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

(21) PI 1013205-8 A2

BRPI1013205A2

(22) Data do Depósito: 05/05/2010

(43) Data da Publicação Nacional: 02/06/2020

(54) Título: MÉTODO E EQUIPAMENTO PARA CLASSIFICAÇÃO DE URLS

(51) Int. Cl.: H04L 29/06; G06F 17/30; H04L 29/08.

(71) Depositante(es): F-SECURE CORPORATION.

(72) Inventor(es): KIMMO MUSTONEN.

(86) Pedido PCT: PCT EP2010056107 de 05/05/2010

(87) Publicação PCT: WO 2010/128082 de 11/11/2010

(85) Data da Fase Nacional: 08/11/2011

(57) Resumo: MÉTODO E EQUIPAMENTO PARA CLASSIFICAÇÃO DE URLs Método para fornecer informação de classificação em relação aos Identificadores de Recursos Uniformes para um terminal de cliente. O método contempla a identificação de um Identificador de Recurso Uniforme no terminal do cliente, enviando uma primeira consulta para um servidor de classificação através de uma rede de IP, a consulta é incluída como uma sequência de consulta de primeiro componente do Identificador de Recurso Uniforme identificado, ou uma derivativa deste primeiro componente, e recebe a primeira consulta no servidor de classificação que determina se existe ou não uma classificação para a sequência de consulta. Uma resposta é enviada pelo servidor para o terminal do cliente, a resposta inclui uma classificação determinada ou uma indicação de que não existe classificação. A resposta é recebida no terminal do cliente e, se a classificação incluída na resposta assim indicar ou, por outro lado, se a resposta assim indicar, então uma consulta adicional é enviada para o servidor de classificação, esta consulta adicional é incluída como uma sequência de consulta do primeiro componente e do segundo componente do Identificador de Recurso Uniforme identificado, ou uma derivativa do primeiro e segundo componentes. Os passos de recepção da primeira (...).

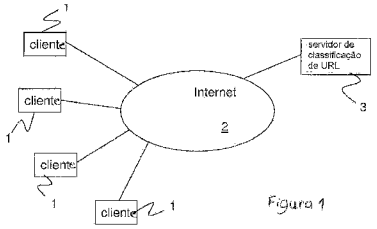

Figura 1

MÉTODO E EQUIPAMENTO PARA CLASSIFICAÇÃO DE URLs

Campo Técnico

A presente invenção refere-se a um método e equipamento para classificação de Localizadores de Recursos Uniformes (URLs). Mais particularmente, a invenção refere-se a método e equipamento para classificação de URLs em uma arquitetura de base em servidor de cliente de acordo com a multiplicidade de solicitações de classificação de URLs dos clientes de um servidor comum ou conjunto de servidores.

Antecedentes

Embora fornecedora de muitos benefícios aos indivíduos e organizações, a Internet pode ser uma fonte de perigos escondidos e conhecidos. Por um lado, a maioria do conteúdo disponível na Internet é de natureza indesejável enquanto que, por outro, páginas da web podem ser uma fonte de softwares mal-intencionados, incluindo spyware, cavalos de troia, vírus, etc. Um mecanismo de defesa contra estes problemas é realizar uma classificação dos Localizadores de Recursos Uniformes (URLs) que entram no navegador da web em um terminal de cliente e apresentar a informação de classificação ao usuário antes de fazer o download da página associada da web e/ou filtrar as solicitações de acesso dependendo dos resultados da classificação.

O US2008/0163380 descreve uma abordagem deste tipo envolvendo classificações de URL. Isto envolve fornecer um Servidor de Classificação de URL que mantenha uma base de dados de URLs conhecidos e suas classificações. Os terminais de cliente solicitam classificações para o Servidor de Classificação, por exemplo, antes de baixar uma página da web. Para reduzir os tempos de espera para o cliente, a cache de URL é construída no terminal do cliente baseada em URLs previamente

acessados e confiáveis, com obtenção de dados de classificação de URL pré-estabelecidos, por exemplo, baseados no perfil do usuário.

As abordagens conhecidas para fornecer classificação de URL usando um servidor de cliente possuem a desvantagem que a base de dados dos URLs classificados deve ser extremamente grande e é extremamente difícil mantê-la confiável dada a natureza dinâmica da Internet. O crescimento massivo dos sítios 2.0 da web em particular causa um problema significativo para os sistemas convencionais de classificação de URL uma vez que conteúdos dúbios e perigosos podem ser introduzidos na Internet e removidos em períodos de tempo muito curtos. Na verdade, as abordagens conhecidas são ineficientes em larga escala em vista da mudança dos padrões de uso vistos hoje na Internet.

Resumo

É um dos objetivos da presente invenção sanar ou, no mínimo, mitigar as desvantagens observadas acima dos métodos conhecidos de classificação de URL. Isso é atingido, pelo menos em parte, procurando identificar URLs classificados que se combinam em parte com os componentes dos URLs que os clientes procuram classificar. Estes URLs classificados podem ser fornecidos para transferir pelo menos uma parte da carga computacional, necessária para classificar um URL, de um servidor de apoio para os clientes.

De acordo com o primeiro aspecto da presente invenção foi fornecido um método de fornecimento da informação de classificação em relação aos Identificadores de Recursos Uniformes para um terminal de cliente, o método abrange:

1) a identificação de um Identificador de Recurso Uniforme no terminal do cliente;

2) o envio de uma primeira consulta a um servidor de classificação através de

uma rede de IP, consulta incluída como sequência de consulta de primeiro componente do Identificador de Recurso Uniforme identificado ou uma derivativa deste primeiro componente;

3) o recebimento da primeira consulta no servidor de classificação e determinar se existe ou não classificação para a sequência de consulta;

4) o envio de uma resposta incluindo uma classificação determinada ou uma indicação de que esta não existe para o terminal do cliente;

5) o recebimento da resposta no terminal do cliente e, se a classificação incluída na resposta assim indicar ou se a resposta, por sua vez, assim indicar, o envio de uma consulta adicional para o servidor de classificação incluída como sequência de consulta do primeiro componente e segundo componente do Identificador de Recurso Uniforme identificado ou uma derivativa do primeiro e segundo componentes;

6) a repetição dos passos 3) a 5) uma ou mais vezes conforme o necessário, adicionando para cada iteração um componente à sequência de consulta.

Modalidades da presente invenção podem, em um grande número de casos, reduzir o tempo para classificar um URL. A classificação integral de um URL pode ser necessária apenas em um número limitado de casos.

As consultas enviadas para o servidor de classificação podem conter derivativas das respectivas sequências de consultas, cada derivativa sendo obtida aplicando uma função de dispersão para a sequência de consulta correspondente.

O primeiro componente pode abranger um nome de domínio registrado ou um endereço de IP registrado. O segundo e qualquer componente adicional podem abranger subcomponentes prefixados ao primeiro componente.

O método pode envolver, no servidor de classificação, se está determinado que

existe uma classificação para a sequência de consulta e que um segundo componente ou adicional é necessário para refinar a classificação, incluindo na resposta uma definição de formato para uma próxima consulta.

