

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 658 980**

51 Int. Cl.:

H04L 12/24 (2006.01)

H04L 29/12 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.11.2014 E 14191409 (3)**

97 Fecha y número de publicación de la concesión europea: **10.01.2018 EP 2869495**

54 Título: **Deduplicación de nodo en un sistema de supervisión de red**

30 Prioridad:

05.11.2013 US 201314072150

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

13.03.2018

73 Titular/es:

**SOLARWINDS WORLDWIDE, LLC (100.0%)
7171 Southwest Parkway, Building 400
Austin, TX 78735, US**

72 Inventor/es:

**ZIZLAVSKY, MAREK;
POSPISIL, TOMAS y
MRKVICKA, TOMAS**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 658 980 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Deduplicación de nodo en un sistema de supervisión de red

5 Antecedentes

Campo

10 Realizaciones de la invención se refieren en general a supervisión de tráfico de red, análisis y/o notificación. Más particularmente, algunas realizaciones se dirigen a métodos, sistemas y programas informáticos para deduplicación de nodo de nodos físicos supervisados por un sistema de supervisión de red, por ejemplo.

Descripción de la técnica relacionada

15 La gestión de redes incluye actividades, métodos, procedimientos y herramientas relacionadas con la operación, administración, mantenimiento y/o suministro de sistemas en red. Funciones que pueden realizarse como parte de gestión de redes pueden incluir, por ejemplo, planificación, control, despliegue, asignación, coordinación y supervisión de los recursos de una red. Funciones adicionales pueden relacionarse con planificación de red, asignación de frecuencia, equilibrio de carga, gestión de configuración, gestión de fallos, gestión de seguridad, gestión de rendimiento, gestión de ancho de banda, analíticas de ruta y gestión de contabilidad.

20 Como se ha mencionado anteriormente, un subconjunto de gestión de redes incluye supervisión de red de tráfico de red. Datos de tráfico de red son de interés para administradores de red por un número de razones, incluyendo el análisis del impacto de una nueva aplicación en la red, solución de problemas de puntos débiles de red, detección de dispositivos de red lentos o en fallo, detección de grandes consumidores de ancho de banda y seguridad de redes. Se han desarrollado diversos protocolos para datos de flujo de tráfico de red. Estos protocolos pueden contener numerosos tipos de información, tales como direcciones de Protocolo de Internet (IP) de origen, dirección IP de destino, puerto de origen, puerto de destino, protocolo IP, interfaz de entrada, Tipo de Servicio IP, momento de inicio y finalización, número de bytes y siguiente salto.

25 A medida que las redes se vuelven más grandes y más complejas, sistemas que supervisan, analizan y notifican sobre datos de flujo de tráfico deben volverse más eficientes en el manejo del número creciente de dispositivos de red y cantidad de información generada sobre los mismos.

30 El artículo titulado "Unique network node identification algorithm" de Hewlett-Packard Company et al., Research disclosure, vol. 582. n.º 62, 1 de octubre de 2012, páginas 838 a 840, ISSN 0374-4353, describe un algoritmo de identificación de nodo único, en el que los dispositivos de red se identifican singularmente usando una combinación de direcciones MAC en el dispositivo, direcciones IP asignadas al dispositivo, la ID de objeto de sistema SNMP del dispositivo, el nombre DNS del dispositivo y el anfitrión al que se asigna el dispositivo.

35 El artículo titulado "Mix-n-Match: Building Personal Libraries from Web Content" de Matthias Geel et al., Theory and Practice of Digital Libraries, 23 de septiembre de 2012, páginas 345 a 365, ISBN 978-3-642-33289-0, describe un enfoque a la agregación de contenido web que permite que información sea recogida de páginas web, independiente de lenguajes de marcas específicos. Se describe que un motor de agregación de contenido se realiza como un marco extensible de una manera que usuarios finales así como desarrolladores pueden usar las herramientas asociadas para crear librerías personales de contenido extraído de la web.

Sumario

40 De acuerdo con la presente invención, se divulgan métodos, aparatos y productos de programa informático para deduplicación de nodo, como se define en las reivindicaciones adjuntas.

45 Ciertas realizaciones se dirigen a métodos, aparatos y productos de programa informático para deduplicación de nodos, que comprenden, se configuran para ejecutar y se configuran para controlar un procesador para ejecutar un proceso que comprende: descubrir, mediante un aparato de supervisión de red, nodos en una red, recoger una lista de direcciones de Protocolo de Internet (IP), direcciones de Control de Acceso al Medio (MAC), nombres de sistema de nombres de dominio (DNS) y sysnames para cada uno de los nodos descubiertos en la red, comparar las direcciones IP de cada uno de los nodos descubiertos con direcciones IP de nodos (supervisados) actuales y otros nodos (anteriormente) descubiertos, comparar las direcciones MAC de cada uno de los nodos descubiertos con direcciones MAC de los nodos (supervisados) actuales y los otros nodos (anteriormente) descubiertos, comparar los nombres DNS de cada uno de los nodos descubiertos con nombres DNS de los nodos (supervisados) actuales y los otros nodos (anteriormente) descubiertos, comparar los sysnames de cada uno de los nodos descubiertos con sysnames de los nodos (supervisados) actuales y los otros nodos (anteriormente) descubiertos, y determinar nodos duplicados que son duplicados de los otros nodos (anteriormente) descubiertos y/o los nodos (supervisados) actuales basándose en la comparación de las direcciones IP, direcciones MAC, nombres DNS y sysnames, en el que, para la determinación, cada uno de los nodos descubiertos se asigna una ID de nodo y se asigna un

ÍndiceConcordancia a cada ID de nodo, indicando el ÍndiceConcordancia una probabilidad de una concordancia entre el nodo descubierto y cualquiera de los nodos (supervisados) actuales y los otros nodos (anteriormente) descubiertos.

5 Breve descripción de los dibujos

Para comprensión apropiada de la invención, debería hacerse referencia a los dibujos adjuntos, en los que:

- 10 la Figura 1 ilustra un diagrama de bloques de un sistema de acuerdo con una realización;
- la Figura 2 ilustra un diagrama de bloques de un sistema de acuerdo con una realización;
- la Figura 3 ilustra un diagrama de flujo de un método de acuerdo con una realización;
- la Figura 4 ilustra un diagrama de flujo de un método de acuerdo con otra realización;
- la Figura 5 ilustra un diagrama de bloques de un aparato de acuerdo con una realización; y
- 15 la Figura 6 ilustra un diagrama de flujo de un método de acuerdo con otra realización.

Descripción detallada

20 Se entenderá fácilmente que los componentes de la invención, como se describe e ilustra de forma general en las figuras en este documento, pueden disponerse y diseñarse en una amplia variedad de diferentes configuraciones. Por lo tanto, la siguiente descripción detallada de las realizaciones de sistemas, métodos, aparatos, y productos de programa informático para deduplicación de nodo, como se representa en las figuras adjuntas, no pretende limitar el alcance de la invención, sino que es meramente representativo de realizaciones seleccionadas de la invención.

25 Las prestaciones, estructuras o características de la invención descritas a lo largo de esta memoria descriptiva pueden combinarse de cualquier manera adecuada en una o más realizaciones. Por ejemplo, el uso de la frase "ciertas realizaciones," "algunas realizaciones" u otro lenguaje similar, a lo largo de esta memoria descriptiva se refiere al hecho de que una prestación, estructura o característica particular descrita en conexión con la realización puede incluirse en al menos una realización de la presente invención. Por lo tanto, apariciones de las frases "en ciertas realizaciones," "en algunas realizaciones," "en otras realizaciones" u otro lenguaje similar, a lo largo de esta memoria descriptiva no necesariamente se refieren todas al mismo grupo de realizaciones y las prestaciones, estructuras o características descritas pueden combinarse de cualquier manera adecuada en una o más realizaciones. Adicionalmente, si se desea, las diferentes funciones analizadas anteriormente pueden realizarse en un orden diferente y/o simultáneamente entre sí. Adicionalmente, si se desea, una o más de las funciones descritas pueden ser opcionales o pueden combinarse. Como tal, la siguiente descripción debería considerarse como
35 meramente ilustrativa de los principios, contenidos y realizaciones de esta invención y no en limitación de la misma.

Se ha de observar que a lo largo de esta memoria descriptiva los términos dispositivos de red y nodos de red, o simplemente dispositivos o nodos, pueden usarse intercambiamente para referirse a cualquier dispositivo físico que es capaz de conectarse a y/o comunicar en una red. Ejemplos de tales dispositivos o nodos pueden incluir, pero
40 sin limitación, encaminadores, conmutadores, servidores, ordenadores, portátiles, tabletas, teléfonos, impresoras, dispositivos móviles y cualquier otro componente actual o futuro capaz de enviar, recibir o reenviar información en un canal de comunicaciones.

