

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 856 500**

51 Int. Cl.:

G06Q 99/00	(2006.01)
G06N 7/00	(2006.01)
H04L 29/06	(2006.01)
H04L 29/08	(2006.01)
G06F 21/31	(2013.01)
H04W 4/20	(2008.01)
H04W 12/12	(2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **29.01.2014 PCT/US2014/013553**
- 87 Fecha y número de publicación internacional: **23.04.2015 WO15057256**
- 96 Fecha de presentación y número de la solicitud europea: **29.01.2014 E 14853974 (5)**
- 97 Fecha y número de publicación de la concesión europea: **14.10.2020 EP 3058472**

54 Título: **Sistema y método para informar sobre agentes navegadores automatizados**

30 Prioridad:

18.10.2013 US 201314057730
02.12.2013 US 201314093964

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
27.09.2021

73 Titular/es:

WHITE OPS, INC. (100.0%)
111 West 33rd Street, 11th Floor
New York, NY 10001, US

72 Inventor/es:

KAMINSKY, DANIEL y
TIFFANY, MICHAEL J.J.

74 Agente/Representante:

SÁNCHEZ SILVA, Jesús Eladio

ES 2 856 500 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

Sistema y método para informar sobre agentes navegadores automatizados

5 Referencia cruzada a las solicitudes relacionadas

Esta solicitud de patente reivindica la prioridad de la Solicitud de Patente Provisional de Estados Unidos Núm. 61/732,368, presentada el 2 de diciembre de 2012. También es una continuación en parte de la solicitud de patente de Estados Unidos 14/057,730 presentada el 18 de octubre de 2013.

10 Campo de la invención

Esta invención se refiere al campo general del software de comunicaciones por Internet y tiene ciertas aplicaciones específicas para la evaluación analítica de las comunicaciones por Internet.

15 Antecedentes de la invención

Por una serie de razones, numerosas personas y organizaciones participan activamente a diario en el envío de tráfico malintencionado y automatizado a páginas web y otros destinos de Internet, y hacen que parezca que el tráfico es humano y no automatizado. Por ejemplo, la gran mayoría de los ingresos que se derivan actualmente del tráfico de Internet proviene de la publicidad pagada. Las empresas y las personas pagan por la colocación de anuncios en Internet, donde pueden ser vistos e interactuar con personas que puedan estar interesadas en conocer y comprar sus productos. Dado que estas interacciones publicitarias tienen lugar de forma electrónica y a distancia, es posible que aquellos interesados en capturar una porción de los ingresos que se gastan en publicidad en Internet empleen agentes de software automatizados para embaucar a quienes pagan por la publicidad. Esto se hace al hacer que parezca que los anuncios han sido vistos por humanos que pueden interesarse en un producto determinado, donde, en realidad, un anuncio determinado sólo ha sido visto o interactuado por softwares maliciosos, que existe sólo con el propósito de cometer tales actos de fraude.

30 Actualmente, existen sistemas y métodos pasivos que detectan diferencias de automatización, o robot, como, por ejemplo, si todo el contenido está cargado o si las tasas de solicitud coinciden con los navegadores legítimos. La detección de estas diferencias es útil desde la perspectiva del hardware de red – uno puede implementar el sistema en una red, no interferir con nada y recuperar datos. Sin embargo, estos datos no son necesariamente de alta calidad porque, por ejemplo, los usuarios humanos legítimos pueden tener patrones de acceso inusuales, las capas de almacenamiento en caché previenen solicitudes como podrían hacerlo los robots automatizados y, lo más importante, los robots están convirtiéndose cada vez más en navegadores completos, por lo que coinciden con muchas de estas métricas pasivas con bastante frecuencia.

40 Los documentos US2012/246293A1, WO2012/073233A1, WO2010/143152A2 representan la técnica anterior relevante.

Resumen de la invención

45 Durante el período de aprendizaje inicial, toda la actividad de navegación en una página (por ejemplo, clics del ratón) puede dividirse en grupos con base en su origen. Por ejemplo, las solicitudes de página que provienen de ordenadores en una red gubernamental protegida probablemente sean enviadas por humanos y se clasificarán como tales. Las solicitudes que provienen de direcciones IP que pertenecen a redes de robots conocidas tienen una baja probabilidad de ser interacción humana y se clasificarán en un grupo separado.

50 La recopilación de datos por parte del servidor de análisis es posible mediante fragmentos de código insertados (o inyectados) en el código de la página por el servidor web antes de que la página se envíe al navegador del usuario. Este código realiza la recopilación de datos sobre la interacción del usuario con la página web y transmite los datos recopilados al servidor de análisis a través de múltiples canales de comunicación.