O método pode envolver, sob recibo de uma resposta no terminal do cliente, a colocação de uma determinada classificação junto com a sequência de consulta associada ou derivativa na cache local. No servidor de classificação, se for determinado que exista uma classificação para a sequência de consulta e que é necessário um segundo componente ou adicional para refinar a classificação, uma definição de formato para uma próxima consulta pode ser incluída na resposta e adicionalmente colocada na cache local.

O método pode abranger, entre os passos 1) e 2), a consulta da cache local para determinar se ela contém ou não uma entrada para um componente do Identificador de Recurso Uniforme ou uma derivativa deste componente e,

se não, incluir este componente ou sua derivativa, como o primeiro componente ou sua derivativa, na primeira consulta, e,

se assim for, apresentar a classificação para um usuário e/ou construir a sequência de consulta da primeira consulta de acordo com uma definição especifica de formato.

Em uma modalidade da invenção, o terminal do cliente abrange um navegador web e o método posteriormente abrange mostrar em uma janela de navegador web uma determinada classificação. O método pode abranger os seguintes passos:

se uma classificação recebida pelo terminal do cliente indicar que o Identificador de Recurso Uniforme é confiável, então se pode baixar os dados daquele Identificador de Recurso Uniforme em uma janela de navegador web e,

se uma classificação recebida pelo terminal do cliente indicar que o

Identificador de Recurso Uniforme é malicioso, ocorre o bloqueio dos dados daquele Identificador de Recurso Uniforme em uma janela de navegador web.

O Identificador de Recurso Uniforme ou uma derivativa deste pode ser incluído na primeira consulta enviada ao servidor de classificação. Neste caso, o método pode abranger, no servidor de classificação, a determinação de se existe ou não uma classificação para o Identificador de Recurso Uniforme ou sua derivativa antes de determinar se existe ou não uma classificação para a sequência de consulta e, se existir, retornar uma resposta contendo a classificação, para o terminal do cliente.

De acordo com um segundo aspecto da presente invenção, é apresentado um método de fornecimento de informação de classificação em relação aos Identificadores de Recursos Uniformes para o terminal do cliente. O método abrange receber, no terminal do cliente, a partir de um servidor na rede, classificações em relação aos Identificadores de Recursos Uniformes e colocar as classificações recebidas no terminal do cliente na memória cache de um cliente. O método ainda permite, no terminal do cliente, que o Identificador de Recurso Uniforme identificado, para o qual uma classificação é solicitada,

consultar a cache do cliente para determinar se ela contém uma classificação para um Identificador de Recurso Uniforme que combina com o primeiro componente do Identificador de Recurso Uniforme identificado,

e dependendo do resultado, estender o primeiro componente com um segundo componente do Identificador de Recurso Uniforme identificado e também conduzir uma consulta adicional da cache do cliente com o Identificador de Recurso Uniforme estendido ou enviar uma consulta para o servidor de classificação.

O Identificador de Recurso Uniforme estendido pode corresponder ao Identificador de Recurso Uniforme identificado.

O método do segundo aspecto pode abranger a extensão adicional do URL estendido dependendo dos resultados e da repetição da consulta da cache do cliente ou enviando uma consulta ao servidor de classificação.

Uma consulta enviada ao servidor de classificação pode conter o Identificador de Recurso Uniforme completo identificado.

De acordo com um terceiro aspecto da presente invenção, é fornecido um programa que faz com que o computador realize os passos de:

1) identificação do Identificador de Recurso Uniforme no terminal do cliente ;

2) envio de uma primeira consulta para um servidor de classificação através de uma rede de IP, a consulta é incluída como uma sequência de consulta para o primeiro componente do Identificador de Recurso Uniforme identificado ou uma derivativa desta primeira parte;

3) receber uma resposta no terminal do cliente e, se a classificação incluída na resposta assim indicar ou se a resposta indicar, enviar uma consulta adicional ao servidor de classificação, que inclui uma sequência de consulta do primeiro e do segundo componente do Identificador de Recurso Uniforme identificado ou uma derivativa do primeiro e segundo componentes;

4) repetir os passos 2) e 3) uma ou mais vezes conforme necessário, adicionando para cada iteração um componente à sequência de consulta;

5) dependendo da classificação final ou da ausência desta, ou permite que o terminal do cliente acesse os dados identificados pelo Identificador de Recurso Uniforme ou previne o acesso a esses dados ou requisita uma decisão do usuário.

De acordo com um quarto aspecto da presente invenção, é fornecida uma mídia de leitura que possui armazenadas instruções para que o computador possa:

1) identificar um Identificador de Recurso Uniforme no terminal do cliente ;

2) enviar uma primeira consulta para um servidor de classificação através de uma rede de IP que é incluída como uma sequência de consulta do primeiro componente do Identificador de Recurso Uniforme identificado ou uma derivativa dessa primeira parte;

3) receber uma resposta no terminal do cliente e, se uma classificação incluída na resposta assim indicar ou se a resposta indicar, enviar uma consulta adicional ao servidor de classificação, a consulta adicional é incluída como uma sequência de consulta ao primeiro e segundo componentes do Identificador de Recurso Uniforme identificado ou uma derivativa do primeiro e segundo componentes;

4) repetir os passos 2) e 3) uma ou mais vezes conforme necessário, adicionando para cada iteração um componente à sequência de consulta;

5) dependendo da classificação final ou da ausência desta, ou permite que o terminal do cliente acesse os dados identificados pelo Identificador de Recurso Uniforme ou previne o acesso a esses dados ou requisita uma decisão do usuário.

De acordo com um quinto aspecto da presente invenção, é apresentado um método de fornecimento de classificações em relação aos Identificadores de Recursos Uniformes para o terminal do cliente, o método abrange:

1) manter em um servidor ou em associação com este uma cache que contenha classificações para os Identificadores de Recursos Uniformes;

2) receber no servidor, a partir do terminal do cliente, uma consulta que contenha um Identificador de Recurso Uniforme consultado;

3) identificar um primeiro componente do Identificador de Recurso Uniforme consultado e determinar se a cache contém ou não uma entrada para um Identificador de Recurso Uniforme que combine com esse primeiro componente;

4) se a cache contém uma entrada que indica que um componente do Identificador de Recurso Uniforme é requisitado a classificar o Identificador de Recurso Uniforme consultado, identificar o segundo componente e determinar se a cache contém ou não uma entrada para um Identificador de Recurso Uniforme que combine com o primeiro e segundo componentes;

5) repetir o passo 4) conforme solicitado com a inclusão de componentes até que uma classificação final seja obtida ou seja determinado que o Identificador de Recurso Uniforme consultado não pode ser classificado;

6) retornar as classificações determinadas ao terminal do cliente incluindo quaisquer classificações determinadas pelos Identificadores de Recursos Uniformes intermediários.

Breve Descrição dos Desenhos

Figura 1 ilustra esquematicamente uma arquitetura de rede através da qual um serviço de classificação de URL é fornecido;

Figura 2 mostra um diagrama de fluxo ilustrando um primeiro procedimento de classificação de URL envolvendo um servidor de classificação analisando uma URL para componentes previamente classificados;

Figura 3 ilustra uma sequência para realizar e melhorar um procedimento de classificação no qual um terminal do cliente analisa um URL e envia subconsultas para um servidor de classificação em relação aos componentes do URL;

Figura 4 ilustra esquematicamente um terminal do cliente configurado para fornecer informação de classificação URL; e

Figura 5 ilustra com mais detalhes o procedimento da Figura 3.