La Figura 1 ilustra un ejemplo de un sistema de acuerdo con una realización. El sistema incluye el sistema de supervisión de red 100, almacenamiento 110 de sistema de supervisión de red, dispositivo de conmutación 130 y uno o más dispositivos de red 120. El almacenamiento 110 de sistema de supervisión de red puede almacenar datos de supervisión de red. El almacenamiento 110 de sistema de supervisión de red puede ser una base de datos o cualquier otro dispositivo de almacenamiento apropiado. Los dispositivos de red 120 pueden ser nodos en la red que se supervisan mediante el supervisor 100 de tráfico de red. Se ha de observar que cualquier número y tipo de dispositivos de red 120 puede soportarse en el sistema. Por consiguiente, realizaciones no se limitan al número y tipo de dispositivos de red ilustrados en la Figura 1. En una realización, el almacenamiento 110 de sistema de supervisión de red puede almacenar tablas de base de datos de resultados de descubrimientos que almacenan y organizan información sobre dispositivos de red descubiertos en la red.

55 En algunos casos, dispositivos de red pueden ser accesibles en múltiples direcciones de Protocolo de Internet (IP). Por ejemplo, algunos dispositivos pueden tener simultáneamente más de una dirección IP. También, algunos nodos son dinámicos en que tienen direcciones IP que varían con el tiempo. Tales direcciones IP dinámicas provocan problemas para lógica de deduplicación porque podría estar concordando una dirección IP primaria anticuada de resultado descubierto con la dirección IP actual de un nodo dinámico.

60 En la actualidad, sistemas generalmente aplican lógica donde una dirección IP iguala a un nodo de red. Como resultado, nodos con múltiples vínculos de dirección IP pueden no ser reconocidos con un único nodo. Este comportamiento puede resultar en un estado en el que un dispositivo físico se supervisa más de una vez, lo que puede provocar obviamente sobrecarga adicional e ineficiencias en el sistema de supervisión de red.

65

Por lo tanto, usuarios normalmente quieren tener dispositivos que respondan en múltiples direcciones IP supervisadas como un único nodo en la red. Realizaciones de la invención implementan un descubrimiento de red automático que es capaz de detectar tal situación y evitar el procesamiento del mismo nodo físico múltiples veces (por ejemplo, cada vez que se usa una dirección IP diferente). Una realización incluye una lógica de deduplicación configurada para identificar automáticamente nodos duplicados de tal forma que un único nodo no se supervisa múltiples veces.

Ciertas realizaciones identifican una o más piezas de información que funcionan como un identificador de nodo necesario para identificar singularmente el nodo de red. Teniendo este identificador de nodo único, ciertas realizaciones pueden proceder a definir la lógica, que detecta duplicados dentro de nodos encontrados durante el proceso de descubrimiento (por ejemplo, mismo nodo de red físico con diferentes direcciones IP) y/o dentro de nodos que ya se han supervisado con una dirección IP diferente.

De acuerdo con ciertas realizaciones, la información usada para identificar singularmente nodos incluye información que puede obtenerse fácilmente, está disponible para una mayoría de dispositivos (es decir, independiente del fabricante), está disponible para nodos de Protocolo de Mensaje de Control de Internet (ICMP) así como otros tipos de nodos (por ejemplo, nodos ICMP pueden considerarse que son nodos no alcanzables en SNMP o WMI) y es un conjunto mínimo de información que aún puede proporcionar resultados precisos mientras es eficiente.

Ciertas realizaciones pueden manejar al menos dos casos de uso típicos. Por ejemplo, un caso de uso puede incluir ejecutar un descubrimiento de red automático en una subred, que contiene dispositivos accesibles en múltiples direcciones IP. Como es conocido, dispositivos que pertenecen a una subred se direccionan con un grupo de bit más significativo idéntico y común en sus direcciones IP. Otro caso de uso puede incluir ejecutar un descubrimiento de red automático en una subred, que ya se supervisa.

Ciertas realizaciones son capaces de identificar nodos de red como duplicados de red, incluso si un conjunto de información se recogió en un periodo de tiempo diferente que el otro. Por ejemplo, esto puede suceder cuando un usuario decide importar resultados de descubrimientos para un perfil de descubrimiento planificado. En un caso de este tipo, algunas realizaciones pueden necesitar trabajar con información posiblemente anticuada recogida durante el descubrimiento y comparar la misma con nueva información, que se recoge continuamente para todos los modos supervisados. Por consiguiente, deduplicación de acuerdo con ciertas realizaciones puede tener lugar durante un trabajo de descubrimiento donde nodos descubiertos se comparan entre sí de tal forma que duplicados se eliminan y/o puede tener lugar durante una importación de resultado de descubrimiento en la que nodos descubiertos se importan y comparan contra nodos existentes (por ejemplo, nodos almacenados en almacenamiento de sistema de supervisión de red o base de datos) de tal forma que nodos duplicados se identifican y eliminan.

De acuerdo con una realización, se usa un conjunto de recopilación de datos compuesto de nombre DNS, Sysname, dirección IP y dirección MAC para cada nodo de red para ayudar a identificar nodos duplicados. Una realización incluye lógica que compara piezas de información del conjunto de recopilación de datos entre sí, por ejemplo, DNS con DNS, MAC con MAC, etc. La lógica puede implementarse en un componente de Detector de Duplicado. Un resultado de la detección es el índice de concordancia, que indica si hay una concordancia, no hay una concordancia o se desconoce. Ciertas realizaciones también proporcionan lógica para agregar resultados parciales de detectores y calcular un veredicto en cuanto a si un nodo es un duplicado o no. Por ejemplo, en una realización, la deduplicación de nodo puede incluir varias subinteracciones que cada una es responsable de descubrir el intervalo IP seleccionado. Al final de cada iteración, se realiza deduplicación para omitir nodos (por ejemplo, puntos de extremo) tan pronto como se crea que son duplicados. En caso de un resultado desconocido o no duplicado para una iteración, el nodo se pasa a la siguiente etapa o iteración para procesamiento.

En una realización, pueden proporcionarse dos conjuntos de detectores de duplicados. Un conjunto de detectores puede usarse durante el proceso de descubrimiento automatizado para filtrar nuevos dispositivos encontrados y eliminar duplicados. Otro conjunto de detectores puede usarse durante importación de resultados de descubrimiento para evitar añadir nodos duplicados en el conjunto de nodos supervisados.

En una realización, el sistema se configura para recoger lista de todos los nombres DNS, Sysnames, direcciones IP y direcciones MAC para todos los nodos descubiertos, y almacenar los mismos, por ejemplo, como parte de un resultado de trabajo de descubrimiento. Esta información puede almacenarse en el almacenamiento de supervisión de sistema de red o base de datos 110. Como se ha mencionado anteriormente, esta información puede usarse durante al menos dos fases de descubrimiento. Por ejemplo, los nombres DNS, Sysnames, direcciones IP y direcciones MAC pueden usarse cuando se ejecuta descubrimiento para comprobar si en la actualidad nodo(s) descubierto(s) son un duplicado de cualquier otro ya encontrado, y/o puede usarse durante importación de resultados de descubrimiento para comparar resultado(s) de descubrimiento con nodos existentes supervisados por el sistema. En una realización, la información MAC se almacena en un almacenamiento persistente, por ejemplo, como parte de resultado de descubrimiento.

La Figura 2 ilustra un sistema 200, de acuerdo con una realización, que puede incluir un detector 201 de duplicado de dirección IP, un detector 203 de duplicado de DNS, un detector 202 de duplicado de dirección MAC y un detector

204 de duplicado de sysname. En una realización, una base de datos 210 puede estar en comunicación con el detector 201 de duplicado de dirección IP, el detector 203 de duplicado de DNS, el detector 202 de duplicado de dirección MAC y el detector 204 de duplicado de sysname de tal forma que cada uno de los detectores puede recoger las direcciones relevantes almacenadas en la base de datos 210. En una realización, la base de datos 210 puede almacenarse en el almacenamiento 110 de datos de tráfico de red ilustrado en la Figura 1. Se ha de observar que el sistema 200 puede implementarse completamente en hardware o en una combinación de hardware y software.

Cada uno de los detectores de duplicados 201, 202, 203, 204 puede tener una prioridad definida de orden de ejecución (por ejemplo, un número bajo indica ejecución previa), una ponderación que indica la fiabilidad del resultado proporcionado por el detector de duplicado (por ejemplo, una ponderación de 0 no tendrá impacto en el resultado final) y un veto que se usa como una prioridad máxima para determinar si un nodo es un duplicado o no.