55 En la etapa de detección de robot, los datos que se transmiten al servidor de análisis se verifican si coinciden con un patrón característico de interacción humana o un patrón de envío de robot automatizado. Los elementos típicos de un patrón de robot incluyen, pero no se limitan a, (1) interacción con elementos invisibles de la página, (2) propiedades faltantes de una interacción (por ejemplo, un clic del ratón), (3) temporización de interacción incorrecta (por ejemplo, una discrepancia entre la marca de tiempo del ratón hacia abajo y hacia arriba), (4) el comportamiento de la interfaz es atípico para los humanos (por ejemplo, el ratón se mueve a lo largo de una línea absolutamente recta), (5) la propiedad del elemento de página incorrecta debido al hecho de que un robot no pudo adivinar correctamente qué datos ingresará un navegador durante la carga de la página, (6) un conjunto de canales de comunicación disponibles no coincide con la característica establecida para el típico ordenador operado por humanos. Los resultados de la detección se proporcionan al cliente del sistema de análisis en tiempo real o, alternativamente, como un informe para un período de tiempo determinado.

Breve descripción de los dibujos

La Figura 1 ilustra un ejemplo del despliegue de la presente invención en un escenario típico de una página web.

La Figura 2 ilustra un ejemplo del proceso empleado por la presente invención para analizar el tráfico de Internet y determinar si un usuario dado es un ser humano o un agente automatizado.

La Figura 3 ilustra el proceso general de recopilación de datos de la presente invención.

Descripción detallada de la modalidad preferida

Definiciones

HTML (Lenguaje de Marcado de Hipertexto). El lenguaje de programación principal que se usa para crear, transmitir y mostrar páginas web y otra información que puede mostrarse en un navegador de Internet.

HTTP (Protocolo de Transferencia de Hipertexto). El protocolo estándar de cliente-servidor de la Red Mundial que se usa para el intercambio de información (como documentos HTML y solicitudes de clientes para dichos documentos) entre un navegador web y un servidor web. HTTP incluye varios tipos diferentes de mensajes que pueden enviarse desde el cliente al servidor para solicitar diferentes tipos de acciones del servidor. Por ejemplo, un mensaje "GET", que tiene el formato GET <URL>, hace que el servidor devuelva el objeto de contenido ubicado en la URL especificada.

Medios para detectar. Este término incluye, pero no se limita a, la inserción activa de un fragmento de código en el código HTML de una página antes de que la página se envíe a un navegador o el monitoreo pasivo de un comportamiento normal. Las inserciones activas de código pueden ser estáticas, lo que significa que contienen completamente la cantidad de material necesario para realizar un análisis completo de acuerdo con la presente invención. O las inserciones activas pueden ser dinámicas, lo que significa que se comunican con la red de detección para recuperar códigos adicionales o claves de descripción, lo que resulta en la compilación de datos estadísticos adicionales. Si bien los siguientes ejemplos hablan de inserción activa, deben leerse para incluir la posibilidad de monitoreo pasivo como un medio alternativo de detección.

Fragmento de código. Aunque el término "fragmento" puede implicar una pequeña porción, este término no debe interpretarse como una limitación - la cantidad de código que se inserta puede variar en tamaño. El fragmento de código está modularizado, con fragmentos de elementos que incluyen, pero no se limita al, análisis DOM del navegador, análisis de temporización flash, captura de eventos del ratón, etc. La capacidad de mutar dinámicamente un fragmento de código dado permite la correlación de tipos y/o clases de robot con los flujos financieros del cliente, por ejemplo, mediante la integración de parámetros ("dimensiones de análisis") de un cliente determinado en un fragmento determinado. La presente invención describe un modelo de sondeo activo para la recopilación de métricas cualitativas que evalúan la actividad de navegación impulsada por humanos frente a la impulsada por agentes automatizados (es decir, impulsada en caliente) a través de una red informática. A través de este modelo de sondeo activo, puede implementarse una reserva mucho más profunda de diferencias entre los dos tipos de actividad (en comparación con las diferencias que se usan en el estado actual de la técnica). A diferencia de los métodos pasivos de recopilación de contenido que ya existe en una red y el contenido que se envía a los sistemas existentes (es decir, los métodos actuales para la detección de robots), el método descrito en la presente descripción carga activamente código adicional y envía contenido adicional por el cable a ubicaciones diferentes y nuevas ("sondeo activo"). JavaScript (JS) y Flash, por ejemplo, pueden probarse activamente por el sistema y método reivindicado para detectar la actividad del robot y elaborar un informe con base en métricas de rendimiento cualitativas.

El sistema y el método reivindicado asumen que los usuarios humanos legítimos, en general, tienen JavaScript y otras tecnologías de secuencias de comandos activas habilitadas, que incluyen pero no se limitan a Flash, y están usando navegadores web completos. Como tal, un robot que no sea un navegador simplemente no podrá ejecutar ninguna consulta que dependa en absoluto de JavaScript. La trampa establecida para los atacantes potenciales es que, para evadir este mecanismo de detección extremadamente confiable, ahora deben emular todas las partes del navegador. Y debido a que un entorno JavaScript real, así como otras tecnologías de secuencias de comandos, tiene una cantidad infinita de propiedades que pueden probarse, el atacante debe emular todas las propiedades potencialmente probadas. Por lo tanto, se expone información previamente desapercibida y las discrepancias resultantes. Por ejemplo, cuando se falsifica un evento de ratón, una marca de tiempo asociada con ese evento puede desaparecer; un campo auxiliar puede establecerse en un valor único e incorrecto; o una tasa de eventos del ratón es demasiado estable o demasiado inestable. Algunos ejemplos de propiedades que pueden probarse incluyen, pero no se limitan a: (1) la relación precisa de los eventos del ratón que se ven en una página (por ejemplo, un clic asociado con un movimiento de ratón hacia arriba o hacia abajo, acuerdo entre las dos marcas de tiempo asociado con cada evento del ratón, como se discutió anteriormente, etc.); (2) la tasa de actualización de Flash (por ejemplo, por segundo) y la confiabilidad de sus llamadas; (3) operación de etapas Flash en todas las ubicaciones de operación (por ejemplo, al operar en sincronía); y (4) la velocidad para completar una actualización gráfica (por