Descrição Detalhada

Está ilustrada na Figura 1 uma arquitetura típica de servidor de cliente para

fornecimento de serviço de classificação de URL. Uma multiplicidade de terminais de clientes 1 (que podem se telefones celulares, PDAs, laptops, PCs etc.) estão habilitados a acessar a Internet 2 através de algumas redes de acesso apropriadas (não mostrado na Figura). Também está acoplado na Internet um servidor de classificação ORSP FE (URL) 3 operado por um provedor de serviço terceirizado, por ex. um vendedor de antivírus e produtos de segurança e serviços. Cada terminal normalmente possui um navegador de web que permite ao usuário baixar, visualizar e interagir com páginas da web. Um navegador de web também possui uma certa nova funcionalidade para otimizar o processo de classificação e aprimorar a experiência do usuário. Essa funcionalidade será descrita.

Todo o URL é feito de alguma combinação do seguinte: um nome de esquema ou tipo de pesquisa, um domínio registrado ou endereço de Protocolo de Internet (IP), o número da porta, o nome do caminho do arquivo a ser afetado ou o programa a ser rodado, a sequência de consulta incluindo parâmetros de consulta e com arquivos html, uma âncora onde a página pode ser mostrada. A sintaxe combinada pode ser como segue:

resource_type://domain:port/filepathname?query_string#anchor.

Como foi observado acima, mesmo quando um nome de domínio ou endereço (IP) é confiável, como é o caso, por exemplo, de "youtube.com" ou "wikipedia.com", o conteúdo por trás do nome do domínio ou do endereço IP pode não ser. Também é possível que a mesma página da web possa ser apontada a um número (grande) de diferentes URLs.

É possível reduzir o número de consultas de classificação que um cliente envia para o servidor de classificação de URL identificando estes componentes de URL por trás em que todos os URLs são confiáveis (ou não confiáveis), e colocar na cache do

cliente. Quando um usuário entra com um URL em um navegador, os componentes apropriados do URL são comparados com os componentes da cache. Se for encontrada uma combinação, a classificação da cache pode ser apresentada para o usuário ou ser tomada uma ação de filtragem, sem a necessidade de que o cliente envie uma consulta ao servidor de classificação. Apenas se não for encontrada combinação uma consulta será enviada.

Seria possível construir a cache no cliente fazendo com que este envie um URL completo para o servidor de classificação no evento de uma combinação não ser encontrada. Este processo é ilustrado na Figura 2, onde (nos passos 1 a 3, 9 e 10) o terminal do cliente determina se a cache de classificação do URL local contém uma entrada para um componente de URL. Se não, o cliente envia o URL ao servidor de classificação (passo 4). O servidor de classificação poderá, então, separar o URL em vários componentes e compará-los aos conteúdos de sua base de dados (passo 5). O servidor de classificação retornará (passo 6) para o cliente um ou mais componentes juntamente com suas classificações. Os resultados são recebidos e colocados na cache pelo cliente (passo 7) e o URL será permitido ou bloqueado de acordo (passo 8). Assim, por exemplo, se um usuário entrar com o URL:

"http://www.marks-clerk.com/uk/attorneys/publications/articles.html"

e não houver combinações para o URL ou para seus componentes na cache local, o URL é enviado para o servidor de classificação de URL. Nesse caso, o servidor de classificação analisa o URL e determina que todas as URLs por trás do domínio "marks-clerk.com" são confiáveis e retorna o resultado para o cliente. Não é necessário que o servidor de classificação de URL mantenha quaisquer dados de classificação para os URLs específicos por trás do domínio marks-clerk.com. O mesmo vale para um nome de domínio malicioso, i.e., é necessário apenas manter

uma entrada para o domínio e não para todos os URLs por trás deste domínio. Para um domínio que seja confiável, mas que os URLs por trás não, por ex. "youtube.com", o servidor de classificação deverá manter classificações para estes URLs (no mínimo que seja estabelecido um nível de confiável/não-confiável).

Velocidades de pesquisa aprimoradas podem ser obtidas através do armazenamento de criptografia de componentes do URL e a geração de criptografias correspondentes da URL recebida. A cache do cliente também contém criptografias. O cliente pode, adicionalmente ao texto integral do URL, incluir na consulta um comando de criptografia dos componentes do URL até o ponto em que não tinha encontrado uma combinação na cache local. O servidor de classificação começa a busca a partir desse comando. Apenas combinações que não são previamente conhecidas ao cliente são retornadas pelo servidor de classificação.

A fim de fornecer uma verificação inicial muito rápida, o cliente pode enviar uma criptografia do URL completo para o servidor de classificação junto com o texto da URL. O servidor de classificação realiza uma rápida busca pelo resumo e, se o resultado é de que o URL é confiável ou não, isso é retornado imediatamente para o cliente. Apenas se uma combinação para o resumo não for encontrada pelo servidor de classificação, a busca realizada usando os componentes do URL.

Embora esta abordagem possa funcionar em princípio, na prática abre uma ameaça à segurança porque é necessário que o cliente envie o texto completo do URL para permitir que o servidor de classificação do URL o analise. Em algumas instâncias, um URL pode conter dados sensíveis como o nome de usuário de banco e a senha de cliente. Apesar de algumas abordagens de classificação de URL requeiram que o cliente codifique o URL enviado, por ex., aplicando uma função de hashing na URL e enviando apenas o resultado, isso não é possível na abordagem descrita no

parágrafo anterior, pois o servidor de classificação do URL ficará impossibilitado de analisar os componentes do URL quando este for previamente desconhecido do servidor de classificação. Se for usada criptografia, o servidor de classificação deve armazenar uma classificação para todo e qualquer URL(consultado).

Uma solução para esse problema é dividir a consulta do URL em um número de subconsultas. Essa abordagem é ilustrada na Figura 3, na qual o usuário entra primeiro com o URL (passo 100) na linha de endereço do navegador (ou clica em um link de web). No caso do exemplo considerado acima e novamente assumindo que o cliente, usando uma cache sua, está impossibilitado de consultar o URL, este determina (passo 101) que deve ser enviada uma consulta ao servidor de classificação contendo o domínio "marks-clerk.com". O cliente aplica um conjunto de regras neste passo para levar em conta em nível superior de hierarquia estruturas de nome de domínio como ".co.uk". Em vez de enviar a parte do nome do domínio em texto, o cliente envia apenas um hash do componente (passo 102). Com o recibo da consulta, um ORSP Front End (FE) no servidor de classificação compara (passo 103) o hash recebido com a base de dados de hashes correspondentes aos URLs consultados. Neste exemplo, o hash de "marks-clerk.com" está presente na base de dados, junto com uma classificação "confiável". O servidor de classificação responde ao cliente (passo 104) com a classificação. O cliente apresenta a classificação ao usuário e baixa a página web requisitada (passo 105). Adicionalmente, o hash da parte do nome do domínio é armazenado junto com a classificação na cache do cliente (passo 106). O usuário pode agora surfar livremente em todos os URLs por trás do domínio "marks-clerk.com" sem que nenhuma consulta adicional tenha que ser feita ao servidor de classificação do URL.