El detector 203 de duplicado de DNS puede configurarse para comparar un DNS de un nodo descubierto con todos los otros nodos descubiertos y nodos actuales que se supervisan. En una realización, el detector 203 de duplicado de DNS puede concluir que un nodo es un duplicado si el DNS de un nodo descubierto es el mismo que un DNS usado por cualquiera de los nodos supervisados o cualquiera de los otros nodos descubiertos.

Las direcciones MAC generalmente son únicas por diseño (aunque existen situaciones en las que se usa la misma dirección MAC en dos diferentes dispositivos, por ejemplo, máquinas virtuales clonadas alojadas en dos anfitriones virtuales separados). El detector 202 de duplicado de dirección MAC puede configurarse para comparar una dirección MAC de un nodo descubierto contra todas las direcciones MAC anteriormente recogidas de nodos, que pueden almacenarse en una tabla de base de datos de direcciones MAC de nodos en la base de datos 210. El detector 202 de duplicado de dirección MAC puede configurarse para concluir que un nodo es un duplicado si un conjunto de direcciones MAC descubiertas es un subconjunto de direcciones MAC actualmente supervisadas por un nodo o si un conjunto de direcciones MAC supervisadas es un subconjunto de direcciones MAC descubiertas.

La Tabla 1 a continuación ilustra un ejemplo en el que los nodos se consideran iguales de acuerdo con el detector 202 de duplicado de dirección MAC, donde A, B, C, ... representan direcciones MAC. Mientras tanto, la Tabla 2 a continuación ilustra un ejemplo en el que los nodos no se consideran iguales (por ejemplo, basándose en direcciones MAC).

TABLA 1

Nodo A	Nodo B	Nodo C	Nodo D
A	A	A	A
B	B	B	B
C	C		C
			D

TABLA 2

Nodo A	Nodo B	Nodo C	Nodo D
A	A	D	A
B	B	E	D
C	C	F	

Dos nodos, que tienen ambos solo dos direcciones MAC, por ejemplo 0000000000000000 y 00000000000000E0 deberían considerarse como iguales. Por ejemplo, Nodo A y Nodo B son iguales cuando y únicamente cuando la lista de direcciones MAC del Nodo A es subconjunto de la lista de direcciones MAC del Nodo B o lista de MAC del Nodo B es subconjunto de la lista de MAC del Nodo A. En ese caso, el sistema puede mirar a un detector de duplicado diferente o método de deduplicación diferente (por ejemplo, concordancia de sysname) ya que, de acuerdo con direcciones MAC, los dos nodos son iguales.

El detector 204 de duplicado de sysname puede configurarse para comparar un sysname de un nodo descubierto con todos los otros nodos descubiertos y nodos actuales que se supervisan. En una realización, el detector 204 de duplicado de sysname puede concluir que un nodo es un duplicado si el sysname de un nodo descubierto es el mismo que un sysname usado por cualquiera de los nodos supervisados o cualquiera de los otros nodos descubiertos. En una realización, el sysname puede ser el sysname para nodos de Protocolo de Gestión de Red Simple (SNMP) o puede ser el nombre del ordenador completo para nodos de instrumental de administración de Windows (WMI), por ejemplo. Se ha de observar que las realizaciones no limitan las fuentes de datos a SNMP y/o WMI y otros tipos de fuentes de datos son igualmente aplicables de acuerdo con ciertas realizaciones (por ejemplo, CLI en SSH en encaminadores/conmutadores o telnet, etc.).

Como se ha mencionado anteriormente, cada detector de duplicado puede cargar una ponderación asociada a partir de una tabla de base de datos de ajustes. La ponderación representa la fiabilidad del resultado proporcionado por el detector de duplicado asociado. Es posible establecer la ponderación a -1 para deshabilitar al detector de duplicado

asociado. De acuerdo con una realización, valores de ponderación pueden oscilar desde 0 a 100 donde 0 representa el menos fiable y 100 representa el más fiable.

5 Todos los detectores de duplicados (d_1, \dots, d_n) pueden ejecutarse secuencialmente con orden definido por prioridad, como se ha analizado anteriormente. Cada uno de los detectores de duplicados puede establecer que una bandera de 'EsAutorizado' sea verdadera, que a continuación puede terminar la ejecución de siguientes detectores de duplicados. En un caso de este tipo, el voto del detector de duplicado con la bandera 'EsAutorizado' establecida a verdadera se considera como final, ignorando todos los demás votos. De acuerdo con una realización, si no existe ninguna bandera 'EsAutorizado' establecida a verdadera, un resultado final en cuanto a si un nodo es un duplicado se calcula como una suma de todos los valores de resultado de voto de detector de duplicado como se indica a continuación:

$$\text{Decisión Final} = d_1.\text{EsDuplicado()} * d_1.\text{Prioridad} + \dots + d_n.\text{EsDuplicado()} * d_n.\text{Prioridad}$$

15 donde d_1 es un primer detector de duplicado, d_2 es un segundo detector de duplicado, ... y d_n es un $n^{\text{ésimo}}$ detector de duplicado. Por consiguiente, $d_n.\text{EsDuplicado}()$ es una función que representa la $n^{\text{ésima}}$ conclusión del detector de duplicado en cuanto a si un nodo es un duplicado.

20 Como se ilustra en la Tabla 3 a continuación, cada detector de duplicado puede devolver una lista de ID de nodo para todos los nodos duplicados que encuentra. Cada ID de nodo puede tener un ÍndiceConcordancia asociado asignado, que indica la probabilidad de la concordancia. En una realización, el intervalo de valores de ÍndiceConcordancia es de 0 a 100, donde 0 indica la menor probabilidad de una concordancia y 100 indica la mayor probabilidad de una concordancia. De acuerdo con una realización, el sistema 200 puede configurarse para agrupar la información de nodo duplicado representada en la Tabla 3 por ID de nodo, y sumar los ÍndicesConcordancia con el mismo ID de nodo. A continuación, el sistema 200 puede configurarse para seleccionar el nodo ID con el mayor ÍndiceConcordancia total sumado para descartar.

TABLE 3

Detector de duplicado	IDNodoDuplicado	ÍndiceConcordancia
Dns	5	90
Mac	1	60
Mac	5	80
Sysname	5	85
Decisión Final:	5	90

30 La Tabla 4 a continuación ilustra una tabla de resultados de ejemplo, de acuerdo con una realización. En este ejemplo, cada fila de la tabla puede representar un nodo. La columna de 'DetectorDuplicadoDns' muestra la conclusión del detector de duplicado de DNS en cuanto a si el nodo es un duplicado o no. De manera similar, la columna de 'DetectorDuplicadoDirecciónMAC' muestra la conclusión del detector de duplicado de dirección MAC en cuanto a si el nodo es un duplicado o no y la columna 'DetectorDuplicadoNombre' muestra la conclusión del detector de duplicado de Sysname en cuanto a si el nodo es un duplicado o no. A continuación, la columna final de 'Resultado Esperado' muestra el resultado esperado para el nodo.

DetectorDuplicadoDns	DetectorDuplicado DirecciónMac	DetectorDuplicadoNombre	Resultado Esperado
Es duplicado	Es duplicado	Es duplicado	Es duplicado
Es duplicado	Es duplicado	No se sabe	Es duplicado
Es duplicado	Es duplicado	No es duplicado	Es duplicado
Es duplicado	No se sabe	Es duplicado	Es duplicado
Es duplicado	No se sabe	No se sabe	Es duplicado
Es duplicado	No se sabe	No es duplicado	Es duplicado
Es duplicado	No es duplicado	Es duplicado	Es duplicado

40

Es duplicado	No es duplicado	No se sabe	Es duplicado
Es duplicado	No es duplicado	No es duplicado	No es duplicado
No se sabe	Es duplicado	Es duplicado	Es duplicado
No se sabe	Es duplicado	No se sabe	Es duplicado
No se sabe	Es duplicado	No es duplicado	Es duplicado
No se sabe	No se sabe	Es duplicado	Es duplicado
No se sabe	No se sabe	No se sabe	No es duplicado
No se sabe	No se sabe	No es duplicado	No es duplicado
No se sabe	No es duplicado	Es duplicado	No es duplicado
No se sabe	No es duplicado	No se sabe	No es duplicado
No se sabe	No es duplicado	No es duplicado	No es duplicado
No es duplicado	Es duplicado	Es duplicado	Es duplicado
No es duplicado	Es duplicado	No se sabe	No es duplicado
No es duplicado	Es duplicado	No es duplicado	No es duplicado
No es duplicado	No se sabe	Es duplicado	No es duplicado
No es duplicado	No se sabe	No se sabe	No es duplicado
No es duplicado	No se sabe	No es duplicado	No es duplicado
No es duplicado	No es duplicado	Es duplicado	No es duplicado
No es duplicado	No es duplicado	No se sabe	No es duplicado
No es duplicado	No es duplicado	No es duplicado	No es duplicado

TABLA 4

5 La Figura 3 ilustra un diagrama de flujo de un método para deduplicación de nodo de acuerdo con una realización. En el ejemplo de la Figura 3, en 300, se descubren uno o más nodos. En 310, se descubre la dirección IP asociada del nodo(s). En 320, se detecta la tecnología soportada usada para obtener información de dispositivo, tales como SNMP, WMI, etc. En 330, se obtiene información sobre el nodo(s) a través de la tecnología soportada (detectada). En 340, se realiza deduplicación del nodo(s) descubierto(s). Se determina si el nodo(s) es un duplicado en 350. Si lo es, el nodo(s) duplicado(s) se omite del resultado descubierto establecido en 360. Si el nodo(s) no es un duplicado, en 370, nodo(s) descubierto(s) se guarda en la base de datos. En 380, datos asociados con nodo(s) descubierto(s) se preparan para importación de descubrimiento.

