelecido para a sessão, e um nó de agrupamento é selecionado em um modo balanceado, e o nó de agrupamento é alocado para a sessão criada. A mensagem Radius é transmitida para o nó de agrupamento. Deste modo, uma mensagem IP subsequente tendo o mesmo IP que a MS pode ser transmitida para este nó de agrupamento durante a sessão. Dependendo do período de tempo limite da sessão, quando a sessão atual expira, uma nova sessão pode ser recriada para a mensagem IP subsequente neste modo e um novo nó de agrupamento pode ser resselecionado.

[00039] Se o número de porta de servidor da mensagem UDP não for de uma porta de servidor Radius, o LB usa o IP de fonte extraído no caso B diretamente como uma senha para criar uma sessão e retém a sessão. Um nó de agrupamento é selecionado em um modo balanceado, e o nó de agrupamento é alocado para a sessão criada. A mensagem UDP é transmitida para o nó de agrupamento. Deste modo, uma mensagem IP subsequente tendo o mesmo IP que a MS pode ser transmitida para este nó de agrupamento durante a sessão.

[00040] 204: A mensagem UDP é distribuída para o nó de agrupamento correspondendo à sessão.

[00041] Descrição é feita para modalidade exposta anteriormente ao pegar a mensagem UDP como um exemplo. Quando a mensagem de sessão recebida pelo LB é uma mensagem TCP, isto é, uma mensagem de tráfego de sessão, o processo é similar ao processo para a mensagem não Radius entre a mensagem UDP.

[00042] Em uma implementação, quando a mensagem de sessão é uma mensagem Radius, o LB pode primeiro extrair o Endereço IP de Quadro e o Número de Diretório de Assinante Internacional de Estação Móvel (MSISDN) (isto é, o ID de Estação de Chamada da mesma). Após uma relação de correspondência ser estabelecida entre o Endereço IP

de Quadro e o ID de Estação de Chamada (por exemplo, uma tabela de dispersão tendo o Endereço IP de Quadro como a senha e tendo o ID de Estação de Chamada como seu valor), o ID de Estação de Chamada pode ser adotado como a senha para criar uma sessão correspondendo a um nó de agrupamento, e a sessão é retida. Quando o LB recebe uma mensagem de sessão de um outro tipo da MS (por exemplo, a camada de aplicação não é uma mensagem Radius), o IP de fonte é extraído da mensagem de sessão. Baseado na relação de correspondência entre Endereço IP de Quadro e ID de Estação de Chamada, um ID de Estação de Chamada correspondendo ao IP de fonte é procurado, e é determinado se o ID de Estação de Chamada descoberto é igual à senha (isto é, o ID de Estação de Chamada) para criar uma sessão. Se eles forem iguais, a mensagem de sessão está associada com a sessão, e a mensagem de sessão é transmitida para o nó de agrupamento correspondendo à sessão. Se eles forem diferentes, a mensagem de sessão é descartada.

[00043] Em uma implementação, quando o LB determina que a mensagem de sessão está associada com uma sessão, o LB pode detectar adicionalmente o status de operação do nó de agrupamento correspondendo à sessão. Mediante detecção de que o nó de agrupamento correspondendo à sessão tem uma falha, uma outra sessão pode ser criada com base na mensagem de sessão e um outro nó de agrupamento sem falha pode ser alocado para a sessão. A mensagem de sessão é transmitida para o nó de agrupamento sem falha. Os detalhes podem se referir à descrição anterior e assim nenhuma repetição é feita aqui.

[00044] Em uma implementação, quando um nó de agrupamento recebe a mensagem de sessão, a informação de sessão na mensagem de sessão (tal como a informação de controle de sessão ou a informação de tráfego de sessão) é gravada localmente. Quando o LB

transmite a mensagem de sessão indicada acima para um nó de agrupamento, ele também pode transmitir a mensagem de sessão indicada acima para um nó de recuperação. A informação de sessão na mensagem de sessão é recuperada no nó de recuperação para subsequente recuperação de informação local no nó de agrupamento. Especificamente, a informação de controle de sessão pode ser a informação em linha e fora de linha de terminal.

[00045] A figura 3 é um diagrama de blocos mostrando um dispositivo de balanceamento de carga de acordo com uma modalidade da invenção. O dispositivo de balanceamento de carga é conectado a diversos dispositivos de nós de agrupamento. O dispositivo de balanceamento de carga inclui uma unidade de pré-processamento 311 e uma unidade de distribuição 312.

[00046] A unidade de pré-processamento 311 é configurada para criar uma sessão com base em um primeiro tipo recebido de mensagem de sessão e aloca um dispositivo de nó de agrupamento correspondendo à sessão. Especificamente, quando uma sessão está sendo criada, um primeiro ID pode ser extraído do primeiro tipo de mensagem de sessão como a senha, a senha é usada para criar a sessão, e a sessão é retida. Desta maneira, a unidade de pré-processamento 311 pode incluir adicionalmente um primeiro módulo de extração configurado para extrair um primeiro ID do primeiro tipo de mensagem de sessão; um módulo de criação configurado para criar a sessão ao usar o primeiro ID como uma senha e reter a sessão; e um módulo de alocação configurado para alocar um nó de agrupamento correspondente para a sessão. A senha pode ser o endereço IP ou o número de um terminal transmitindo ou recebendo o primeiro tipo de mensagem de sessão. Por exemplo, quando o primeiro tipo de mensagem de sessão é recebido do GGSN, o endereço IP ou o número do terminal transmitindo o primeiro tipo de mensagem de sessão pode

ser usado como a senha. Quando o primeiro tipo de mensagem de sessão é recebido de um dispositivo de nó de agrupamento, o endereço IP ou o número do terminal recebendo o primeiro tipo de mensagem de sessão pode ser usado como a senha. Alocação de um nó de agrupamento para a sessão pelo módulo de alocação pode depender da carga, do status de operação ou de coisa parecida do dispositivo de nó de agrupamento. Se o primeiro tipo recebido de mensagem de sessão for transmitido de um dispositivo de nó de agrupamento, o dispositivo de nó de agrupamento pode ser alocado para a sessão.

[00047] A unidade de distribuição 312 é configurada para: ao receber um segundo tipo de mensagem de sessão associado com a sessão, distribuir o segundo tipo de mensagem de sessão para o dispositivo de nó de agrupamento correspondendo à sessão. Durante a determinação para verificar se o segundo tipo de mensagem de sessão está associado com a sessão, um segundo ID pode ser extraído do segundo tipo de mensagem de sessão para determinar se o segundo ID é igual à senha para criar a sessão. Se o segundo ID for igual à senha, o segundo tipo de mensagem de sessão está associado com a sessão. Desta maneira, a unidade de distribuição 312 pode incluir adicionalmente um módulo de recepção configurado para receber um segundo tipo de mensagem de sessão; um segundo módulo de extração configurado para extrair um segundo ID do segundo tipo de mensagem de sessão; um módulo de determinação configurado para determinar se o segundo ID é igual à senha para criar a sessão, o segundo tipo de mensagem de sessão sendo associado com a sessão se o segundo ID for igual à senha; e um módulo de transmissão configurado para distribuir o segundo tipo de mensagem de sessão associado com a sessão para o nó de agrupamento correspondendo à sessão, quando o módulo de determinação determina que o segundo tipo de mensagem de sessão está associado com a sessão. Aqui, o segundo ID pode ser o endereço

IP ou o número de um terminal transmitindo ou recebendo o segundo tipo de mensagem de sessão. Na modalidade exposta anteriormente, tanto o primeiro ID quanto o segundo ID são o endereço IP do terminal ou o número do terminal.