Considere agora o caso em que o usuário entra com um URL de um sítio da web de

uma comunidade (web 2.0) no navegador. A sequência seguinte ilustra essa situação:

1 . O usuário vai para http://www.youtube.com/watch?v=1234.

2. O cliente consulta a raiz "." do domínio com a cache do cliente. A cache do cliente mantém um conjunto de regras para manipular essas consultas. Neste caso, retorna uma instrução para consultar novamente com o formato "<top-level domain>".

3. O cliente consulta a cache do cliente novamente usando o nível de domínio superior ".com", e a cache do cliente retorna uma instrução para consultar novamente com o formato "<xxx>.com".

4. O cliente consulta a cache do cliente novamente usando o domínio ".youtube.com".

5. A cache do cliente retorna o resultado de que o hash do domínio não está contido dentro da cache.

6. O cliente então direciona a consulta para o servidor ORSP FE, o qual retorna a resposta de que o domínio "*.youtube.com*" é confiável por padrão, mas requisita a primeira parte do caminho".

7. O cliente direciona uma consulta para o servidor ORSP FE contendo um hash de "*.youtube.com/watch*" e recebe uma resposta de que "*.youtube.com/watch*" é confiável por padrão, mas requisita o parâmetro V ".

8. O cliente direciona uma consulta para o servidor ORSP FE contendo um hash de "*.youtube.com/watch?v=1234*", e recebe uma resposta de que "*.youtube.com/watch?v=1234*" é malicioso.

9. O cliente armazena cada resposta na cache (incluindo o hash e a classificação associada) e retorna a resposta final para o navegador plug-in

10. Subsequentemente, a cada vez que o usuário surfar em "youtube.com" ou em um vídeo em nível de domínio, apenas uma consulta simples necessita ser enviada ao servidor ORSP FE relacionada ao vídeo em questão.

Figura 4 ilustra esquematicamente um terminal do cliente 10 apropriado para o uso da abordagem descrita acima. Como exemplo, o terminal pode ser um PC, um laptop ou um telefone celular. É composto de um display 11, uma unidade de entrada do usuário 12, por exemplo, um teclado e uma interface de Internet 13. O terminal posteriormente compreende componentes de hardware e de software 14 para a implementação de um navegador web browser 15, um plug-in de navegador específico de classificação de URL 16, um cliente ORSP 17 e uma memória cache 18 disponível para o ORSP do cliente. O plug-in de navegador 16 é configurado para implementar a funcionalidade do ORSP do cliente descrita acima, enquanto que a memória 18 está configurada para armazenar a cache do cliente.

Considerando os componentes do cliente 10 com mais detalhes, o plug-in do navegador 16 representa um componente "fino" cujo papel é identificar e extrair URLs do navegador 15, passar consultas para o ORSP do cliente 17, receber respostas de classificação do ORSP do cliente e atualizar o display do navegador de acordo com as classificações recebidas (incluindo baixar/bloquear páginas da web). O ORSP do cliente 17 é responsável por analisar URLs, criptografar componentes, consultar a cache local 18 e o servidor de classificação. O ORSP do cliente pode ser um serviço fornecido na plataforma Windows™ e comunica-se com o plug-in do navegador através de uma Interface de Programação de Aplicação (API) apropriada. A cache local (cliente) é armazenada pelo ORSP do cliente 17 na memória cache 18.

Figura 5 ilustra a sinalização associada com este exemplo, onde o navegador, o

plug-in do navegador, o ORSP do cliente (Adaptador NRS-ORSP) e a cache do cliente são todos componentes de software dentro do cliente e o FE-Servidor e o FE DB (base de dados) são componentes, possivelmente distribuídos, do servidor de classificação. Neste exemplo, uma mensagem redirecionada faz com que o ORSP do cliente reconsulte tanto a cache local como o servidor de classificação.

Seria bom que qualquer um que espiasse o tráfego entre o terminal do cliente e o servidor de classificação visse apenas os componentes criptografados do URL, e não os próprios componentes. No evento de o servidor de classificação possuir uma classificação para um hash recebido, o servidor de classificação estaria habilitado a identificar o componente do URL correspondente. Neste caso, é improvável que a revelação do componente do URL seja da alçada do usuário. Componentes sensíveis do URL são muito improváveis de serem conhecidos do servidor de classificação.

Também seria apreciado pela pessoa com habilidade na tecnologia que várias modificações possam ser feitas na modalidade descrita acima sem afastar-se do escopo da presente invenção. Por exemplo, uma subconsulta inicial enviada pelo cliente ao servidor de classificação poderia incluir adicionalmente um hash do URL completo. O servidor de classificação poderia, então, realizar uma "busca rápida" inicial para esse hash, retornando um resultado imediatamente ao cliente se existir uma classificação para esse URL. Apenas se uma classificação não for encontrada, o servidor prosseguirá a realização de uma pesquisa baseada no nível de domínio.

O ORSP do cliente pode ser implementado como uma aplicação dedicada no terminal do cliente. Alternativamente, o ORSP pode ser implementado como um plug-in de navegador, applet etc.

O servidor de classificação de URL pode manter um conjunto de categorias de

classificação indicativas da confiabilidade associada com um URL classificado. Por exemplo, as seis categorias seguintes podem ser usadas:

1 Não testado [CINZA]

2 Malicioso [VERMELHO]

3 Inseguro [VERMELHO]

4 Suspeito [AMARELO]

5 Seguro [VERDE]

6 Confiável [VERDE]

A cor associada com cada categoria define a cor que é mostrada em um "semáforo" indicador de confiabilidade adicionado aos botões do navegador pelo ORSP do cliente.

Uma classificação fornecida para o cliente pelo servidor de classificação de URL pode incluir um valor de "time-to-live" (tempo de vida útil) (TTL). Isto indica para o cliente uma duração para que a classificação possa ser confiável e é armazenada na cache do cliente. Quando uma URL é inserida no navegador combina ou inclui um componente que combina com o hash pelo qual o TTL expira, o cliente repetirá o procedimento de consulta com o servidor de classificação atualizando a entrada na cache de acordo com o resultado. O cliente pode também, periodicamente, iniciar atualizações de entradas expiradas no cache.

Atualizações críticas de cache podem ser solicitadas obrigatoriamente ao cliente pelo servidor de classificação. Isto pode aplicar-se, por exemplo, a nomes de domínio maliciosos recentemente descobertos. Alternativamente, esse tipo de atualização crítica pode ser fornecido em bloco na próxima vez que um cliente fizer uma consulta comum no servidor de classificação.

Seria de muito valor que o procedimento de classificação descrito aqui possa ser

aplicado para classificar URLs associados com outras aplicações/serviços que não um navegador da web. Por exemplo, o procedimento pode ser aplicado para classificar URLs e dados da web contidos dentro de emails. Neste caso, um componente fino tipo "plug-in" pode ser implementado em um email de cliente como o Microsoft Outlook™, com o componente comunicando-se com o ORSP do cliente. O ORSP do cliente pode ser compartilhado por aplicações e serviços múltiplos.