15 La Figura 4 ilustra un diagrama de flujo de un método para deduplicación de nodo de acuerdo con otra realización. En el ejemplo de la Figura 4, en 400, se guarda un resultado de un trabajo de descubrimiento. En 405, el resultado de descubrimiento se carga y paraleliza de la base de datos. En 410, detección de duplicado se hace esencialmente con dirección IP primaria asociada con el nodo. En 415, se detecta si nodo con misma IP ya se supervisa por cualquier motor. Si nodo con misma IP ya se supervisa, la información de nodo existente se actualiza en 420. Si nodo con misma IP no se supervisa ya, se realiza detección de duplicados contra nodos ignorados en 425. En 430, se determina si nodo es un duplicado. Si es así, las condiciones se registran y el nodo duplicado se descarta en 435. Si no se determina como un duplicado, se realiza una comprobación de duplicado contra todos los modos supervisados en 440. En 445, se determina si nodo es un duplicado. Si se determina como duplicado, las condiciones se registran y el nodo duplicado se descarta en 450. Si no se determina como un duplicado, información de nodo(s) se guarda en almacenamiento persistente en 455. En 460, resultado de importación de descubrimiento es una lista de nodos con direcciones IP y direcciones MAC asociadas.

La Figura 5 ilustra un diagrama de bloques de un aparato 10 que puede implementar una realización de la invención. El aparato 10 puede incluir un bus 12 u otro mecanismo de comunicaciones para comunicar información entre componentes del aparato 10. El aparato 10 también incluye un procesador 22, acoplado al bus 12, para procesar información y ejecutar instrucciones u operaciones. El procesador 22 puede ser cualquier tipo de procesador de fin general o específico. El aparato 10 incluye adicionalmente una memoria 14 para almacenar información e instrucciones a ejecutar por el procesador 22. La memoria 14 puede constar de cualquier combinación de Memoria de Acceso Aleatorio ("RAM"), memoria de sólo lectura ("ROM"), almacenamiento estático tal como un disco magnético u óptico, o cualquier otro tipo de máquina o medio legible por ordenador. El aparato 10 adicionalmente incluye un dispositivo de comunicación 20, tal como una tarjeta de interfaz de red u otra interfaz de comunicaciones, para proporcionar acceso a una red. Como resultado, un usuario puede interactuar con el aparato 10 directamente o remotamente a través de una red o cualquier otro método.

Medio legible por ordenador puede ser cualquier medio disponible que puede accederse mediante el procesador 22 e incluye tanto medios volátiles como no volátiles, memoria extraíble y no extraíble y medios de comunicación. Medios de comunicación pueden incluir instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos en una señal de datos modulada tal como una onda portadora u otro mecanismo de transporte e incluye cualquier medio de distribución de información.

El procesador 22 se acopla adicionalmente a través del bus 12 a un dispositivo de presentación 24, tales como un visualizador, monitor, pantalla o navegador web, para visualizar información a un usuario, tales como información de tráfico de red. Un componente de entrada de usuario 26, tales como un teclado, ratón informático o navegador web, se acoplan adicionalmente al bus 12 para habilitar que un usuario interactúe con el aparato 10. El procesador 22 y memoria 14 también pueden acoplarse a través del bus 12 a un sistema de base de datos 30 y, por lo tanto, pueden ser capaces de acceder y recuperar información almacenada en el sistema de base de datos 30. En una realización, el sistema de base de datos es el almacenamiento 110 de sistema de supervisión de red ilustrado en la Figura 1. Aunque únicamente se ilustra una única base de datos en la Figura 5, puede usarse cualquier número de bases de datos de acuerdo con ciertas realizaciones.

En una realización, la memoria 14 almacena módulos de software que proporcionan funcionalidad cuando se ejecuta por el procesador 22. Los módulos pueden incluir un sistema operativo 15 que proporciona funcionalidad de sistema operativo para el aparato 10. La memoria también puede almacenar uno o más detectores de duplicado 16, que soportan una funcionalidad de deduplicación de nodo, como se ha analizado anteriormente. El uno o más detectores de duplicado 16 pueden incluir, por ejemplo, el detector 201 de duplicado de dirección IP, el detector 203 de duplicado de DNS, el detector 202 de duplicado de dirección MAC y el detector 204 de duplicado de sysname, como se representa en la Figura 2 analizada anteriormente. El aparato 10 también puede incluir uno o más otros módulos funcionales 18 para proporcionar funcionalidad adicional.

El sistema de base de datos puede incluir un servidor de base de datos y cualquier tipo de base de datos, tales como una base de datos relacional o de archivo plano. El sistema de base de datos puede almacenar datos relacionados con flujo de tráfico de red de cada una de las entidades en la red y/o cualquier dato asociado con el aparato 10 o sus módulos y componentes asociados.

En ciertas realizaciones, el procesador 22, el detector(s) de duplicado 16 y otros módulos funcionales 18 pueden implementarse como unidades físicas separadas y lógicas o pueden implementarse en una única unidad física o lógica. Adicionalmente, en algunas realizaciones, el procesador 22, el detector(s) de duplicado 16 y otros módulos funcionales 18 pueden implementarse en hardware o como cualquier combinación adecuada de hardware y software.

En algunas realizaciones, el procesador 22 se configura para controlar el aparato 10 para descubrir nodos en una red. De acuerdo con una realización, información que identifica los nodos descubiertos puede almacenarse en base de datos 110, por ejemplo. El procesador 22 puede configurarse para controlar el aparato 10 para recoger una lista de direcciones IP, direcciones MAC, nombres DNS y sysnames para cada uno de los nodos descubiertos en la red.

De acuerdo con una realización, el procesador 22 puede configurarse para controlar el aparato 10 para ejecutar un detector de duplicado de IP configurado para comparar las direcciones IP de cada uno de los nodos descubiertos con direcciones IP de nodos actuales y otros nodos descubiertos, un detector de duplicado de MAC configurado para comparar las direcciones MAC de cada uno de los nodos descubiertos con direcciones MAC de los nodos actuales y los otros nodos descubiertos, un detector de duplicado de DNS configurado para comparar los nombres DNS de cada uno de los nodos descubiertos con nombres DNS de los nodos actuales y los otros nodos descubiertos, y un detector de duplicado de nombre configurado para comparar los sysnames de cada uno de los nodos descubiertos con sysnames de los nodos actuales y los otros nodos descubiertos. El procesador 22 puede configurarse a continuación para controlar el aparato 10 para determinar nodos duplicados que son duplicados de los otros nodos descubiertos y/o los nodos actuales basándose en el resultado de comparación del detector de duplicado de IP, el detector de duplicado de MAC, el detector de duplicado de DNS y el detector de duplicado de nombre.

En una realización, el procesador 22 puede configurarse para controlar el aparato 10 para descartar los nodos duplicados. De acuerdo con una realización, el procesador 22 puede configurarse para controlar el aparato 10 para asignar una prioridad a cada uno del detector de duplicado de IP, el detector de duplicado de MAC, el detector de duplicado de DNS y el detector de duplicado de nombre que determina un orden de ejecución. El aparato 10 puede controlarse para determinar los nodos duplicados, por ejemplo, ejecutando la siguiente fórmula:

$$d_1.EsDuplicado()*d_1.Prioridad+\dots+d_n.EsDuplicado()*d_n.Prioridad.$$

En una realización, a cada uno de los nodos descubiertos puede asignarse una ID de nodo. El procesador 22 puede configurarse para controlar el aparato 10 para asignar un ÍndiceConcordancia a cada ID de nodo, donde el ÍndiceConcordancia indica una probabilidad de una concordancia entre el nodo descubierto y cualquiera de los nodos actuales y los otros nodos descubiertos. De acuerdo con una realización, el procesador 22 puede configurarse para controlar el aparato 10 para agrupar los nodos duplicados por ID de nodo y sumar los ÍndicesConcordancia con el mismo ID de nodo. Adicionalmente, una ponderación se asigna a cada uno del detector de duplicado de IP, el detector de duplicado de MAC, el detector de duplicado de DNS y el detector de duplicado de nombre. La ponderación indica la fiabilidad del resultado proporcionado por los respectivos detectores de duplicados.