ejemplo, a un elemento <CANVAS>), que podría indicar el tipo de hardware que se usa o la actualización activa de una pantalla de usuario real.

La presente invención permite diferenciar los agentes automatizados maliciosos de los humanos mediante la recopilación y procesamiento de elementos de la interacción de un usuario dado con una página web que ocurre después de que el usuario ha cargado una página web, y al comparar esos elementos con resultados de referencia que se extraen de un grupo de control. Esto se logra en parte al colocar ciertos elementos dentro del código de una página web antes de que un usuario determinado la cargue, de manera que esos elementos puedan evaluarse después de que el usuario haya cargado esa página web.

Los elementos monitoreados y evaluados se dividen en dos clases principales de datos: (1) contenido que existe (o está ausente, es decir, no existe) al cargar la página, y (2) contenido que se genera con el tiempo (o temporización) como la página persiste en forma potencialmente usable. El contenido que existe en la carga de la página incluye bits, o partes de código, que son accesibles o visibles aunque no deberían serlo. Este contenido consta de elementos JavaScript ("DOM") que existen (o no existen) debido a la forma en que se aloja el navegador. Por ejemplo, si los carga un usuario humano, algunos bits serían inaccesibles por razones de seguridad u otras razones; sin embargo, si se carga por un agente automatizado o en caliente, los mismos bits serían accesibles). Por otro ejemplo, los agentes automatizados también inyectan de forma constante y activa configuraciones específicas de robot de maneras diferentes al comportamiento del navegador o del sitio que está monitoreándose. En general, los aspectos de un intérprete de comandos (por ejemplo, Internet Explorer, Firefox, Safari, Chrome) están expuestos al entorno JavaScript en un motor (por ejemplo, Trident, Gecko, Webkit) y los robots, que son intérpretes de comandos en sí mismos, exponen demasiada información o demasiado poca información, y las discrepancias se capturan por el modelo de sondeo activo de la presente invención. Estas características capturadas incluyen, pero no se limitan a, el cumplimiento de los estándares HTML5, patrones en el manejo de errores (incluida información sobre el idioma al que se traducen los errores) y elementos del navegador que se inyectan por el intérprete de comandos del navegador en lugar del objeto nativo (diferentes objetos se inyectan o no se inyectan con base en el receptor, que podría ser, por ejemplo, Internet Explorer o una estructura de agente automatizado (es decir, robot)).

La segunda clase de datos, el contenido que se genera con el tiempo (o temporización), generalmente se refiere a elementos que varían debido a la interacción con un usuario humano. Estos pueden ser eventos que toman cantidades incorrectas de tiempo, entre sí, porque no hay un ser humano real para quien se realizan los eventos. Los ataques de temporización funcionan contra más que solo sistemas criptográficos. A menudo es más rápido, pero a veces mucho más lento, expresar el resultado de una operación del navegador (de las cuales hay cientos de miles) cuando no hay pantalla para actualizar ni usuario para informar. Por ejemplo, los mensajes de error pueden suprimirse o el hardware de gráficos puede notar que ningún píxel requiere actualización. Al medir las diferencias de temporización absolutas y relativas, los robots se exponen al sistema y al método reivindicado. Las pruebas se generan en un número infinito de tales diferencias, alojadas con poca frecuencia (dado que el propósito de los robots es operar a escala, esto no tiene que ocurrir a menudo), y por lo tanto, un desarrollador atacante enfrenta el obstáculo de falsificar credenciales que no necesariamente sabe de antemano.

La presente invención recopila datos que se relacionan con la interacción de cualquier usuario con una página web después de que se haya cargado. Estos datos incluyen, pero no se limitan a, la actividad del ratón (dónde se encuentra el ratón, el número de actualizaciones por segundo, la geometría del movimiento del ratón, los datos auxiliares a los datos de eventos del ratón, es decir, los metadatos asociados con un clic del ratón, desplazarse hacia arriba, desplazarse hacia abajo o desplazarse sobre la correlación entre eventos del ratón, etc.), datos faltantes cuando un evento se sintetiza incorrectamente, actividad del teclado, datos del acelerómetro, eventos de desplazamiento, tiempo promedio de lectura y visita, tasa de actualización de la página (la tasa de animación tiene una fuerte correlación con visibilidad de una página) y protocolos de red y estándares web compatibles (los robots pueden romper las vías de comunicación).