REIVINDICAÇÕES

1. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE, caracterizado** por compreender:

1) identificar um Identificador de Recurso Uniforme no terminal do cliente;

2) enviar uma primeira consulta a um servidor de classificação através de uma rede IP, incluída como uma sequência de consulta de primeiro componente do Identificador de Recurso Uniforme identificado ou uma derivativa deste primeiro componente;

3) receber a primeira consulta no servidor de classificação e determinar se existe ou não uma classificação para a sequência de consulta;

4) enviar uma resposta incluindo uma determinada classificação ou uma indicação de que esta não existe para o terminal do cliente;

5) receber a resposta no terminal do cliente e, se uma classificação incluída na resposta indicar ou se a resposta indicar, enviar uma consulta adicional para o servidor de classificação, a consulta adicional é incluída como uma sequência de consulta do primeiro componente e do segundo componente do Identificador de Recurso Uniforme identificado ou uma derivativa do primeiro e segundo componentes;

6) repetir os passos 3) a 5) uma ou mais vezes conforme o necessário adicionando para cada iteração um componente à sequência de consulta.

2. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE**, de acordo com a reivindicação 1, **caracterizado** pelas consultas conterem derivativas das respectivas sequências de consulta, cada

derivativa obtida pela aplicação de uma função hashing à sequência de consulta correspondente.

3. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE**, de acordo com a reivindicação 1 ou 2, **caracterizado** pelo primeiro componente abranger um nome de domínio ou endereço de IP registrados.

4. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com a reivindicação 3, **caracterizado** pelo segundo e qualquer componente adicionado abranger subcomponentes adicionais pré-fixados ao primeiro componente.

5. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com qualquer uma das reivindicações precedentes, **caracterizado** por compreender, no servidor de classificação, se for determinado que exista uma classificação para a sequência de consulta e que um segundo componente ou adicional é necessário para refinar a classificação, incluir na resposta uma definição de formato para uma consulta adicional.

6. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com qualquer uma das reivindicações precedentes, **caracterizado** por compreender, ao receber uma resposta no terminal do cliente, colocar na cache local qualquer consulta determinada junto com a sequência de consulta associada ou derivativa.

7. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO**

REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE, de acordo com a reivindicação 6, **caracterizado** por compreender, no servidor de classificação, se for determinado que exista uma classificação para a sequência de consulta e que um segundo componente ou adicional é necessário para refinar a classificação, incluir na resposta uma definição de formato para uma consulta adicional e adicionalmente colocar na cache local a definição de formato.

8. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com a reivindicação 6 ou 7, **caracterizado** por compreender, entre os passos 1) e 2), consultar a cache local para determinar se ela contém ou não uma entrada para um componente do Identificador de Recurso Uniforme ou uma derivativa deste componente e, se não, incluir este componente ou sua derivativa como primeiro componente ou sua derivativa na primeira consulta, e, assim apresentar a consulta a um usuário e/ou construir a sequência de consulta da primeira consulta de acordo com uma definição de formato especificada.

9. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo terminal do cliente compreender um navegador da web e o método adicionalmente abrange mostra em uma janela de navegador da web uma classificação determinada.

10. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com a reivindicação 9, **caracterizado** por

compreender: se uma classificação recebida pelo terminal do cliente indica que o Identificador de Recurso Uniforme é confiável, os dados deste Identificador de Recurso Uniforme são baixados em uma janela de navegador da web e, se uma classificação recebida pelo terminal do cliente indicar que o Identificador de Recurso Uniforme é malicioso, bloqueia o recebimento de dados deste Identificador de Recurso Uniforme na janela de navegador da web.

11. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com qualquer uma das reivindicações precedentes, **caracterizado** por compreender a inclusão de um Identificador de Recurso Uniforme ou uma derivativa deste na primeira consulta enviada ao servidor de classificação.

12. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO REFERENTE AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com a reivindicação 11, **caracterizado** por determinar no servidor de classificação se existe ou não uma classificação para o Identificador de Recurso Uniforme ou sua derivativa antes de determinar se existe ou não uma classificação para a sequência de consulta e, se existir, retornar uma resposta contendo a classificação para o terminal do cliente.

13. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO EM RELAÇÃO AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** dito método **caracterizado** por compreender: receber, no terminal do cliente a partir de um servidor de rede, classificações em relação aos Identificadores de Recursos Uniformes; colocar as classificações recebidas no terminal do cliente dentro da memória cache; e, para um Identificador de Recurso

Uniforme identificado no terminal do cliente, para o qual uma classificação é requerida, consultar a cache do cliente para determinar se esta contém ou não uma classificação para um Identificador de Recurso Uniforme e combinar um primeiro componente do Identificador de Recurso Uniforme identificado dependendo do resultado, estender o primeiro componente com um segundo do Identificador de Recurso Uniforme identificado e também conduzir uma consulta adicional da cache do cliente com essa extensão ou enviar uma consulta ao servidor de classificação.

14. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO EM RELAÇÃO AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com a reivindicação 13, **caracterizado** pelo Identificador de Recurso Uniforme estendido corresponder ao Identificador de Recurso Uniforme identificado.

15. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO EM RELAÇÃO AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com a reivindicação 13 ou 14, **caracterizado** por abranger extensão adicional do URL estendido dependendo dos resultados adicionais e da repetição da consulta da cache do cliente ou o envio de uma consulta ao servidor de classificação.

16. **MÉTODO DE FORNECIMENTO DE INFORMAÇÃO DE CLASSIFICAÇÃO EM RELAÇÃO AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA UM TERMINAL DE CLIENTE,** de acordo com qualquer uma das reivindicações 13 a 15, **caracterizado** por uma consulta enviada ao servidor de classificação conter o Identificador de Recurso Uniforme identificado completo.

17. **PROGRAMA DE COMPUTADOR, caracterizado** por fazer com que o computador realize os passos de:

1) identificar um Identificador de Recurso Uniforme no terminal do cliente ;

2) enviar uma primeira consulta a um servidor de classificação através de uma rede de IP incluída como uma sequência de consulta de um primeiro componente do Identificador de Recurso Uniforme identificado ou uma derivativa desta primeira parte;

3) receber uma resposta no terminal do cliente e, se uma classificação incluída na resposta indicar ou se a resposta indicar, enviar uma consulta adicional ao servidor de classificação incluída como uma sequência de consulta do primeiro e segundo componentes do Identificador de Recurso Uniforme identificado ou uma derivativa destes;

4) repetir os passos 2) e 3) uma ou mais vezes, conforme o necessário, adicionando para cada iteração um componente adicional à sequência de consulta;

5) dependendo da classificação final ou da sua ausência, permite que o terminal do cliente acesse os dados identificados pelo Identificador de Recurso Uniforme ou evita o acesso a esses dados ou requisita uma decisão do usuário.