La Figura 6 ilustra un diagrama de flujo de ejemplo de un método, de acuerdo con una realización. El método incluye, en 600, descubrir, por ejemplo, mediante un aparato de supervisión de red, nodos en una red. El método puede a continuación incluir, en 610, recoger una lista de direcciones de Protocolo de Internet (IP), direcciones de Control de Acceso al Medio (MAC), nombres de sistema de nombres de dominio (DNS) y sysnames para cada uno de los nodos descubiertos en la red. En 620, el método incluye comparar las direcciones IP de cada uno de los nodos descubiertos con direcciones IP de nodos actuales y otros nodos descubiertos, en 630, el método incluye comparar las direcciones MAC de cada uno de los nodos descubiertos con direcciones MAC de los nodos actuales y los otros nodos descubiertos. En 640, el método incluye comparar los nombres DNS de cada uno de los nodos descubiertos con nombres DNS de los nodos actuales y los otros nodos descubiertos. En 650, el método incluye comparar los sysnames de cada uno de los nodos descubiertos con sysnames de los nodos actuales y los otros nodos descubiertos. El método puede incluir adicionalmente, en 660, determinar nodos duplicados que son duplicados de los otros nodos descubiertos y/o los nodos actuales basándose en la comparación de las direcciones IP, direcciones MAC, nombres DNS y sysnames.

En algunas realizaciones, la funcionalidad de cualquiera de los métodos descritos en este documento, tales como los de las Figuras 3, 4, y 6, pueden implementarse mediante software y/o código de programa informático almacenado en memoria u otro medio legible por ordenador o tangible y ejecutado por un procesador. En otras realizaciones, la funcionalidad puede realizarse mediante hardware, por ejemplo, a través del uso de un circuito integrado específico de la aplicación (ASIC), una matriz de puertas programables (PGA), un campo de matriz de puertas programables (FPGA) o cualquier otra combinación de hardware y software.

Un experto en la materia entenderá fácilmente que la invención como se ha analizado anteriormente puede practicarse con etapas en un orden diferentes y/o con elementos de hardware en configuraciones que son diferentes de las divulgadas. Por lo tanto, aunque la invención se ha descrito basándose en estas realizaciones preferidas, sería evidente para los expertos en la materia que ciertas modificaciones, variaciones y construcciones alternativas serían evidentes, sin alejarse de la presente invención como se define en las reivindicaciones adjuntas.

Como es evidente a partir de lo anterior, existen sistemas divulgados, métodos, aparatos y productos de programa informático para deduplicación de nodo. Un método incluye descubrir, mediante un aparato de supervisión de red, nodos en una red y recoger una lista de direcciones de Protocolo de Internet (IP), direcciones de Control de Acceso al Medio (MAC), nombres de sistema de nombres de dominio (DNS) y sysnames para cada uno de los nodos descubiertos en la red. El método también puede incluir comparar la lista recogida de información para cada uno de los nodos descubiertos con información correspondiente para nodos actuales y otros nodos descubiertos. El método puede a continuación incluir determinar nodos duplicados que son duplicados de los otros nodos descubiertos y/o los nodos actuales basándose en la comparación de las direcciones IP, direcciones MAC, nombres DNS y sysnames.

REIVINDICACIONES

1. Un método, que comprende:

5 descubrir (600), mediante un aparato de supervisión de red, nodos en una red;
 recoger (610) una lista de direcciones de protocolo de internet, IP, direcciones de control de acceso al medio, MAC, nombres de sistema de nombre de dominio, DNS, y sysnames para cada uno de los nodos descubiertos en la red;
 10 comparar (620) las direcciones IP de cada uno de los nodos descubiertos con direcciones IP de nodos supervisados y nodos anteriormente descubiertos;
 comparar (630) las direcciones MAC de cada uno de los nodos descubiertos con direcciones MAC de los nodos supervisados y los nodos anteriormente descubiertos;
 comparar (640) los nombres DNS de cada uno de los nodos descubiertos con nombres DNS de los nodos supervisados y los nodos anteriormente descubiertos;
 15 comparar (650) los sysnames de cada uno de los nodos descubiertos con sysnames de los nodos supervisados y los nodos anteriormente descubiertos;
 determinar (660) nodos duplicados que son duplicados de los nodos anteriormente descubiertos y/o los nodos supervisados basándose en la comparación de las direcciones IP, direcciones MAC, nombres DNS y sysnames, en el que, para la determinación, cada uno de los nodos descubiertos se asigna una ID de nodo y el método comprende adicionalmente asignar un ÍndiceConcordancia a cada ID de nodo, indicando el ÍndiceConcordancia una probabilidad de una concordancia entre el nodo descubierto y cualquiera de los nodos supervisados y los nodos anteriormente descubiertos.

2. El método de acuerdo con la reivindicación 1, comprendiendo además descartar (435) los nodos duplicados.

3. El método de acuerdo con la reivindicación 1, comprendiendo además asignar una prioridad de cada una de las etapas de comparación que determina un orden de ejecución.

4. El método de acuerdo con la reivindicación 1, en el que la determinación se realiza usando una pluralidad de detectores de duplicados, estando cada detector de duplicado configurado para realizar una de las comparaciones, y comprende ejecutar la siguiente fórmula para calcular un resultado final en cuanto a si un nodo descubierto es un duplicado:

$$d_1.EsDuplicado()*d_1.Prioridad + \dots + d_n.EsDuplicado()*d_n.Prioridad$$

donde d_1 es un primer detector de duplicado, d_2 es un segundo detector de duplicado, ..., y d_n es un $n^{\text{ésimo}}$ detector de duplicado, $d_n.EsDuplicado()$ es una función que representa la conclusión del $n^{\text{ésimo}}$ detector de duplicado en cuanto a si el nodo descubierto es un duplicado y $d_n.Prioridad$ es una prioridad asignada a cada uno del $n^{\text{ésimo}}$ detector de duplicado.

5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, comprendiendo además agrupar los nodos duplicados por ID de nodo y sumar los ÍndicesConcordancia con el mismo ID de nodo.

6. Un aparato, que comprende:

al menos un procesador (22) y al menos una memoria (14) que incluyen código de programa informático, la al menos una memoria y el código de programa informático se configuran, con el al menos un procesador, para provocar que el aparato al menos descubra (600) nodos en una red;
 50 recoger (610) una lista de Direcciones de Protocolo de Internet, IP, direcciones de Control de Acceso al Medio, MAC, nombres de sistema de nombre de dominio, DNS, y sysnames para cada uno de los nodos descubiertos en la red;
 en el que el al menos un procesador se configura adicionalmente para ejecutar:

un detector de duplicado de IP (16, 201) configurado para comparar (620) las direcciones IP de cada uno de los nodos descubiertos con direcciones IP de nodos supervisados y nodos anteriormente descubiertos;
 un detector de duplicado de MAC (16, 202) configurado para comparar (630) las direcciones MAC de cada uno de los nodos descubiertos con direcciones MAC de los nodos supervisados y los nodos anteriormente descubiertos;
 60 un detector de duplicado de DNS (16, 203) configurado para comparar (640) los nombres DNS de cada uno de los nodos descubiertos con nombres DNS de los nodos supervisados y los nodos anteriormente descubiertos;
 un detector de duplicado de nombre (16, 204) configurado para comparar (650) los sysnames de cada uno de los nodos descubiertos con sysnames de los nodos supervisados y los nodos anteriormente descubiertos;

en el que la al menos una memoria y el código de programa informático se configuran adicionalmente, con el al menos un procesador, para provocar que el aparato al menos

5 determine (660) nodos duplicados que son duplicados de los nodos anteriormente descubiertos y/o los nodos supervisados basándose en el resultado de comparación del detector de duplicado de IP, el detector de duplicado de MAC, el detector de duplicado de DNS y el detector de duplicado de nombre, en el que, para la determinación, a cada uno de los nodos descubiertos se asigna una ID de nodo y la al menos una memoria y el código de programa informático se configuran adicionalmente, con el al menos un procesador, para provocar que el aparato al menos asigne un ÍndiceConcordancia a cada ID de nodo, 10 indicando el ÍndiceConcordancia una probabilidad de una concordancia entre el nodo descubierto y cualquiera de los nodos supervisados y los nodos anteriormente descubiertos.