[00048] Quando a senha para criar a sessão (isto é, o primeiro ID) é o número do terminal e o segundo ID é o endereço IP do terminal, a unidade de pré-processamento 311 também deve extrair um terceiro ID do primeiro tipo de sessão, e estabelecer uma relação de correspondência entre o terceiro ID e o primeiro ID, o terceiro ID sendo o endereço IP do terminal. Durante a determinação para verificar se o segundo tipo de mensagem de sessão está associado com a sessão, a relação é primeiro pesquisada com base no segundo ID, para descobrir o número de terminal correspondendo ao segundo ID. Então, comparação é executada com base no número de terminal encontrado e na senha para criar a sessão (isto é, o número de terminal). Se eles forem iguais, o segundo tipo de mensagem de sessão está associado com a sessão. Portanto, o diagrama de blocos do dispositivo de balanceamento de carga nesta modalidade está mostrado na figura 4. A unidade de pré-processamento também pode incluir um terceiro módulo de extração, configurado para extrair um terceiro ID do primeiro tipo de mensagem de sessão, o terceiro ID sendo o endereço IP de um terminal transmitindo ou recebendo o primeiro tipo de mensagem de sessão; e um módulo de relação configurado para criar uma relação de correspondência entre o endereço IP do terminal transmitindo ou recebendo o primeiro tipo de mensagem e o número de terminal. A unidade de distribuição pode incluir adicionalmente um módulo de recepção configurado para receber um segundo tipo de mensagem de sessão; um segundo módulo de extração configurado para extrair um segundo ID do segundo tipo de mensagem de sessão; um módulo de pesquisa configurado para: com base na relação no módulo de

correspondência, procurar por um número de terminal correspondendo ao segundo ID; um módulo de determinação configurado para determinar se o número de terminal encontrado é igual à senha para criar a sessão, o segundo tipo de mensagem de sessão sendo associado com a sessão se o número de terminal for igual à senha; e um módulo de transmissão configurado para distribuir o segundo tipo de mensagem de sessão associado com a sessão para o nó de agrupamento correspondendo à sessão, quando o módulo de determinação determina que o segundo tipo de mensagem de sessão está associado com a sessão.

[00049] Na modalidade exposta anteriormente, as mensagens de sessão (tais como uma mensagem IP) podem ser de tipos diferentes, por exemplo, podem ser divididas na mensagem de controle de sessão e na mensagem de tráfego de sessão. O primeiro tipo de mensagem de sessão é uma mensagem de controle de sessão e o segundo tipo de mensagem de sessão é uma mensagem de tráfego de sessão, ou o primeiro tipo de mensagem de sessão é uma mensagem de tráfego de sessão e o segundo tipo de mensagem de sessão é uma mensagem de controle de sessão.

[00050] Na modalidade exposta anteriormente, a unidade de pré-processamento 311 retém uma sessão e aloca um dispositivo de nó de agrupamento correspondendo à sessão. Então, a unidade de distribuição 312 pode distribuir tipos diferentes de mensagens de sessão para o dispositivo de nó de agrupamento correspondendo à sessão com base na associação entre os tipos diferentes de mensagens de sessão e a sessão. Armazenamento e recuperação de tipos diferentes de informação de sessão podem ser alcançados no mesmo dispositivo de nó de agrupamento. Tipos diferentes de informação de sessão podem ser obtidos facilmente, e tipos diferentes de informação de sessão podem ser distribuídos em um modo balanceado. Expansão

de capacidade uniforme e dinâmica pode ser alcançada para o agrupamento, o desempenho de sistema pode ser aprimorado e manutenção de sistema pode ser facilitada.

[00051] Descrições detalhadas serão dadas a seguir para a unidade de pré-processamento 311 e para a unidade de distribuição 312 do dispositivo de balanceamento de carga. Na descrição a seguir, por exemplo, a mensagem de sessão é uma mensagem de Protocolo de Datagramas de Usuário (UDP) (o número de protocolo da mensagem IP sendo 17 indica uma mensagem UDP) e o endereço IP da MS é a senha da sessão.

[00052] Após o módulo de recepção receber a mensagem UDP, o segundo módulo de extração extrai um segundo ID da mensagem UDP. Quando o número de porta de servidor da mensagem UDP é de uma porta de servidor Radius (isto é, a mensagem UDP é uma mensagem Radius), o segundo módulo de extração extrai um Endereço IP de Quadro (isto é, o IP da MS) dela. Quando o número de porta de servidor da mensagem UDP não é de uma porta de servidor Radius (isto é, a mensagem UDP é uma mensagem não Radius), o segundo módulo de extração extrai um IP de fonte (isto é, o IP da MS) dela. Depois, o módulo de determinação determina se existe uma sessão tendo uma senha igual ao Endereço IP de Quadro ou IP de fonte (isto é, o IP da MS) extraído. Se existir uma sessão como esta, a mensagem UDP está associada com a sessão, e o módulo de transmissão distribui a mensagem UDP para o dispositivo de nó de agrupamento correspondendo à sessão. Se não existir uma sessão como esta, não existe sessão associada com a mensagem UDP, e o primeiro módulo de extração extrai o Endereço IP de Quadro ou o IP de fonte (isto é, o IP da MS) da mensagem UDP com base no tipo da mensagem UDP. O módulo de criação usa o endereço IP extraído da MS como uma senha para criar uma sessão, e um dispositivo de nó de agrupamento

correspondente é alocado para a sessão criada. Um período de tempo limite pode ser estabelecido para a sessão. Quando a sessão atual expira, uma nova sessão pode ser recriada para uma mensagem IP subsequente neste modo e um novo dispositivo de nó de agrupamento pode ser resselecionado.

[00053] Além disso, quando a mensagem de sessão recebida pelo dispositivo de balanceamento de carga é uma mensagem TCP (isto é, uma mensagem de tráfego de sessão), o processo é similar ao processo para uma mensagem não Radius entre mensagem UDP.