Ambas clases de métricas de rendimiento obligan a un sistema dado a seguir rutas de código que difieren en dependencia si la interacción del navegador es automatizada o humana. Los datos de medición de temporización son detectables y diferenciables porque las operaciones aún se completan, solo toman más o menos tiempo, en dependencia del tipo de usuario. También existe la posibilidad de superposición entre las familias (es decir, clases de datos). Por ejemplo, un control de seguridad diferente puede fallar (o no fallar) bajo automatización, al producir la misma reacción a un código de cliente, pero un 10 % más lento. Además, repetir la prueba muchas veces permite amplificar incluso las diferencias de temporización muy pequeñas hasta el punto de una capacidad de distinción confiable.

El proceso con respecto a una sola métrica dada puede trazarse generalmente mediante las siguientes etapas: (1) obtener la diferencia; (2) medir en el sistema cliente (o primer servidor, servidor de análisis), en circunstancias ampliadas si es necesario; (3) enviar al servidor de informes (o segundo servidor), ofuscado si es posible; y (4) adjuntar el hallazgo al identificador de sesión, creando así una unidad de datos de interacción del usuario, donde una compilación de unidades de datos del usuario forma un informe.

El siguiente es un ejemplo de ejecución que usa el método actualmente reivindicado: Se ejecutan pequeños

fragmentos de JavaScript. Si los datos resultantes son grandes, se completa algún análisis y compresión del lado del cliente (por ejemplo, en una métrica de 'cuántas veces se llama a Flash por segundo', las actualizaciones se envían solo cuando la tasa de llamadas cambia en más del 10 %, y las actualizaciones que se envían pueden contener algunas estadísticas precalculadas por un cliente determinado). El envío de comentarios se vuelve ortogonal a la recopilación. Independientemente del subsistema del que se extraigan los datos, uno de los objetivos finales es la minimización y agregación de las devoluciones de datos.

Los elementos de datos de interacción del usuario se comparan con los resultados de referencia extraídos de un conjunto de tres grupos de control diferentes: (1) aquellas interacciones que se cree que se realizan por agentes automatizados o robots, (2) aquellas interacciones que se cree que se realizan por un ser humano, y (3) aquellas interacciones que no están claras en cuanto a si las realiza un humano o un robot. Los mejores grupos de control para conjuntos de elementos de verdadera interacción humana surgen de los navegadores web impulsados desde ubicaciones autenticadas en lugares sin motivo de fraude publicitario. Los mejores grupos de control para conjuntos de elementos de comportamiento de robots surgen de "zoológicos de robots" u otras redes de agentes automatizados.

Antes de que comience el proceso de diferenciación, debe insertarse un fragmento de código individualizado en el código HTML de una página web determinada. Cuando este fragmento de código está presente en el código de una página web determinada y se accede a esa página, las métricas de rendimiento se envían a servidores de análisis remotos a través de publicaciones HTTP asincrónicas. Estas métricas evalúan el comportamiento y el rendimiento de la entidad que vio o está viendo la página web dada y cómo se cargó esa página. El fragmento de código se inyecta como JavaScript junto con un anuncio u otro evento de carga de secuencia de comandos. Como Internet se compone de muchas de esas cargas (o inyecciones), esta invención crea simplemente una más. Por ejemplo, una métrica de rendimiento con base en un evento de ratón puede recopilarse de la siguiente manera: (1) Los controladores y oyentes se registran para un evento de ratón; (2) El controlador recibe las distintas marcas de tiempo y valores asociados con el evento del ratón; (3) El sistema luego emite las marcas de tiempo y valores sin procesar, o un resumen de los mismos, a través de la red. Si no se registra ningún oyente, sería imposible recuperar estos datos del tráfico ambiental.

Las métricas de rendimiento para varios visitantes de una página web determinada que contiene el fragmento de código, así como las de todas las páginas web que contienen fragmentos de código similares, se compilan y agregan por los servidores de análisis remotos en métricas que pueden informarse, que a su vez están disponibles para el operador de una página web determinada en varios medios de presentación de informes, que incluyen, pero no se limitan a, paneles de control HTML interactivos protegidos con contraseña, documentos de hojas de cálculo exportables e informes en PDF y por correo electrónico basados en suscripción, y pueden usarse en tiempo real para controlar el acceso a una página web determinada.

Las métricas de rendimiento que pueden informarse incluyen, pero no se limitan a, el origen y el destino de un visitante, la probabilidad de que el visitante fuera un agente automatizado o un ser humano, y una variedad de variables que identifican información, como puntos de datos publicitarios, que incluyen, pero no se limitan a, el código específico de la campaña publicitaria, el medio publicitario, el ID de la fuente y el proveedor de publicidad.