7. El aparato de acuerdo con la reivindicación 6, en el que la al menos una memoria y el código de programa informático se configuran adicionalmente, con el al menos un procesador, para provocar que el aparato al menos descarte (435) los nodos duplicados. 15

8. El aparato de acuerdo con la reivindicación 6, en el que la al menos una memoria y el código de programa informático se configuran adicionalmente, con el al menos un procesador, para provocar que el aparato al menos asigne una prioridad a cada uno del detector de duplicado de IP, el detector de duplicado de MAC, el detector de duplicado de DNS y el detector de duplicado de nombre que determina un orden de ejecución. 20

9. El aparato de acuerdo con la reivindicación 6, en el que la al menos una memoria y el código de programa informático se configuran adicionalmente, con el al menos un procesador, para provocar que el aparato al menos determine los nodos duplicados ejecutando la siguiente fórmula para calcular un resultado final en cuanto a si un nodo descubierto es un duplicado: 25

$$d_1.\text{EsDuplicado}()*d_1.\text{Prioridad}+\dots+d_n.\text{EsDuplicado}()*d_n.\text{Prioridad}$$

30 donde d_1 es un primer detector de duplicado, d_2 es un segundo detector de duplicado, ..., y d_n es un $n^{\text{ésimo}}$ detector de duplicado, $d_n.\text{EsDuplicado}()$ es una función que representa la concusión del $n^{\text{ésimo}}$ detector de duplicado en cuanto a si el nodo descubierto es un duplicado y $d_n.\text{Prioridad}$ es una prioridad asignada a cada uno del $n^{\text{ésimo}}$ detector de duplicado.

10. El aparato de acuerdo con una cualquiera de las reivindicaciones 6 a 9, en el que la al menos una memoria y el código de programa informático se configuran adicionalmente, con el al menos un procesador, para provocar que el aparato al menos agrupe los nodos duplicados por ID de nodo y sume los ÍndicesConcordancia con el mismo ID de nodo. 35

11. El aparato de acuerdo con la reivindicación 6, en el que se asigna una ponderación a cada uno del detector de duplicado de IP, el detector de duplicado de MAC, el detector de duplicado de DNS y el detector de duplicado de nombre, en el que la ponderación indica la fiabilidad del resultado proporcionado por los respectivos detectores de duplicados. 40

12. Un programa informático, incorporado en un medio legible por ordenador, que comprende instrucciones que, cuando el programa informático se ejecuta mediante un procesador, provocan que el procesador efectúe las etapas de: 45

descubrir (600), mediante un aparato de supervisión de red, nodos en una red; 50 recoger (610) una lista de direcciones de protocolo de internet, IP, direcciones de control de acceso al medio, MAC, nombres de sistema de nombre de dominio, DNS, y sysnames para cada uno de los nodos descubiertos en la red; comparar (620) las direcciones IP de cada uno de los nodos descubiertos con direcciones IP de nodos supervisados y nodos anteriormente descubiertos; comparar (630) las direcciones MAC de cada uno de los nodos descubiertos con direcciones MAC de los nodos 55 supervisados y los nodos anteriormente descubiertos; comparar (640) los nombres DNS de cada uno de los nodos descubiertos con nombres DNS de los nodos supervisados y los nodos anteriormente descubiertos; comparar (650) los sysnames de cada uno de los nodos descubiertos con sysnames de los nodos supervisados y los nodos anteriormente descubiertos; y 60 determinar (660) nodos duplicados que son duplicados de los nodos anteriormente descubiertos y/o los nodos supervisados basándose en la comparación de las direcciones IP, direcciones MAC, nombres DNS y sysnames, en el que, para la determinación, a cada uno de los nodos descubiertos se asigna una ID de nodo y el proceso comprende además asignar un ÍndiceConcordancia a cada ID de nodo, indicando el ÍndiceConcordancia una probabilidad de una concordancia entre el nodo descubierto y cualquiera de los nodos supervisados y los nodos 65 anteriormente descubiertos.