18. **MÍDIA DE REGISTRO COM INSTRUÇÕES ARMAZENADAS, caracterizada** para que o computador seja capaz de:

1) identificar um Identificador de Recurso Uniforme no terminal do cliente ;

2) enviar uma primeira consulta a um servidor de classificação através de uma rede de IP incluída como sequência de consulta de um primeiro componente do Identificador de Recurso Uniforme identificado ou uma derivativa desta primeira parte;

3) receber uma resposta no terminal do cliente e, se uma classificação incluída na resposta indicar ou se a resposta indicar, enviar uma consulta adicional ao servidor de classificação incluída como uma sequência de consulta do primeiro e do segundo

componentes do Identificador de Recurso Uniforme identificado, ou uma derivativa dos primeiro e segundo componentes;

4) repetir os passos 2) e 3) uma ou mais vezes, conforme necessário, adicionando para cada iteração um componente adicional à sequência de consulta;

5) dependendo da classificação final ou da sua ausência, permite que o terminal do cliente acesse os dados identificados pelo Identificador de Recurso Uniforme ou evita o acesso a esses dados ou requisita uma decisão do usuário

19. **MÉTODO DE FORNECIMENTO DE CLASSIFICAÇÕES EM RELAÇÃO AOS IDENTIFICADORES DE RECURSOS UNIFORMES PARA OS TERMINAIS DOS CLIENTES**, dito método **caracterizado** por compreender:

1) manter em um servidor ou em associação com este, uma cache contendo classificações para Identificadores de Recursos Uniformes;

2) receber no servidor, a partir de um terminal de cliente, uma consulta contendo um Identificador de Recurso Uniforme consultado;

3) identificar um primeiro componente do Identificador de Recurso Uniforme consultado e determinar se a cache contém ou não uma entrada para um Identificador de Recurso Uniforme combinando com o primeiro componente;

4) se a cache contém uma entrada que indica que um componente adicional do Identificador de Recurso Uniforme é necessário para classificar o Identificador de Recurso Uniforme consultado, então identificar este segundo componente e determinar se a cache contém ou não uma entrada para um Identificador de Recurso Uniforme combinando com o primeiro e segundo componentes;

5) repetir o passo 4), conforme requerido, com a inclusão de componentes adicionais até que seja obtida uma classificação final ou seja determinado que o Identificador de Recurso Uniforme consultado não pode ser classificado;

6) retornar as consultas determinadas para o terminal do cliente incluindo quaisquer classificações determinadas para Identificadores de Recursos Uniformes intermediários.