Estas métricas son evaluadas de tal manera por los servidores de análisis remoto que la información presentada al operador de una página web determinada que incluye un fragmento de código se presenta con una evaluación cualitativa de si una determinada visita a esa página web fue o no fue realizada por un agente automatizado. Este proceso de evaluación implica lo siguiente: el fragmento de código envía "eventos de emisión" desde varios "complementos". Estas emisiones (es decir, "eventos de emisión") se envían a través de una variedad de canales de red, no todos los cuales están siempre disponibles. Los canales que se usan actualmente son etiquetas , XMLHttpRequests con CORS (Intercambio de Recursos de Origen Cruzado) y eventos de Publicaciones de Formulario IFrame. Inicialmente, se usan las Publicaciones de Formulario IFrame, ya que son las más compatibles. En segundo lugar, si CORS es compatible, el sistema puede actualizarse a CORS. Otros canales incluyen WebSockets y Mismo Dominio XMLHttpRequest (que requiere el uso de un iframe local que se configura para hablar entre dominios, a través de un kit de herramientas como EasyXDM).

Además, el proceso computacional que se requiere para determinar las métricas de rendimiento anteriores y, en última instancia, evaluar si un visitante es automático o humano, puede implementarse mediante procesamiento por lotes o mediante procesamiento de flujo. El procesamiento por lotes puede ser más eficiente y puede recopilar métricas en varios eventos. El procesamiento de flujo puede escalar mejor que el procesamiento por lotes, pero no puede, por ejemplo, usar datos futuros para informar impresiones pasadas de normalidad (porque, en el momento de la decisión, el evento futuro aún no ha ocurrido). Con el procesamiento de flujo, puede lograrse una evaluación casi en tiempo real de un usuario determinado. Por lo tanto, aunque las métricas de normalidad están determinadas solo por el pasado, el procesamiento de flujo permite el uso de identificadores de transacciones incrustados en un evento de medición particular para evaluar, dentro de los treinta segundos posteriores a la última interacción de un usuario dado, si ese usuario era o no un robot o un humano.

La Figura 1 da un ejemplo de cómo puede implementarse la presente invención en un escenario típico de una

página web. Primero, se inserta un fragmento de código que contiene una identificación única en la página web **100**. Un usuario (humano o automatizado) solicita la página web que contiene el fragmento de código **101**. El usuario carga la página web que contiene el fragmento de código **102**. Y a medida que el usuario continúa navegando normalmente **103**, los datos relacionados con la interacción del usuario con la página web se envían al servidor de análisis **104**, donde el servidor de análisis analiza además los datos del usuario de forma cualitativa **105**.

La Figura 2 muestra una aplicación de ejemplo del proceso repetible que se emplea por la presente invención para analizar el tráfico de Internet. El proceso ilustrado consta de las siguientes etapas: Declarar o recopilar el identificador del cliente (es decir, el cliente), identificador de pares (es decir, contra quién el cliente le gustaría probar, por ejemplo, editor, ubicación del anuncio, intercambio secundario, etc.), y el identificador de la transacción (es decir, la vista del anuncio particular) **200**; Cargar el GS del Cargador **201** desde el servidor de análisis; carga de secuencia de comandos del GIF de Destello de Señal **202** desde el servidor de análisis; cargar GIF de Destello de Señal **203** desde el servidor de análisis; cargar el monitor humano (pagespeed.js) **204** desde el servidor de análisis; la carga del Informe se realizó correctamente, en el estado "init" con todas las métricas disponibles al servidor de análisis **205**; Si se detecta un acto humano **206**, emitir inmediatamente un segundo informe (estado "primero") **207**, esperar seis (6) segundos **208** y emitir un informe final (estado "verificación de estado") **209**; Si no se detecta ningún acto humano **210**, las etapas **207**, **208** y **209** no ocurren; Realizar un análisis cualitativo de métricas e informes disponibles, si los hay **211**; e Informar un puntaje cualitativo para el ID de cliente (sesión) **212**.

El proceso descrito anteriormente y que se ilustra por la Figura 2 es un ejemplo del proceso más general que se emplea y se reivindica por la presente invención. Específicamente, este proceso más amplio, que se muestra en la Figura 3, ocurre de la siguiente manera: Primero, los identificadores de clientes, pares y transacciones se recopilan **300**; A continuación, estos identificadores se incrustan en una sonda activa, donde la sonda activa (1) recupera un estado adicional del ambiente de ejecución del cliente y (2) transmite datos a través de múltiples canales **301**; En tercer lugar, estas características probadas activamente se miden frente a huellas de robots conocidas (es decir, características de robots) **302**. Las dos clases principales de características que se prueban y se analizan son (1) qué canales o información está disponible y/o ausente (nota: la presencia o ausencia de un canal es, por sí misma, una fuente de huella de robot) y (2) el tiempo que tarda en probarse las propiedades/características. El análisis realizado mide el grado/cantidad de automatización, así como el grado-cantidad de la verdadera interacción humana. Finalmente, se emiten informes (1) al usuario/cliente, informando sobre el porcentaje de automatización/caliente **303**, de acuerdo con las dimensiones dadas en el identificador de pares, y (2) al servidor para un análisis adicional y características adicionales para una mayor generación de huellas de robots **304**.