[00054] Em uma modalidade, quando o módulo de criação cria a sessão ao usar o número da MS como a senha e retém a sessão (isto é, o primeiro ID extraído pelo primeiro módulo de extração é o ID de Estação de Chamada na mensagem de sessão), o terceiro módulo de extração extrai o Endereço IP de Quadro da mensagem de sessão. O módulo de relação estabelece uma relação de correspondência entre o Endereço IP de Quadro e o ID de Estação de Chamada (por exemplo, uma tabela de dispersão tendo o Endereço IP de Quadro como a senha e tendo o ID de Estação de Chamada como seu valor). Quando o módulo de recepção recebe uma mensagem de sessão de um outro tipo da MS (por exemplo, a camada de aplicação não é uma mensagem Radius), o segundo módulo de extração extrai o IP de fonte da mensagem de sessão. Com base na relação entre Endereço IP de Quadro e ID de Estação de Chamada, o módulo de pesquisa procura por um ID de Estação de Chamada correspondendo ao IP de fonte, e o módulo de determinação determina se o ID de Estação de Chamada descoberto pelo módulo de pesquisa é igual à senha (isto é, o ID de Estação de Chamada) para criar uma sessão. Se eles forem iguais, a mensagem de sessão está associada com a sessão, e o módulo de transmissão transmite a mensagem de sessão para o dispositivo de nó de agrupamento correspondendo à sessão.

[00055] Em uma modalidade, o dispositivo de balanceamento de carga também pode incluir uma unidade de detecção de falha. Quando o módulo de determinação indicado acima determina que a mensagem de sessão está associada com uma sessão, a unidade de detecção de falha detecta o status de operação do dispositivo de nó de agrupamento correspondendo à sessão. Mediante detecção de que o dispositivo de nó de agrupamento correspondendo à sessão tem uma falha, a unidade de pré-processamento 311 pode ser disparada para criar uma outra sessão com base na mensagem de sessão e alocar um outro dispositivo de nó de agrupamento sem falha para a sessão. Detalhes podem se referir à descrição anterior e assim nenhuma repetição é feita aqui.

[00056] Em uma modalidade, o dispositivo de balanceamento de carga também pode incluir uma unidade de recuperação. A mensagem de sessão indicada acima é transmitida para um dispositivo de nó de recuperação, e a informação de controle de sessão ou a informação de tráfego de sessão na mensagem de sessão é recuperada no dispositivo de nó de recuperação para subsequente recuperação de informação. Especificamente, a informação de controle de sessão pode ser a informação em linha e fora de linha de terminal.

[00057] A figura 5 é um diagrama de blocos mostrando um dispositivo de nó de agrupamento de acordo com uma modalidade da invenção. Referindo-se à figura, o dispositivo de nó de agrupamento inclui uma base de dados de informação de usuário (UIDB) 51 e uma WAP GW correspondente 52.

[00058] A UIDB 51 executa um processo de controle de sessão em uma sessão e armazena a informação de controle de sessão para a sessão, tal como a informação em linha e fora de linha de terminal. A WAP GW 52 executa um processo de tráfego de sessão na mesma sessão processada pela UIDB 51. Quando a WAP GW 52 pesquisa para a informação de controle de sessão da sessão, a WAP GW 52 transmite

uma solicitação de pesquisa para a informação de controle de sessão para a UIDB 51 e recebe a informação de controle de sessão em uma resposta da UIDB 51.

[00059] Em uma modalidade, a WAP GW 52 pode incluir uma unidade de recepção configurada para receber uma mensagem de sessão de tráfego; uma unidade de pesquisa local configurada para transmitir uma solicitação de pesquisa para a informação de sessão de controle para a UIDB 51 e receber informação de controle de sessão em uma resposta da UIDB 51; e uma unidade de processamento configurada para processar a mensagem de sessão de tráfego com base na informação de sessão de controle descoberta. A WAP GW 52 também pode incluir uma unidade de pesquisa externa configurada para: quando a unidade de pesquisa local transmite uma solicitação de pesquisa para a informação de controle de sessão para a UIDB 51 e recebe uma resposta indicando que nenhuma informação de controle de sessão foi descoberta, procurar pela informação de sessão de controle da sessão recuperada em um dispositivo de nó de recuperação fora do dispositivo de nó de agrupamento.

[00060] É para ser notado que a UIDB 51 e a WAP GW 52 podem ser fornecidas separadamente em diferentes dispositivos físicos ou podem ser fornecidas no mesmo dispositivo físico.

[00061] De acordo com as modalidades expostas anteriormente, recuperação de tipos diferentes de informação de sessão pode ser alcançada no mesmo dispositivo de nó de agrupamento, de maneira que tipos diferentes de informação de sessão podem ser obtidos de forma conveniente.

[00062] A figura 6 é um diagrama de blocos mostrando um sistema de balanceamento de carga de acordo com uma modalidade da invenção. Referindo-se à figura, o sistema inclui principalmente um dispositivo de balanceamento de carga 61 e uma pluralidade dos

dispositivos de nós de agrupamento 62. A pluralidade dos dispositivos de nós de agrupamento pode formar um agrupamento.

[00063] O dispositivo de balanceamento de carga 61 é configurado para criar uma sessão com base em um primeiro tipo recebido de mensagem de sessão, alocar um dispositivo de nó de agrupamento correspondendo à sessão e, ao receber um segundo tipo de mensagem de sessão associado com a sessão, distribuir o segundo tipo de mensagem de sessão para um dispositivo de nó de agrupamento correspondendo à sessão.

[00064] O dispositivo de nó de agrupamento 62 é configurado para executar um primeiro tipo de processo de sessão na sessão com base no primeiro tipo recebido de mensagem de sessão e executar um segundo tipo de processo de sessão na sessão com base no segundo tipo recebido de mensagem de sessão. Por exemplo, quando o primeiro tipo de mensagem de sessão é uma mensagem de tráfego de sessão, o dispositivo de nó de agrupamento 62 executa um processo de tráfego de sessão com base na mensagem de tráfego de sessão recebida, e quando o segundo tipo de mensagem de sessão é uma mensagem de controle de sessão, o dispositivo de nó de agrupamento 62 executa um processo de controle de sessão com base na mensagem de controle de sessão recebida.

[00065] Descrições serão feitas a seguir para dois cenários de aplicação da modalidade exposta anteriormente da invenção.

[00066] A figura 7 é um diagrama de blocos mostrando um sistema de balanceamento de carga no modo proxy de acordo com uma modalidade da invenção. Referindo-se à figura, o sistema inclui um Cliente 71, um LB 72, um agrupamento de Porta de Comunicação de Serviço Proxy Transparente (TPSG) 73 (não limitado ao número tal como representado), um LB 74 e um Servidor de Aplicação (APPServer) 75. O LB 72, o LB 74 e o agrupamento TPSG 73 são transparentes para

o Cliente 71 e para o Servidor de Aplicação 75. O Cliente 71 é conectado ao LB 72. O LB 72 e o LB 74 são conectados ao mesmo agrupamento 73. O LB 74 é conectado ao Servidor de Aplicação 75. A função de cada dispositivo é como se segue.

(1) O Cliente 71 transmite uma mensagem UDP para o Servidor de Aplicação 75.

[00067] Especificamente, o Cliente 71 transmite uma mensagem UDP para o LB 72.