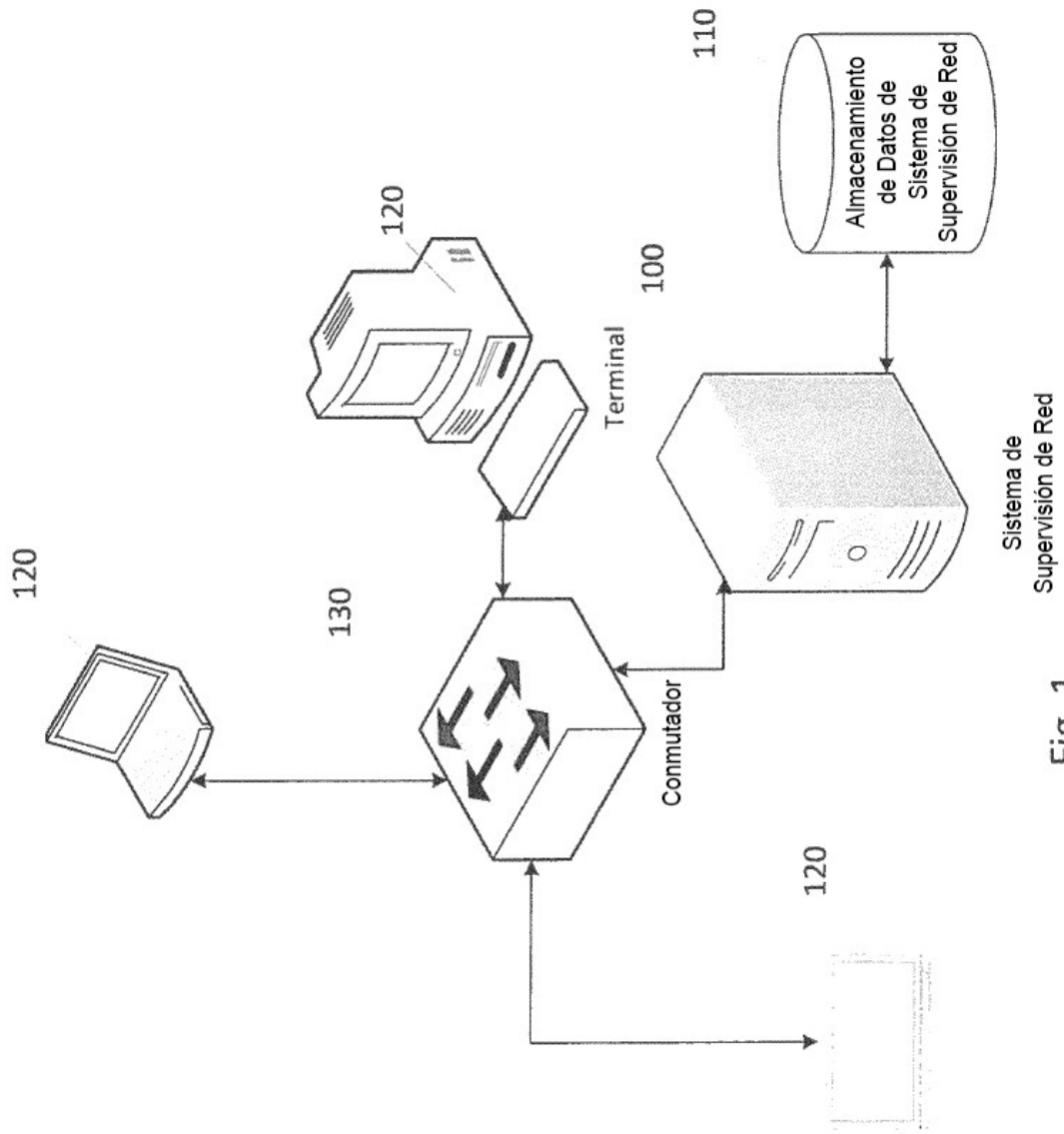


Fig. 1

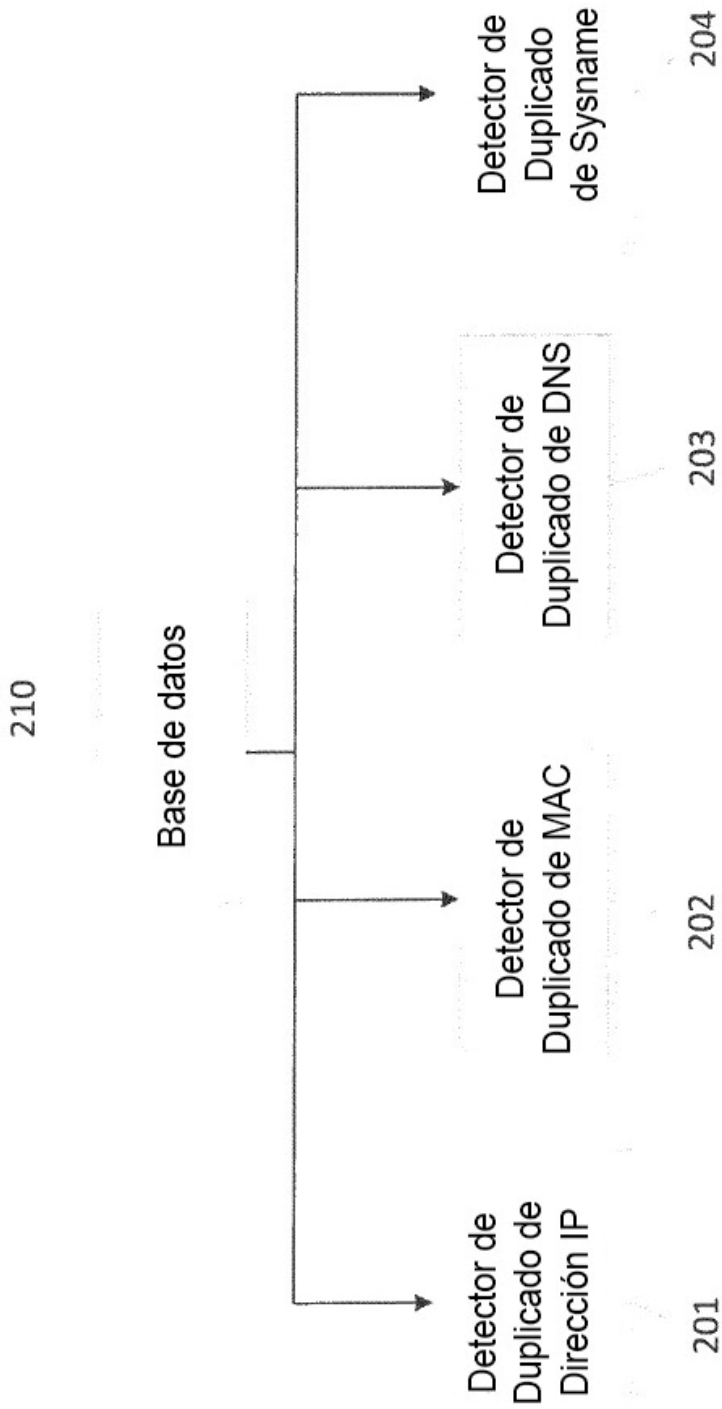


Fig. 2

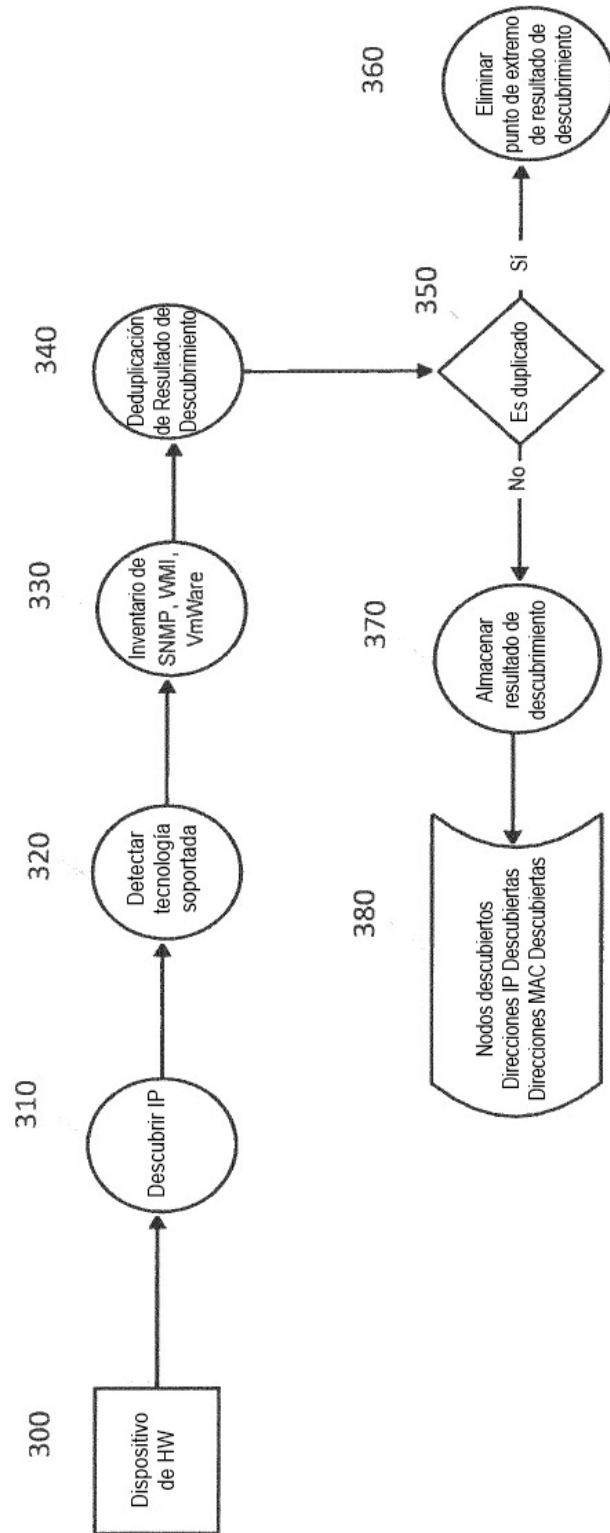


Fig. 3