A continuación, se exponen algunos ejemplos de cómo pueden evaluarse métricas específicas para lograr resultados que pueden informarse:

Evaluación de la Ubicación: Al usar los datos que se recopilan como se establece anteriormente, se ha inventado un método para evaluar probabilística, estadística y directamente la ubicación de los clics en una página web determinada que se ejecutan durante una visita determinada a una página web y, al hacerlo, evaluar, o contribuir a un modelo estadístico con el fin de evaluar si esa visita determinada fue o no realizada por un agente automatizado.

Evaluación de la Temporización entre Clics: Al usar los datos que se recopilan como se establece anteriormente, se ha inventado un método para evaluar probabilística, estadística y directamente la temporización entre los clics en una página web determinada durante una visita determinada, así como para usar dicha temporización entre clics para identificar o determinar información sobre un usuario o clase de usuarios determinados. Dicha temporización puede proporcionar una "huella digital" del escritorio de un usuario determinado y/o patrones de navegación en Internet con el fin de evaluar o contribuir a un modelo estadístico que se diseña para evaluar si una visita determinada fue o no realizada por un agente automatizado, así como para muchos otros propósitos.

Evaluación de la Temporización entre Clics del Escritorio Remoto y la VPN: Al usar los datos que se recopilan como se establece anteriormente, se ha inventado un método para realizar la Evaluación de la Temporización entre Clics, incluso si una sesión de navegación determinada atraviesa una red privada virtual y/o una conexión de escritorio remoto basándose en el hecho que los comandos de ratón, teclado y clic deben transmitirse a través de dichas conexiones a una velocidad de lectura fija.

Detección de Agentes Automatizados Móviles Relacionados con el Movimiento y el Estado: Al usar los datos que se recopilan como se establece anteriormente, se han inventado varios métodos para determinar si una sesión de navegación determinada que se origina o parece originarse desde un navegador o aplicación que se ejecuta en un dispositivo móvil, como un teléfono inteligente o tableta, se lleva a cabo total o parcialmente por un agente automatizado. Por ejemplo, HTML5 permite que las lecturas del giroscopio y del acelerómetro se tomen "sin hacer clic" o sin que el usuario se comprometa activamente con una página web, y la información de desplazamiento puede leerse de manera similar. La mera presencia de información como esta, relacionada con la posición del dispositivo móvil en el espacio, y el compromiso del usuario con la interfaz del dispositivo móvil, determina si un humano está presente o no. Los cambios en información como esta, y la naturaleza de dichos cambios, pueden reflejar el entorno preciso en el que el dispositivo reivindica estar en, y la evaluación de dicha información, su presencia, ausencia o naturaleza cambiante, puede usarse con el propósito de evaluar o contribuir a un modelo estadístico que se diseña para evaluar si una visita determinada fue o no realizada por un agente automatizado, así

como para muchos otros propósitos.

5 Detección de Agentes Automatizados Móviles Relacionados con el IP y Geolocalización: Las metodologías establecidas anteriormente pueden complementarse adicionalmente mediante la evaluación de la dirección IP, la supuesta geolocalización y otros datos más estáticos relacionados con un dispositivo determinado y su usuario, tanto por sí mismos como en referencia a los datos recopilados en la Detección de Agentes Automatizados Móviles Relacionados con el Movimiento y el Estado, con el fin de evaluar o contribuir a un modelo estadístico diseñado para evaluar si una visita determinada fue o no realizada por un agente automatizado, así como para muchos otros fines.

10 Detección de Agentes Automatizados Móviles Relacionados con el IP y Geolocalización basada en el tiempo: La información de Detección de Agentes Automatizados Móviles Relacionados con el IP y Geolocalización establecida anteriormente puede evaluarse más a fondo durante períodos de tiempo prolongados y compararse con otros datos similares, con el fin de evaluar o contribuir a un modelo estadístico diseñado para evaluar si una visita determinada fue o no realizada por un agente automatizado, así como para muchos otros fines.

15 Ocultación y Separación de Datos: Quizás el mecanismo más eficiente para implementar código con el fin de determinar si una sesión de navegación determinada está realizándose por un agente automatizado, así como para realizar muchos otros tipos de evaluaciones útiles de eventos de navegación web, es causar una página web para, a su vez, hacer que se realice un procesamiento evaluativo en el ordenador u otro dispositivo que de hecho esté navegando, y una vez que se complete dicho procesamiento, transmitir sus resultados a una máquina remota para una evaluación adicional. En lugar de ser máximamente eficiente, se ha inventado una metodología que, aunque menos eficiente, es más segura y menos probable de detectarse, en donde una variedad de métricas, útiles para el propósito actual, pero también útiles para una serie de otros propósitos analíticos normales, se recopilan y transmiten al servidor remoto para su evaluación. Por lo tanto, se crea incertidumbre en cuanto a qué aspectos
20 específicos de los datos están evaluándose realmente y con qué propósito, y aquellos actores maliciosos que se involucran en la creación y el uso de agentes de navegación automatizados tienen menos probabilidades y
25 requerirán más recursos para determinar que tal evaluación está teniendo lugar.