[00068] Mediante recebimento da mensagem UDP transmitida pelo Cliente 71, o LB 72 extrai um segundo ID da mensagem UDP. Quando o número de porta de servidor da mensagem UDP é de uma porta de servidor Radius (isto é, a mensagem UDP é uma mensagem Radius), um Endereço IP de Quadro (isto é, o IP da MS) é extraído dela. Quando o número de porta de servidor da mensagem UDP não é de uma porta de servidor Radius (isto é, a mensagem UDP é uma mensagem não Radius), um IP de fonte (isto é, o IP da MS) é extraído dela. Depois, uma vez que o LB 72 tenha criado uma sessão ao pegar o IP da MS na mensagem de sessão recebida anteriormente como a senha e retido a sessão, o LB 72 determina se existe uma sessão tendo uma senha igual ao Endereço IP de Quadro ou IP de fonte (isto é, o IP da MS) extraído. Se existir uma sessão como esta, a mensagem UDP está associada com a sessão, e o LB 72 distribui a mensagem UDP para a TPSG correspondendo à sessão. Se não existir sessão associada com a mensagem UDP, o LB 72 extrai o Endereço IP de Quadro ou IP de fonte (isto é, o IP da MS) da mensagem UDP com base no tipo da mensagem UDP. O endereço IP extraído da MS é usado como uma senha para criar uma sessão, e uma TPSG correspondente é alocada para a sessão criada. Um período de tempo limite pode ser estabelecido para a sessão. Quando a sessão atual expira, uma nova sessão pode ser recriada para uma mensagem IP subsequente neste modo e uma nova

TPSG pode ser resselecionada.

[00069] Após uma TPSG no agrupamento TPSG 73 correspondendo à sessão indicada acima receber a mensagem UDP, a mensagem UDP é transmitida para o LB 74.

[00070] Mediante recebimento da mensagem UDP transmitida pela TPSG, quando a mensagem UDP é uma mensagem Radius, o LB 74 extrai o Endereço IP de Quadro da mensagem UDP (isto é, o IP do Cliente 71), extrai o endereço MAC do último salto da mensagem UDP (isto é, o endereço MAC da TPSG transmitindo a mensagem UDP), e estabelece uma relação de correspondência entre o Endereço IP de Quadro e o endereço MAC (por exemplo, uma tabela de dispersão tendo o Endereço IP de Quadro como a senha e tendo o ID de Estação de Chamada como seu valor). Depois, o LB 74 transmite a mensagem UDP para o Servidor de Aplicação 75.

[00071] O Servidor de Aplicação 75 recebe a mensagem UDP e executa um processo correspondente.

(2) Após o Servidor de Aplicação 75 receber a mensagem UDP e executar um processo correspondente, o Servidor de Aplicação 75 retorna a mensagem UDP para o Cliente 71.

[00072] Especificamente, o Servidor de Aplicação 75 retorna a mensagem UDP para o LB 74.

[00073] Mediante recebimento da mensagem UDP retornada pelo Servidor de Aplicação 75, o LB 74 extrai um segundo ID da mensagem UDP e extrai o IP de destino dela (isto é, o IP da MS). Depois, o LB 74 determina se existe uma sessão tendo uma senha igual ao IP de destino extraído. Se existir uma sessão como esta, a mensagem UDP está associada com a sessão, e então LB 74 pesquisa a relação entre o Endereço IP de Quadro e o endereço MAC, descobre o endereço MAC da TPSG correspondendo ao IP de destino, e distribui a mensagem UDP para a TPSG correspondendo ao endereço MAC (isto é, a TPSG

recebendo e transmitindo a mensagem UDP no caso (1) mencionado anteriormente). Se não existir uma sessão como esta, não existe sessão associada com a mensagem UDP e o LB 74 descarta a mensagem UDP.

[00074] Após a TPSG no agrupamento TPSG 73 correspondendo à sessão indicada acima receber a mensagem UDP, a mensagem UDP é transmitida para o LB 72.

[00075] Após o LB 72 receber a mensagem UDP transmitida pela TPSG, a mensagem UDP é transmitida para o Cliente 71 correspondendo ao IP de destino.

[00076] Na modalidade exposta anteriormente, uma sessão é retida em duas direções (a direção do Cliente 71 para o Servidor de Aplicação 75 e a direção do Servidor de Aplicação 75 para o Cliente 71), e um dispositivo de nó de agrupamento correspondente é alocado para a sessão. Com base na associação entre tipos diferentes de mensagens de sessão e a sessão, tipos diferentes de mensagens de sessão podem ser distribuídos para o dispositivo de nó de agrupamento correspondendo à sessão para envio. Processamento de tipos diferentes de informação de sessão pode ser alcançado no mesmo dispositivo de nó de agrupamento. Tipos diferentes de informação de sessão podem ser distribuídos em um modo balanceado. Expansão de capacidade uniforme e dinâmica pode ser alcançada para o agrupamento, o desempenho de sistema pode ser aprimorado e manutenção de sistema pode ser facilitada. Além disso, o primeiro tipo de mensagem de sessão e o segundo tipo de mensagem do mesmo usuário podem ser distribuídos para o mesmo nó de agrupamento. Assim, balanceamento de carga baseado em usuário pode ser alcançado, e desta maneira controle de tráfego baseado em usuário e gerenciamento de faturamento podem ser alcançados.

[00077] A figura 8 é um diagrama de blocos mostrando um sistema

de balanceamento de carga no modo servidor de acordo com uma modalidade da invenção. O sistema inclui um LB 81, os dispositivos de nó de aplicação 82 e um dispositivo de nó de recuperação 83. O LB 81 é conectado a diversos dispositivos de nó de aplicação no agrupamento de dispositivo de nó de aplicação. O LB 81 inclui uma unidade de pré-processamento 811, uma unidade de distribuição 812 e uma unidade de recuperação 813. Cada dispositivo de nó de aplicação 82 tem uma UIDB 821 para executar processo de controle de sessão na sessão e uma WAP GW 822 para executar processo de tráfego de sessão na mesma sessão processada pela UIDB 821. O número de dispositivos de nó de recuperação pode ser determinado de acordo com exigências práticas. Neste exemplo, existe um dispositivo de nó de recuperação. Referindo-se à figura 8, os dispositivos e unidades funcionais do sistema são como se segue.