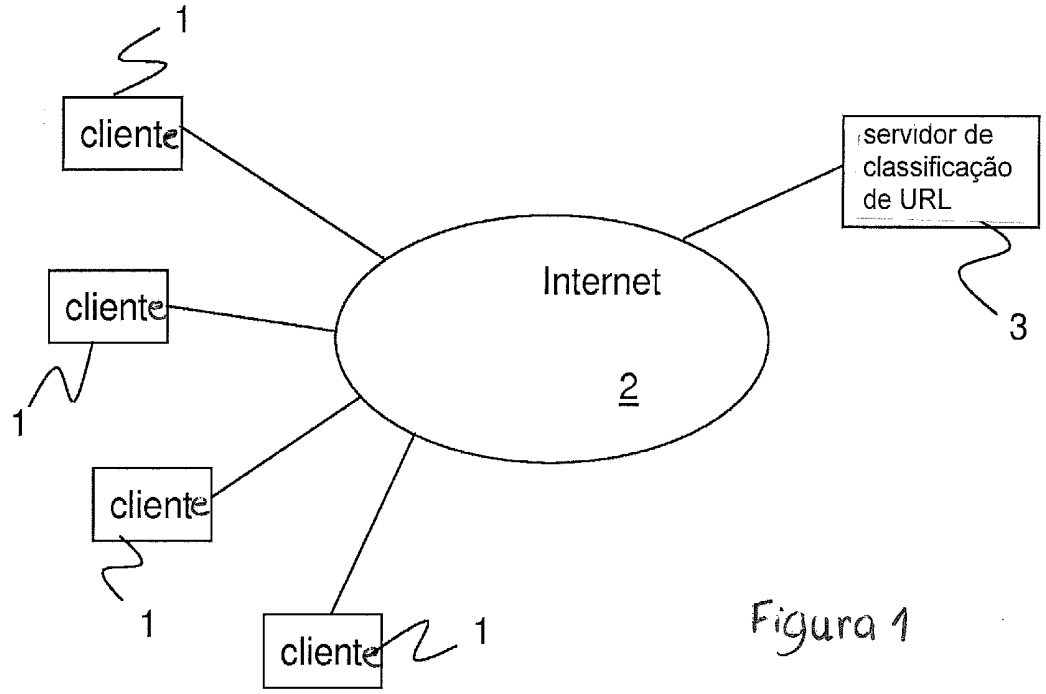

Figura 1

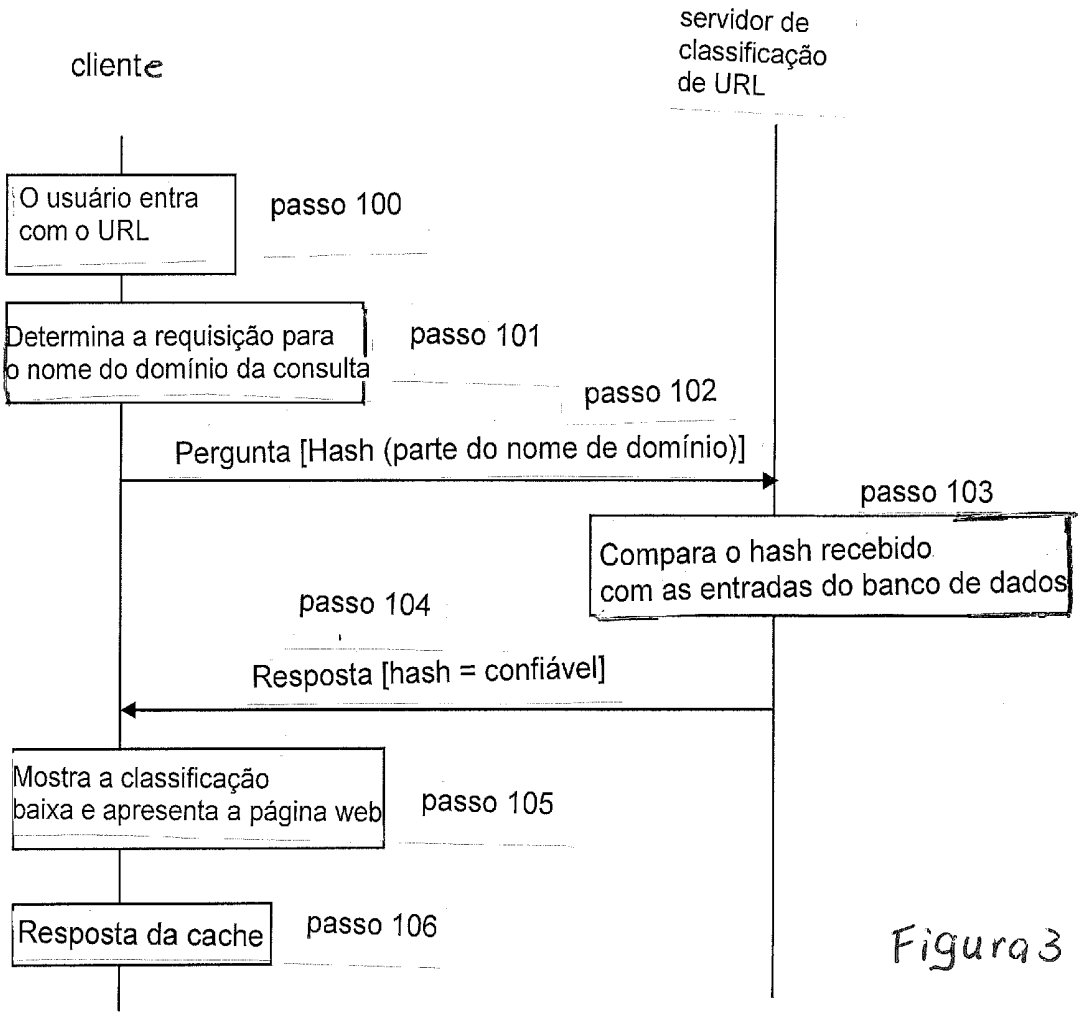

Figura 3

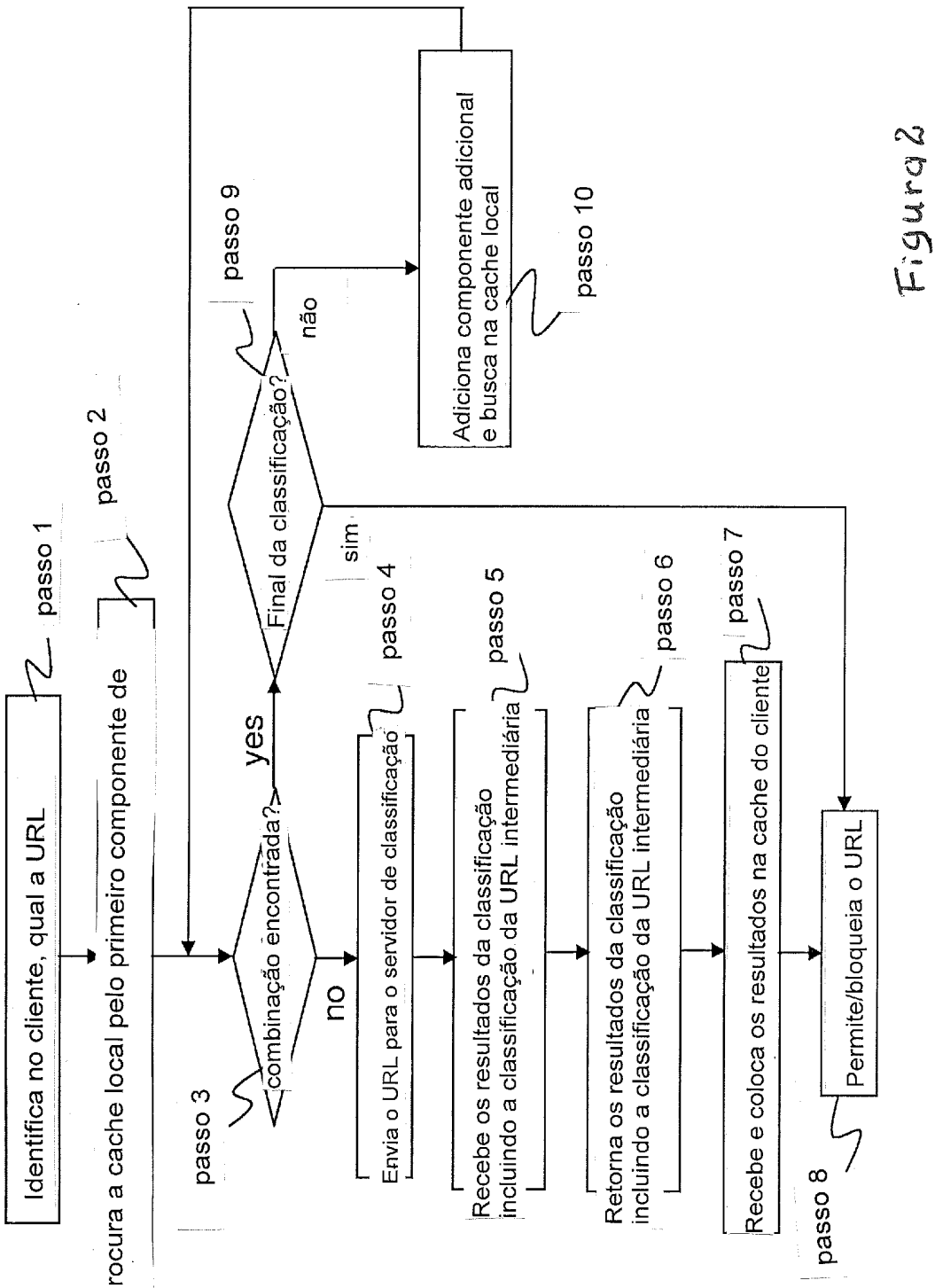
Identifica no cliente, qual a URL
passo 1
Procura a cache local pelo primeiro componente de
passo 2
passo 3
combinação encontrada?
yes
no
Final da classificação?
sim
não
passo 9
Envia o URL para o servidor de classificação
passo 4
Recebe os resultados da classificação incluindo a classificação da URL intermediária
passo 5
Retorna os resultados da classificação incluindo a classificação da URL intermediária
passo 6
Recebe e coloca os resultados na cache do cliente
passo 7
Permite/bloqueia o URL
passo 8
Adiciona componente adicional e busca na cache local
passo 10