30 Evaluación Diferencial de la Prestación: Además de evaluar la interacción del usuario, también es posible evaluar cuánto tardan en ejecutarse varias acciones. Cuando un ser humano está en el bucle, es necesario que un navegador web active ciertos aspectos del hardware de un dispositivo informático, que incluyen el hardware de gráficos, el hardware de sonido y similares. La cantidad de tiempo para completar ciertas acciones depende de si dicho hardware está realmente activado y en qué grado (por ejemplo, si la acción gráfica es opaca o semitransparente). Ciertos factores diferencian aún más la cantidad de tiempo que se tarda, como si el navegador
35 debe "rediseñar" la página o no, lo que resulta en una secuencia predecible de eventos de redibujado. Esta cantidad de tiempo varía con base en la naturaleza de la pantalla y, lo que es más importante, puede usarse para diferenciar entre una pantalla no completada (un "búfer de fotografías virtual") o una pantalla real.

40 Evaluación de inestabilidad: La cantidad de "inestabilidad" (en oposición al tiempo absoluto) observada es una indicación adicional de si un sistema determinado está realizando una tarea determinada en primer plano o en segundo plano.

45 Análisis de Distribución de Tiempo de la VM: Es posible determinar si la Distribución de Tiempo de la Máquina Virtual está produciéndose mediante la evaluación de los retrasos de la prestación (es decir, mediante la cuantificación de los potenciales de tiempo, como puede verse a través de llamadas repetidas a temporizadores de milisegundos en JavaScript).

50 Validación de Caché: Es posible usar el comportamiento de las cookies y cachés del navegador web, especialmente a lo largo del tiempo, para diferenciar entre navegadores humanos y automatizados, especialmente si un navegador se utiliza en muchos destinos.

55 Hay muchas aplicaciones para la invención actualmente reivindicada. En una aplicación, la tecnología actual se integra con la lucha contra el fraude financiero (en un contexto de "envío de dinero" o "pago del carrito de la compra"). Otra aplicación de la presente invención es para un auditor de registro preCAPTCHA. Cabe señalar que el sistema reivindicado no bloquea directamente un registro; en su lugar, marca las cuentas que los sistemas CAPTCHA no están notando o detectando. La invención reivindicada opera como una métrica independiente. También funciona como un sistema excelente para encontrar malware en las redes empresariales internas, ya que la mayoría de las intranets usan sitios internos que los atacantes exploran de forma remota. El sistema puede detectar que los atacantes no son en realidad los usuarios que reivindican ser, incluso si y especialmente si se conectan a
60 través de un túnel a través de una máquina en la red corporativa.

A continuación se exponen ejemplos adicionales de otras aplicaciones generales ilustrativas de la presente invención, aplicables a una amplia gama de campos e industrias:
65 Evaluación del Compromiso: los datos que se recopilan como se establece anteriormente son especialmente útiles como herramienta para determinar si un agente automatizado está llevando a cabo una sesión de navegación determinada. Sin embargo, este no es su único uso. Los datos que se recopilan por cada una de las metodologías

establecidas en la presente descripción también pueden usarse cuando un navegador está siendo manejado por un ser humano y no un agente automatizado para determinar cómo ese usuario interactúa con una página web y sus diversos aspectos, lo que resulta en una medida del compromiso de ese usuario con esa página web y sus diversos aspectos, tanto en una sesión de navegación determinada, como en comparación con sesiones de navegación anteriores y futuras.

Huella de Robot: Diferentes agentes automatizados se exponen de diferentes maneras. La evaluación de la información expuesta por diferentes agentes automatizados y que se recopila como se establece anteriormente, y/o que se recopila por cualquier número de otros métodos, como direcciones IP, modos de falla, nivel de soporte de JavaScript, permite su comparación y las comparaciones de las firmas de todos los agentes evaluados. Dichas "Huellas de Robot" pueden usarse para evaluar tendencias en el uso de agentes automatizados, para rastrear su desarrollo y propagación, y para cualquier número de otros fines.

Evaluación de Errores del Navegador: La información que se proporciona por los métodos establecidos anteriormente en relación con los errores del navegador puede usarse de manera efectiva para determinar si un agente automatizado está llevando a cabo una sesión de navegación determinada. Por ejemplo, es posible extraer el exceso de métricas y provocar intencionalmente errores de JavaScript de manera que las respuestas de error que se generan puedan usarse para distinguir entre agentes automatizados y navegadores humanos. Cuando falla un comando, esta falla se detecta, dentro de una construcción de captura/prueba. Esta información se captura por el JavaScript de la invención actualmente reivindicada en lugar de transmitirse a la consola del desarrollador. Supongamos, por ejemplo, que un navegador es un navegador de habla china pero oculta el hecho de que hablan chino. Los errores de navegación que se detectan por el sistema y método actuales seguirán estando en ese idioma (es decir, chino).