[00078] A unidade de pré-processamento 811 é configurada para criar uma sessão com base em uma mensagem Radius recebida e alocar um dispositivo de nó de aplicação correspondente para a sessão. Especificamente, o IP da MS na mensagem Radius recebida anteriormente pode ser usado como uma senha para criar a sessão e a sessão pode ser retida. Após a unidade de distribuição 812 receber uma mensagem UDP encaminhada pelo GGSN, um segundo ID é extraído da mensagem UDP. Quando o número de porta de servidor da mensagem UDP é de uma porta de servidor Radius (isto é, a mensagem UDP é uma mensagem Radius), o segundo ID extraído pela unidade de distribuição 812 é um Endereço IP de Quadro (isto é, o IP da MS) disto. Quando o número de porta de servidor da mensagem UDP não é de uma porta de servidor Radius (isto é, a mensagem UDP é uma mensagem não Radius), o segundo ID extraído pela unidade de distribuição 812 é um IP de fonte (isto é, o IP da MS) disto. Depois, a unidade de distribuição 812 determina se existe uma sessão tendo uma

senha igual ao Endereço IP de Quadro ou IP de fonte (isto é, o IP da MS) extraído. Se existir uma sessão como esta, a mensagem UDP está associada com a sessão, e a unidade de distribuição 812 distribui a mensagem UDP para o dispositivo de nó de aplicação correspondendo à sessão. Entretanto, a unidade de recuperação 813 transmite a mensagem UDP para o dispositivo de nó de recuperação 83. Se não existir uma sessão como esta, não existe sessão associada com a mensagem UDP, e a unidade de pré-processamento 811 extrai o Endereço IP de Quadro ou o IP de fonte (isto é, o IP da MS) da mensagem UDP com base no tipo da mensagem UDP. O endereço IP extraído da MS é usado como uma senha para criar uma sessão, e um dispositivo de nó de aplicação correspondente é alocado para a sessão criada. Um período de tempo limite pode ser estabelecido para a sessão. Quando a sessão atual expira, uma nova sessão pode ser recriada para a mensagem IP subsequente neste modo e um novo dispositivo de nó de aplicação pode ser resselecionado.

[00079] A UIDB 821 recebe a mensagem Radius transmitida pelo LB 81 e armazena a informação de controle de sessão na mesma (por exemplo, a informação em linha e fora de linha MS). A WAP GW 822 recebe uma mensagem não Radius entre a mensagem UDP ou qualquer outra mensagem de tráfego de sessão (tal como uma mensagem TCP) transmitida pelo LB 81, e armazena a informação de tráfego de sessão na mesma.

[00080] O dispositivo de nó de recuperação 83 pode recuperar a informação de controle de sessão ou informação de tráfego de sessão na mensagem UDP.

[00081] Quando a WAP GW 822 pesquisa para a informação de controle de sessão de uma sessão, ela transmite uma solicitação de pesquisa para a informação de controle de sessão para a UIDB 821. A UIDB 821 responde à solicitação de pesquisa com a informação de

controle de sessão armazenada localmente da sessão. Quando a WAP GW 822 falha para obter a informação de controle de sessão da UIDB 821, ela pesquisa para a informação de controle de sessão da sessão recuperada no dispositivo de nó de recuperação 83.

[00082] Além disso, a UIDB 821 pode pesquisar na WAP GW 822 para a informação de tráfego de sessão em outras modalidades da invenção.

[00083] Na modalidade exposta anteriormente, uma sessão é retida, um nó de aplicação correspondente é alocado para a sessão, e tipos diferentes de mensagens de sessão podem ser distribuídos para o nó de aplicação correspondendo à sessão com base na associação entre os tipos diferentes de mensagens de sessão e a sessão. Armazenamento e recuperação de tipos diferentes de informação de sessão podem ser alcançados no mesmo nó de aplicação. Tipos diferentes de informação de sessão podem ser obtidos mais facilmente, e tipos diferentes de informação de sessão podem ser distribuídos em um modo balanceado. Expansão de capacidade uniforme e dinâmica pode ser alcançada para o agrupamento de nó de aplicação, o desempenho de sistema pode ser aprimorado e manutenção de sistema pode ser facilitada.

[00084] Deve-se notar que a portadora relacionada para a informação de controle de sessão indicada acima pode ser uma mensagem Radius ou uma mensagem Diameter. O tráfego de sessão correspondendo à informação de tráfego de sessão pode envolver aplicações IP tais como a aplicação WAP, a aplicação de mídia de transferência em fluxo contínuo, a aplicação de Voz sobre Protocolo de Internet (VoIP), ou a aplicação Não Hierárquica (P2P).

[00085] Além disso, os versados na técnica podem perceber que algumas ou todas as etapas do método de acordo com as modalidades expostas anteriormente podem ser implementadas por meio de um

programa instruindo o hardware relacionado. O programa pode ser armazenado em uma mídia de armazenamento legível por computador. Quando o programa é executado, as etapas dos métodos mencionados anteriormente de acordo com as várias modalidades podem ser envolvidas. A mídia de armazenamento pode ser um disco magnético, um disco ótico, uma Memória Somente de Leitura (ROM) ou uma Memória de Acesso Aleatório (RAM), etc.

[00086] Descrições detalhadas foram feitas anteriormente para implementações específicas da invenção. Para os versados na técnica, várias mudanças e melhoramentos podem ser feitos sem divergir do princípio da invenção, e estas mudanças e melhoramentos são pretendidos para estar incluídos no escopo da invenção.

REIVINDICAÇÕES

1. Método para distribuição de mensagens que compreende as etapas de:

criar (101) uma sessão com base em um primeiro tipo recebido de mensagem de sessão, e alocar um nó de agrupamento correspondente para a sessão;

caracterizado pelo fato de que: a etapa de criar (101) a sessão com base no primeiro tipo recebido de mensagem de sessão compreende as etapas de:

extrair um primeiro ID do primeiro tipo de mensagem de sessão;

em que o primeiro ID é um número de um terminal transmitindo ou recebendo o primeiro tipo de mensagem de sessão; e

criar a sessão ao usar o primeiro ID como uma senha;

reter a sessão;

extrair um terceiro ID do primeiro tipo de mensagem de sessão, o terceiro ID sendo o endereço IP de um terminal transmitindo ou recebendo o primeiro tipo de mensagem de sessão, e criar uma relação de correspondência entre o endereço IP e o número de terminal; e

o método ainda compreendendo as etapas de:

ao receber um segundo tipo de mensagem de sessão, extrair um segundo ID do segundo tipo de mensagem de sessão, o segundo ID sendo o endereço IP de um terminal transmitindo ou recebendo o segundo tipo de mensagem de sessão;

com base na relação de correspondência entre o endereço IP e o número de terminal, procurar por um número de terminal correspondendo ao segundo ID;

determinar se o número de terminal encontrado é o mesmo que a senha para criar a sessão; e

se o número de terminal encontrado for o mesmo que a senha, o segundo tipo de mensagem de sessão sendo associado com a sessão, distribuir (102) o segundo tipo de mensagem de sessão para o nó de agrupamento correspondendo à sessão.

2. Método, de acordo com a reivindicação 1, **caracterizado pelo fato de que** ainda compreende:

detectar se o nó de agrupamento tem uma falha; e

se o nó de agrupamento tiver uma falha, criar uma outra sessão com base no segundo tipo de mensagem de sessão e alocar um outro nó de agrupamento sem falha correspondente para a sessão.

3. Método, de acordo com a reivindicação 1 ou 2, **caracterizado pelo fato de que** o primeiro tipo de mensagem de sessão é uma mensagem de controle de sessão e o segundo tipo de mensagem de sessão é uma mensagem de tráfego de sessão, ou o primeiro tipo de mensagem de sessão é uma mensagem de tráfego de sessão e o segundo tipo de mensagem de sessão é uma mensagem de controle de sessão.