Figura 2

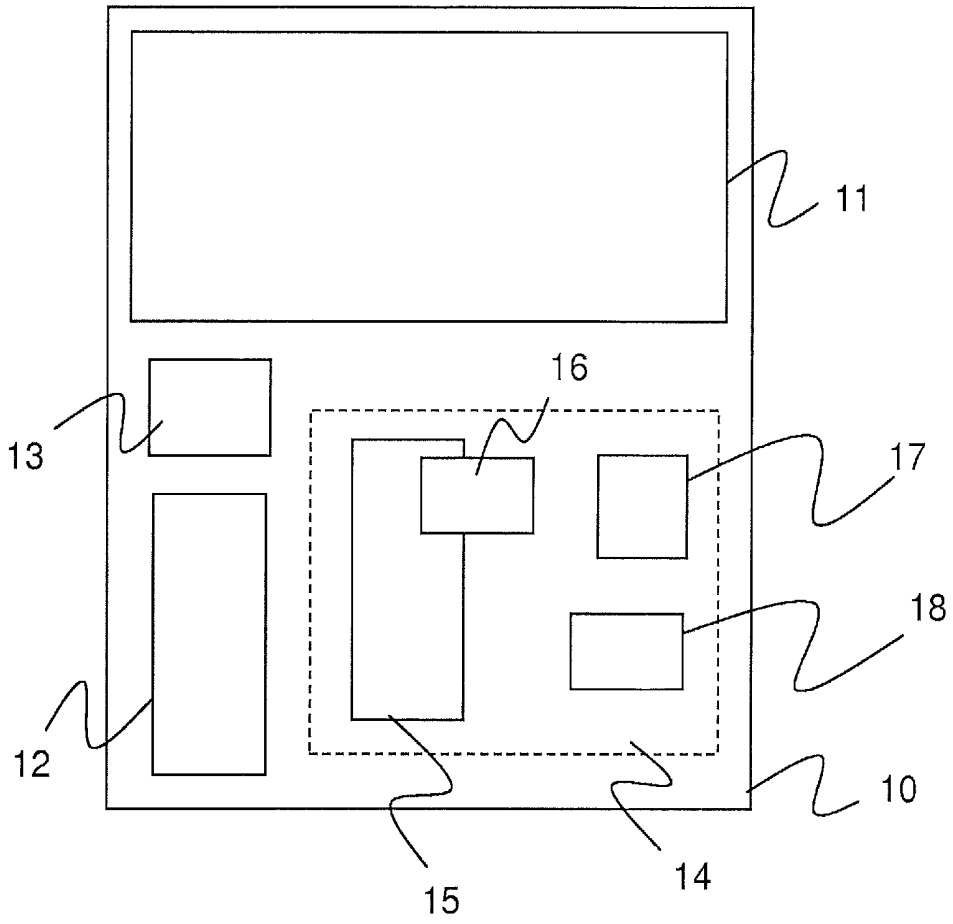

Figura 4

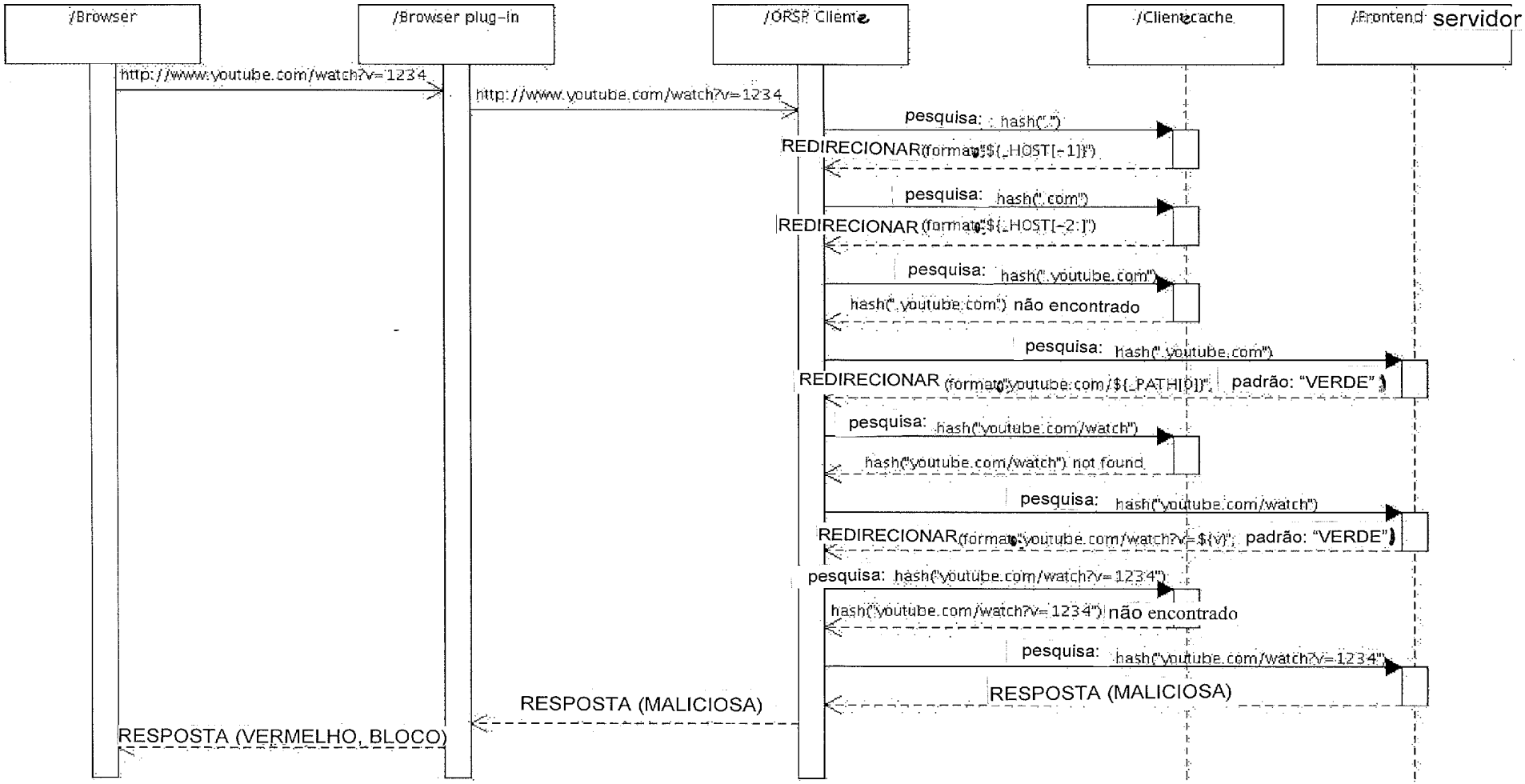
/Browser
/Browser plug-in
/ORSP Cliente
/Clientecache
/Frontend servidor
http://www.youtube.com/watch?v=1234
http://www.youtube.com/watch?v=1234
pesquisa: hash(".")
REDIRECIONAR(formato"${_HOST[-1]}")
pesquisa: hash(".com")
REDIRECIONAR (formato"${_HOST[-2:]}")
pesquisa: hash(".youtube.com")
hash(".youtube.com") não encontrado
pesquisa: hash(".youtube.com")
REDIRECIONAR (formato"youtube.com/${_PATH[0]}", padrão: "VERDE")
pesquisa: hash("youtube.com/watch")
hash("youtube.com/watch") not found
pesquisa: hash("youtube.com/watch")
REDIRECIONAR(formato"youtube.com/watch?v=${v}", padrão: "VERDE")
pesquisa: hash("youtube.com/watch?v=1234")
hash("youtube.com/watch?v=1234") não encontrado
pesquisa: hash("youtube.com/watch?v=1234")
RESPOSTA (MALICIOSA)
RESPOSTA (MALICIOSA)
RESPOSTA (VERMELHO, BLOCO)

Figura5

RESUMO

MÉTODO E EQUIPAMENTO PARA CLASSIFICAÇÃO DE URLs. Método para fornecer informação de classificação em relação aos Identificadores de Recursos Uniformes para um terminal de cliente. O método contempla a identificação de um Identificador de Recurso Uniforme no terminal do cliente, enviando uma primeira consulta para um servidor de classificação através de uma rede de IP, a consulta é incluída como uma sequência de consulta de primeiro componente do Identificador de Recurso Uniforme identificado, ou uma derivativa deste primeiro componente, e recebe a primeira consulta no servidor de classificação que determina se existe ou não uma classificação para a sequência de consulta. Uma resposta é enviada pelo servidor para o terminal do cliente, a resposta inclui uma classificação determinada ou uma indicação de que não existe classificação. A resposta é recebida no terminal do cliente e, se a classificação incluída na resposta assim indicar ou, por outro lado, se a resposta assim indicar, então uma consulta adicional é enviada para o servidor de classificação, esta consulta adicional é incluída como uma sequência de consulta do primeiro componente e do segundo componente do Identificador de Recurso Uniforme identificado, ou uma derivativa do primeiro e segundo componentes. Os passos de recepção da primeira consulta no servidor de classificação, enviando uma resposta e recebendo-a no terminal do cliente, são repetidos uma ou mais vezes, conforme o necessário, adicionando para cada iteração um componente a mais na sequência de consulta.