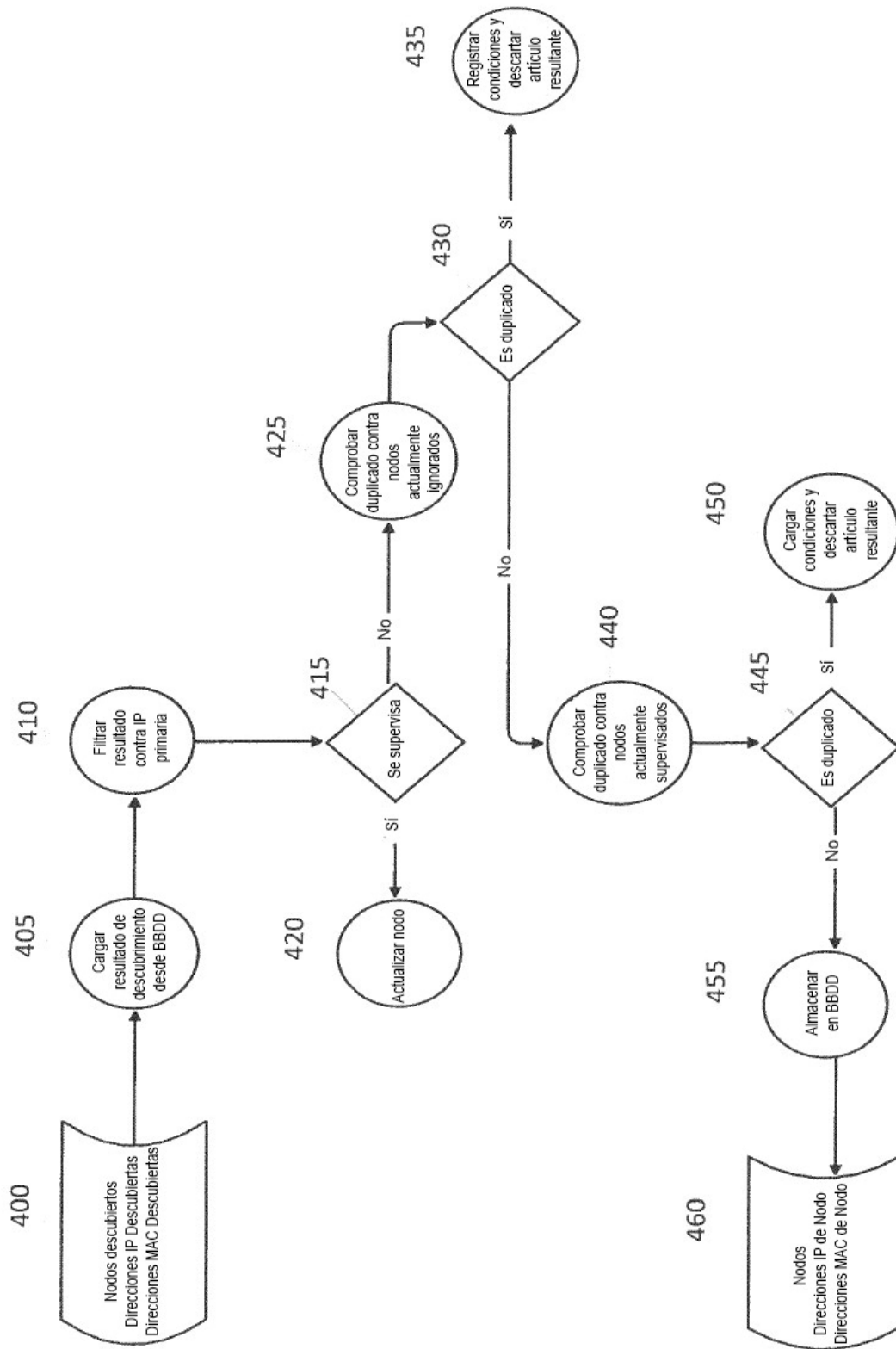


Fig. 4

10

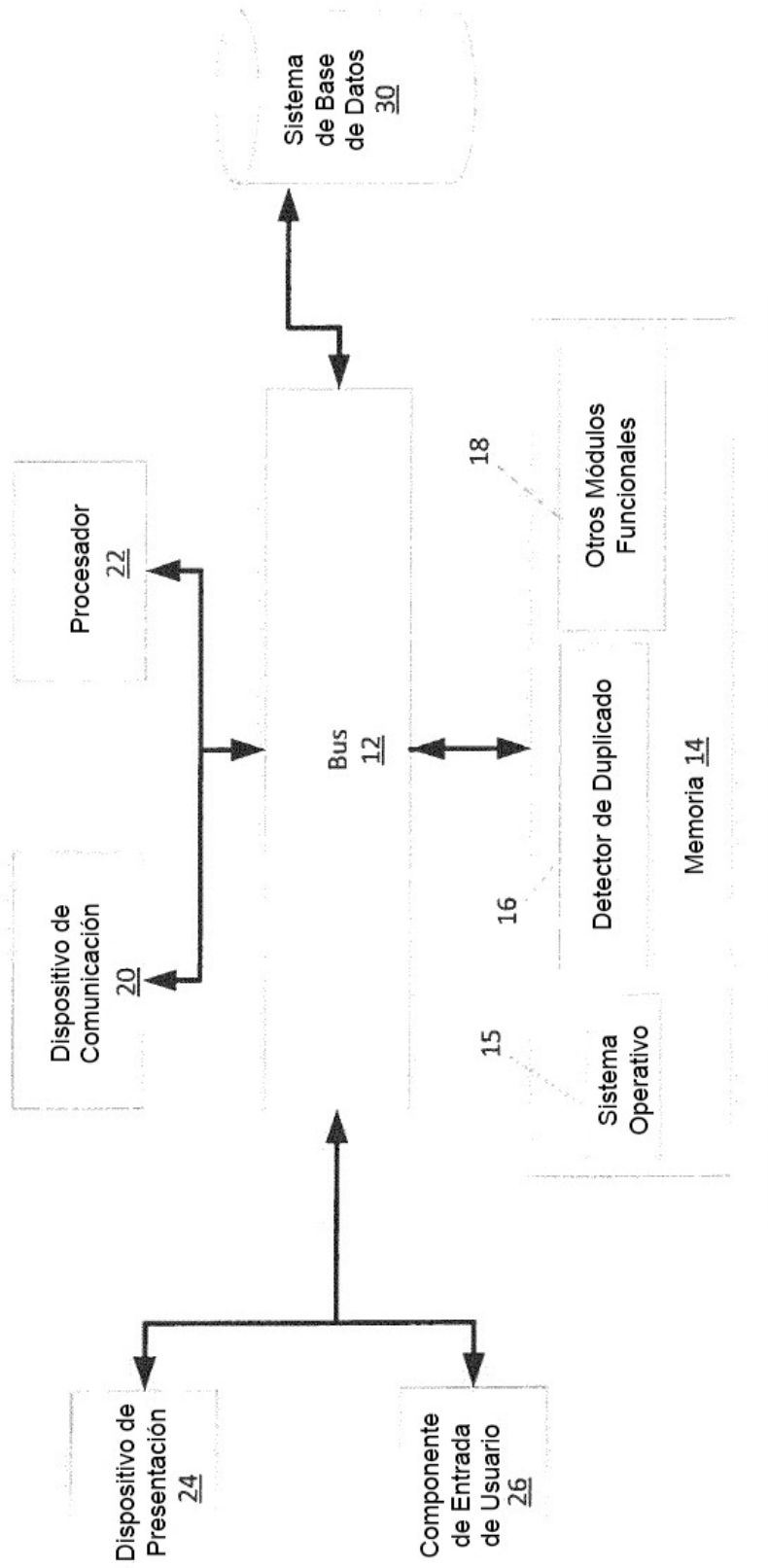


Fig. 5

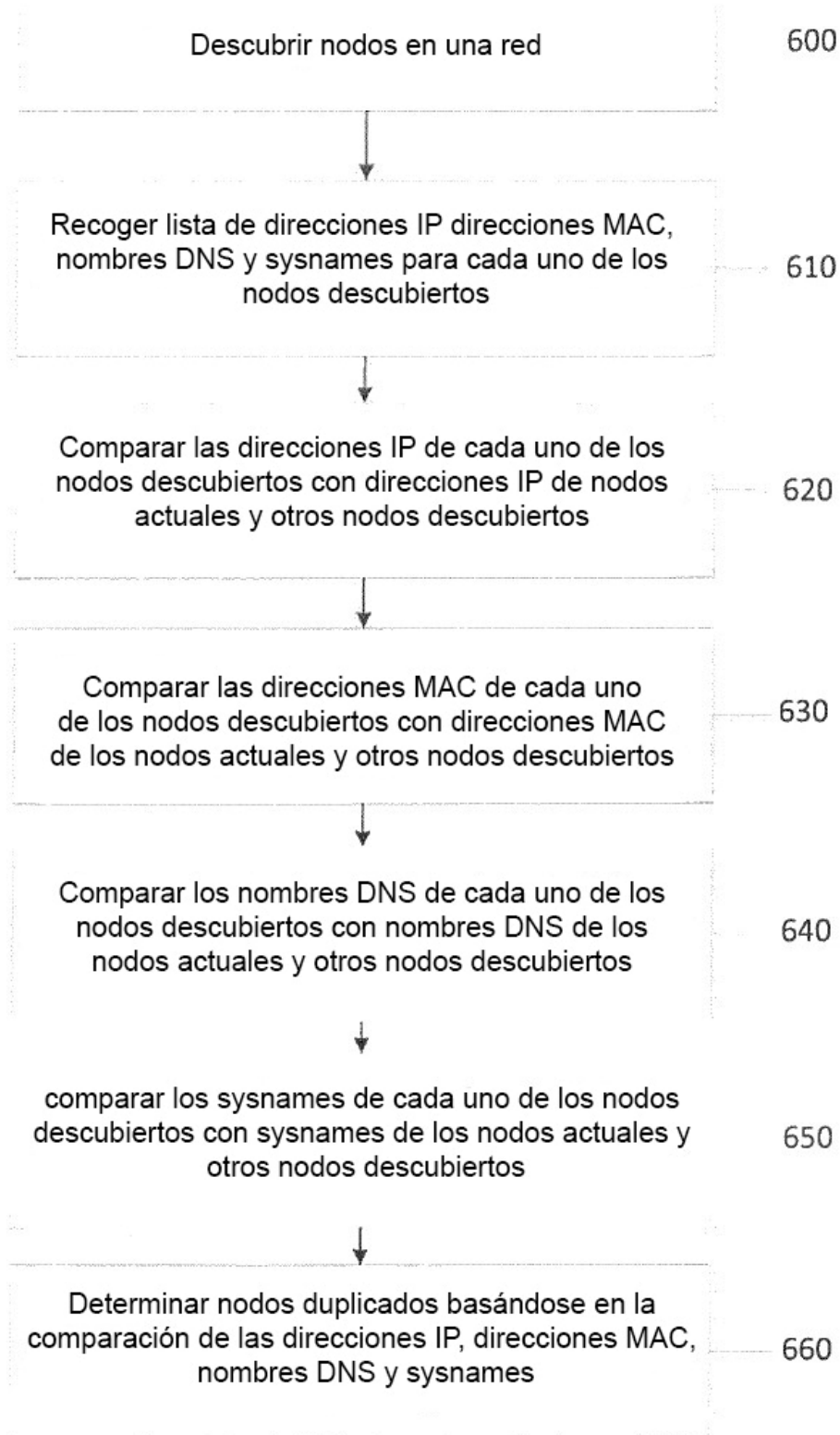


Fig. 6