4. Dispositivo de balanceamento de carga (61,72,74,81) que compreende:

uma unidade de pré-processamento (311), configurada para criar uma sessão com base em um primeiro tipo recebido de mensagem de sessão e alocar um nó de agrupamento correspondente para a sessão; e

uma unidade de distribuição (312) configurada para: ao receber um segundo tipo de mensagem de sessão associado com a sessão, distribuir o segundo tipo de mensagem de sessão para o nó de agrupamento correspondendo à sessão; **caracterizado pelo fato de que:**

a unidade de pré-processamento (311) compreende:

um primeiro módulo de extração, configurado para extrair um

primeiro ID do primeiro tipo de mensagem de sessão, o primeiro ID sendo o número de um terminal transmitindo ou recebendo o primeiro tipo de mensagem de sessão;

um módulo de criação, configurado para criar a sessão ao usar o primeiro ID como uma senha, e reter a sessão;

um módulo de alocação, configurado para alocar um nó de agrupamento correspondente para a sessão;

um terceiro módulo de extração, configurado para extrair um terceiro ID do primeiro tipo de mensagem de sessão, o terceiro ID sendo o endereço IP de um terminal transmitindo ou recebendo o primeiro tipo de mensagem de sessão; e

um módulo de relação, configurado para criar uma relação de correspondência entre o endereço IP do terminal transmitindo ou recebendo o primeiro tipo de mensagem e o número de terminal;

em que a unidade de distribuição (312) compreende:

um módulo de recepção, configurado para receber um segundo tipo de mensagem de sessão;

um segundo módulo de extração, configurado para extrair o segundo ID do segundo tipo de mensagem de sessão, o segundo ID sendo o endereço IP de um terminal transmitindo ou recebendo o segundo tipo de mensagem de sessão;

um módulo de pesquisa, configurado para procurar por um número de terminal correspondendo ao segundo ID, com base na relação de correspondência no módulo de relação,;

um módulo de determinação, configurado para determinar se o número de terminal é o mesmo que a senha para criar a sessão, o segundo tipo de mensagem de sessão sendo associado com a sessão se o número de terminal for o mesmo que a senha; e

um módulo de transmissão, configurado para distribuir o segundo tipo de mensagem de sessão associado com a sessão para o

nó de agrupamento correspondendo à sessão quando o módulo de determinação determina que o segundo tipo de mensagem de sessão está associado com a sessão.

5. Dispositivo (61,72,74,81), de acordo com a reivindicação 4, **caracterizado pelo fato de que** ainda compreende:

uma unidade de detecção de falha, configurada para detectar se o nó de agrupamento tem uma falha e, se o nó de agrupamento tiver uma falha, disparar a unidade de pré-processamento (311) para criar uma outra sessão com base no segundo tipo de mensagem de sessão e alocar um outro nó de agrupamento sem falha correspondente para a sessão.

6. Dispositivo (61,72,74,81), de acordo com a reivindicação 4 ou 5, **caracterizado pelo fato de que** o primeiro tipo de mensagem de sessão é uma mensagem de controle de sessão e o segundo tipo de mensagem de sessão é uma mensagem de tráfego de sessão, ou o primeiro tipo de mensagem de sessão é uma mensagem de tráfego de sessão e o segundo tipo de mensagem de sessão é uma mensagem de controle de sessão.

7. Sistema de balanceamento de carga **caracterizado pelo fato de que** compreende um dispositivo de balanceamento de carga (61,72,74,81) conforme definido em qualquer uma das reivindicações 4 a 6 e um dispositivo de nó de agrupamento (62), em que:

o dispositivo de nó de agrupamento (62) é configurado para: receber um primeiro tipo de mensagem de sessão e um segundo tipo de mensagem de sessão, em que o primeiro tipo de mensagem de sessão é uma mensagem de controle de sessão e o segundo tipo de mensagem de sessão é uma mensagem de tráfego de sessão, ou o primeiro tipo de mensagem de sessão é uma mensagem de tráfego de sessão e o segundo tipo de mensagem de sessão é uma mensagem de controle de sessão;

armazenar a informação de controle de sessão na mensagem de controle de sessão; e

processar a mensagem de tráfego de sessão com base na informação de controle de sessão.

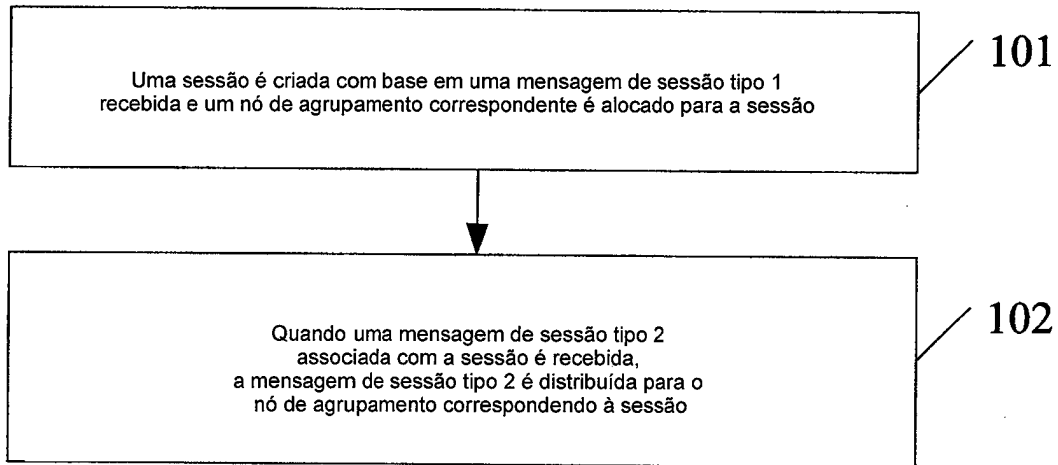


FIG.1

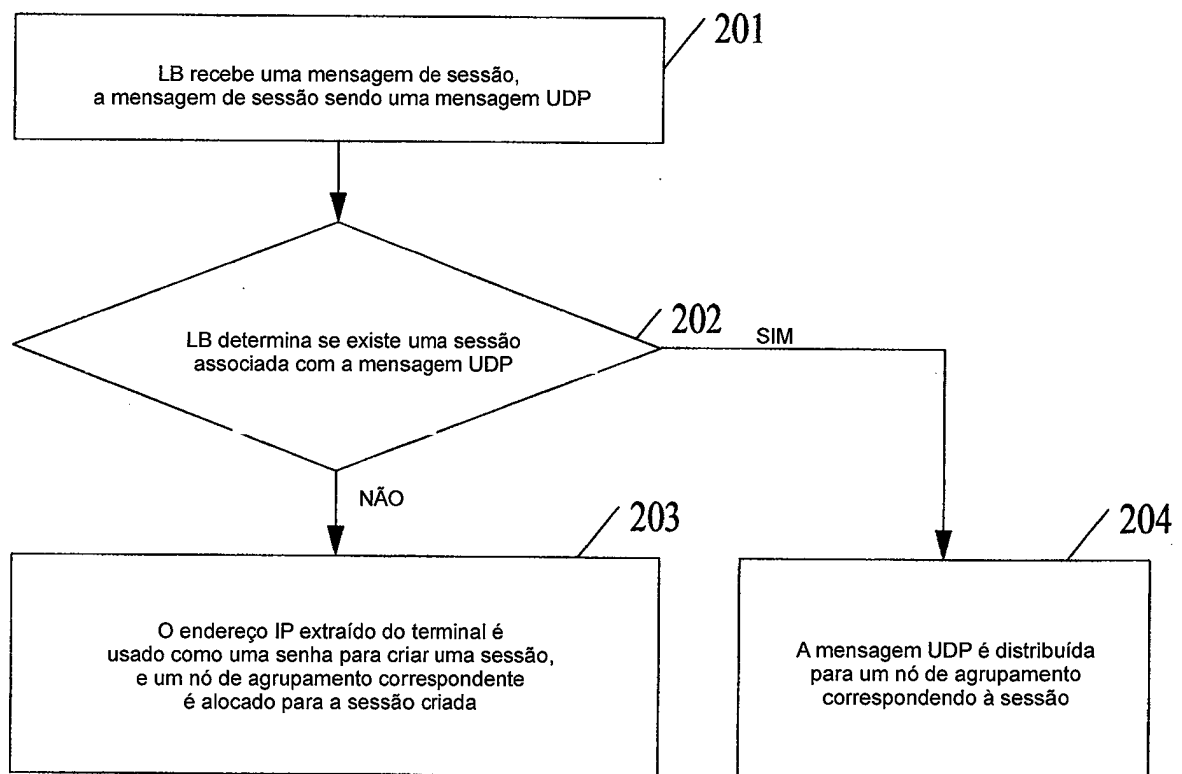


FIG.2

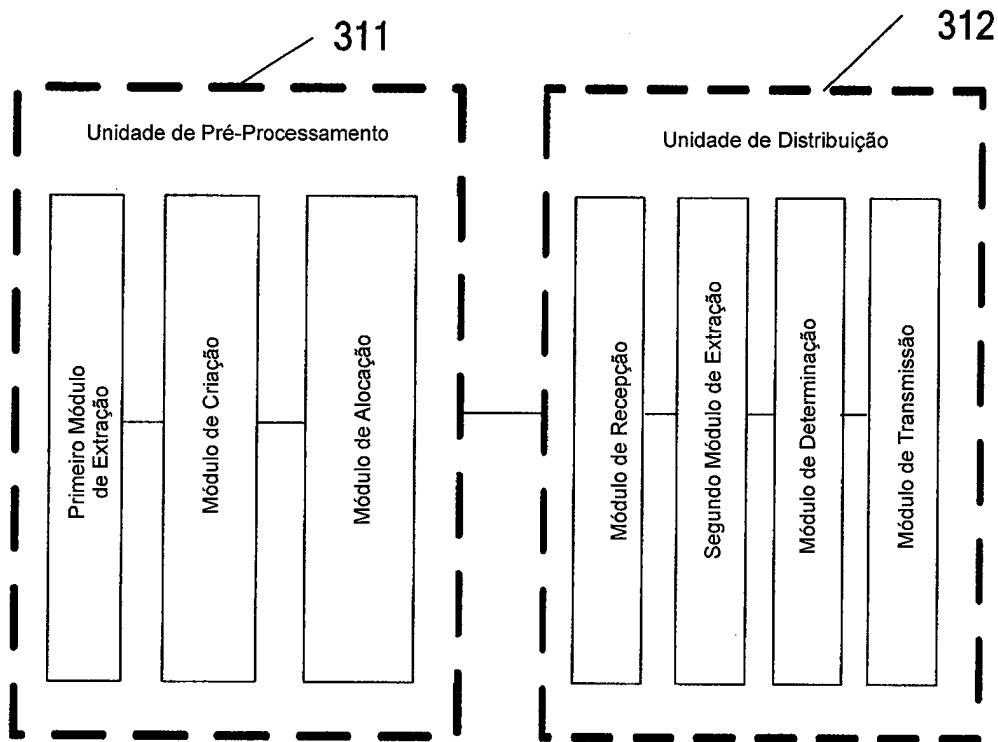


FIG. 3

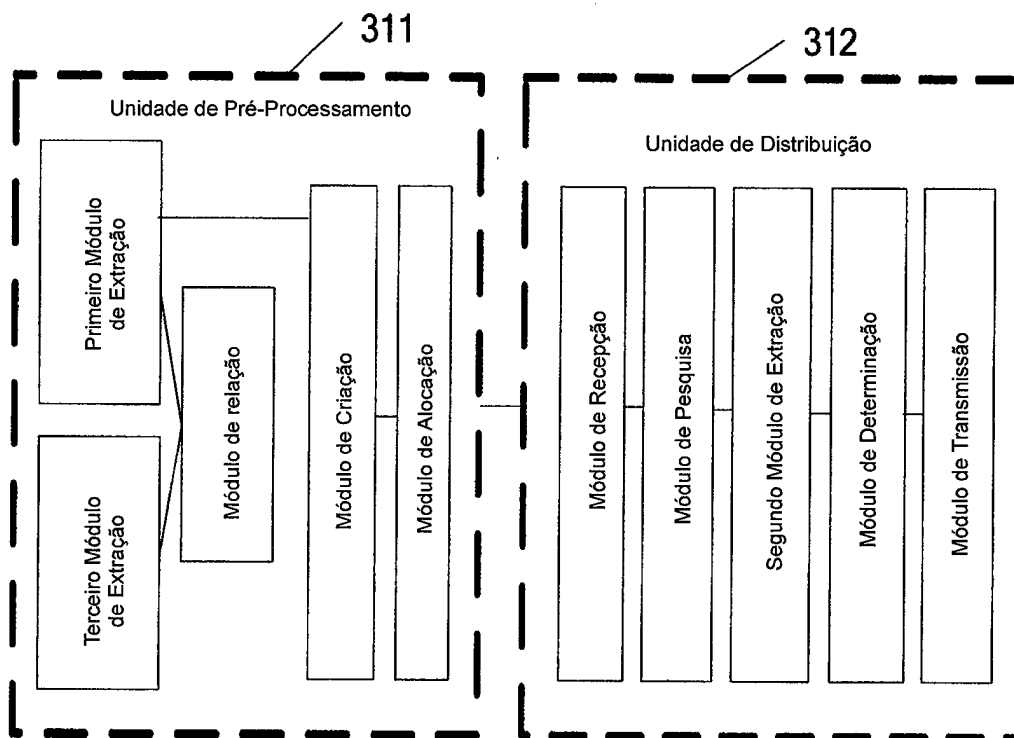


FIG. 4

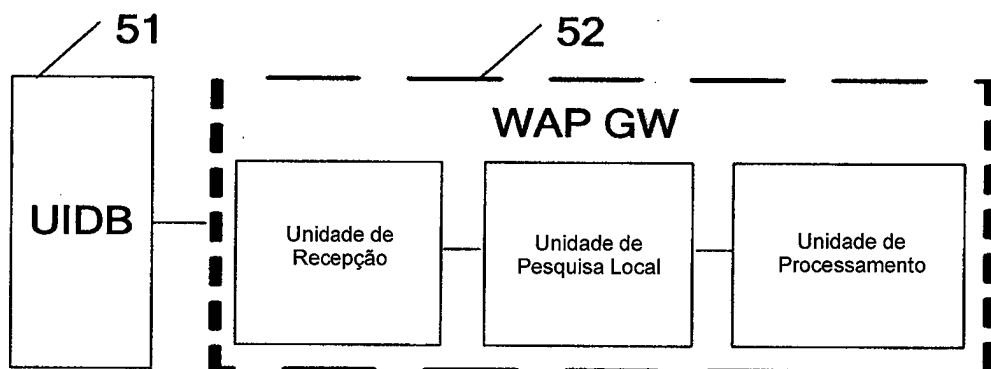


FIG.5

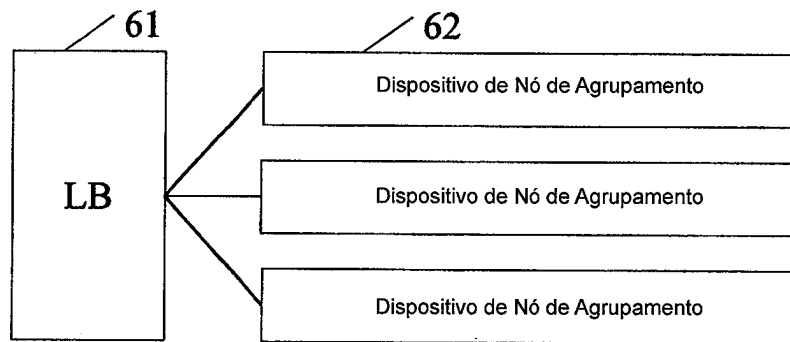


FIG.6

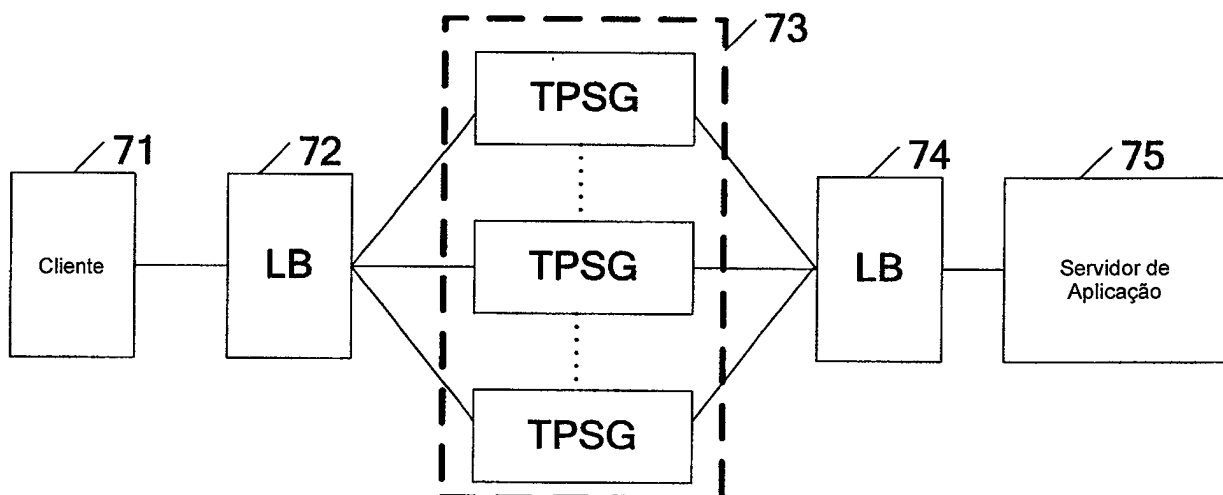


FIG.7

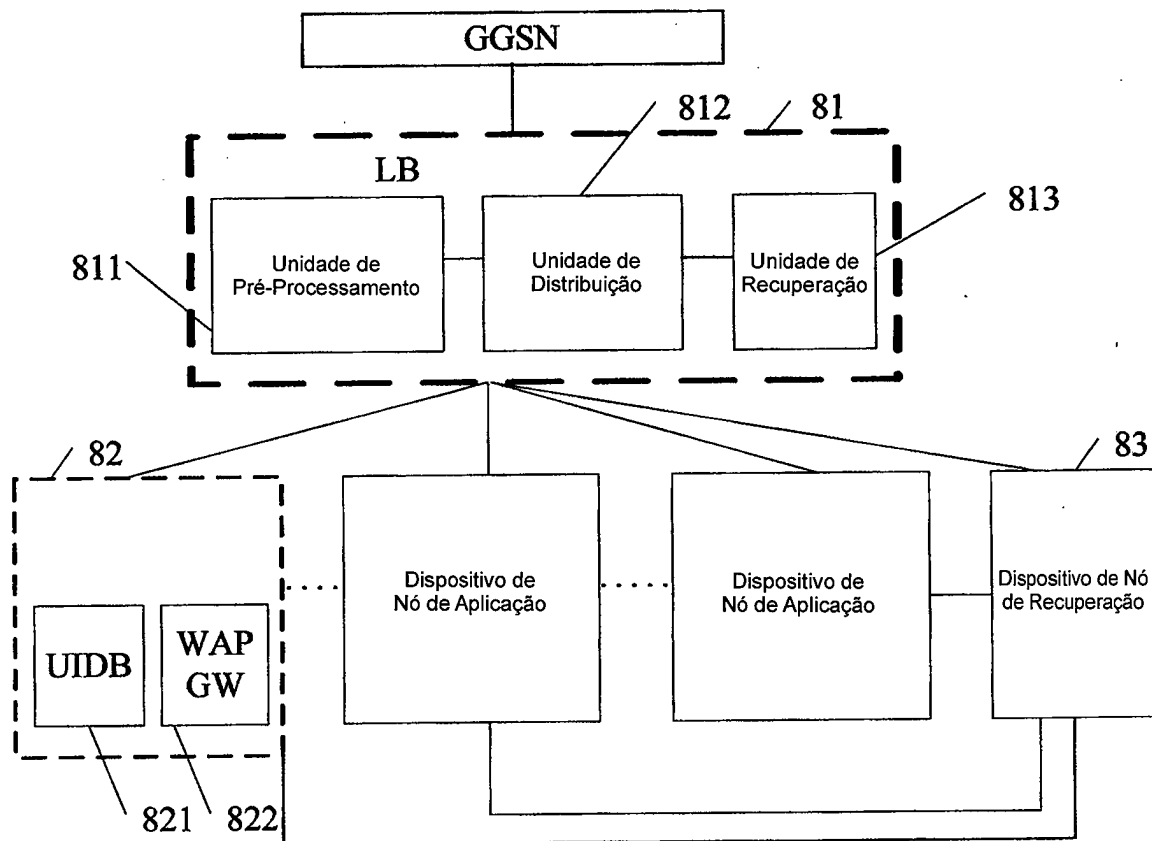


FIG.8