Evaluación A-B: Se sabe que diferentes campañas tienen diferente efectividad en diferentes audiencias. Sin embargo, los agentes automatizados no son impulsados por los mismos factores que los seres humanos y no responderán a diferentes campañas de la misma manera que lo harán los seres humanos. Cuando la tecnología establecida en la presente descripción se implementa en diferentes campañas publicitarias, la comparación de diferentes respuestas por diferentes fuentes de tráfico de navegación puede usarse como un mecanismo activo para detectar o complementar la detección de comportamiento automatizado. Esta comparación sigue siendo eficaz incluso cuando se usan seres humanos en lugar de agentes automatizados con el fin de realizar fraudes publicitarios.

Evaluación de Firma Estocástica: Las metodologías de detección de agentes automatizados establecidas en la presente descripción no necesitan estar expuestas en cada página web o cada carga de una página determinada, ni es necesario usar los mismos mecanismos cada vez o todo el tiempo. La mutación del JavaScript implementado, tanto en ubicación como en estilo, aumenta significativamente el costo de éxito de los operadores de los agentes automatizados y limita su capacidad para desarrollar e implementar contramedidas efectivas a los métodos establecidos en la presente descripción.

Evaluación en términos de Costo por Humano: En lugar de evaluar el tráfico web y específicamente las campañas publicitarias en términos de métricas como costo por clic, costo por mil clics o costo por acción, la presente invención permite y contribuye a la evaluación de dicho tráfico en términos de una métrica mucho más significativa: el costo por humano ("CPH"). En lugar de medir los clics u otros eventos que pueden o no generarse por un agente automatizado, la evaluación de CPH permite una determinación mucho más significativa de la efectividad de las cantidades gastadas para atraer tráfico a una página web determinada. El CPH es una métrica mejor y más significativa porque el objetivo final de la publicidad en línea no es ofrecer "impresiones" de por sí, sino mostrar impresiones publicitarias específicamente a seres humanos. El CPH refleja el costo de llegar a humanos reales al calcular los costos publicitarios en términos de dólares que se gastan por humano alcanzado, en lugar de dólares que se gastan por impresión servida a cualquier cosa, humana o robot. El CPH puede calcularse como sigue, por ejemplo. $CPH = \text{gasto total en publicidad} / \text{total de impresiones humanas que se obtienen con ese gasto}$, multiplicado por 1000 (mil) para escalar a la medida tradicional, CPM (costo por M, es decir, costo por mil). Si un anuncio se mostrara 1000 veces por \$10, el CPM de esas impresiones sería igual a \$10. Si, de esas 1000 impresiones, 600 se mostrarán a robots y solo 400 a humanos, el CPH equivaldría a \$25.

Evaluación de la Firma del Mapa de Calor: Cuando un ser humano está presente en una sesión de navegación, la invención contenida en la presente descripción puede usarse para evaluar los patrones de uso del ratón y el teclado, de manera que para cada usuario, pueda determinarse una firma de patrón, asumiendo que la configuración del navegador de esa persona permite que se recopile dicha información. Dichas firmas pueden usarse para varios fines, como dirigir contenido específico a usuarios humanos específicos.

Correlación de Firmas de Mapas de Calor: Con una cantidad suficiente de firmas de mapas de calor recopiladas, es posible comparar modelos de uso en un gran número de sitios web y, por lo tanto, detectar modelos de variación insuficientes o no humanos, con más datos de los que puede poseer un operador de agentes automatizados. Cabe señalar que, si bien la recopilación de firmas de mapas de calor con respecto a dónde hace clic un navegador determinado puede ser ampliamente conocida, el análisis muy detallado de los eventos reales del ratón es mucho

menos conocido en el campo de esta invención. Además, aunque algunos expertos pueden conocer la colección de patrones de movimiento inhumanos y eventos incompletos (como el ratón hacia abajo y hacia arriba, pero sin hacer clic en un dispositivo que no es móvil), la recopilación de tasas de eventos del ratón y eventos mal formados es novedosa en el campo.

5 Visibilidad Global: Con la implementación generalizada de las metodologías establecidas en la presente descripción, no solo en los sitios de destino, sino también en el JavaScript que aloja los clics en sí, es posible medir las tasas de acción de los agentes automatizados no solo en los sitios que han implementado activamente el fragmento de código necesario, sino para todos los sitios que pueden implementar dicho fragmento de código. Si se realiza
10 correctamente, esta metodología puede proporcionar una muestra estadísticamente significativa de todo el fraude de clics en Internet y, por lo tanto, proporcionar visibilidad global con respecto a la acción automatizada del navegador, y no solo visibilidad de los sitios que ejecutan nuestro fragmento de código.

15 Incorporación de la Página de Origen: Al incorporar las invenciones establecidas en la presente descripción en la página desde la que se origina un clic determinado (la "página de origen"), la interacción está garantizada independientemente de la naturaleza del visitante, ya que, por definición, un clic requiere interacción. La incrustación de la página de origen externa a un iframe, o marco incorporado, permite además el monitoreo de otras campañas publicitarias o contenido que se coloca en una página de origen determinada sin requerir la participación de las partes que colocan dicho contenido.

20 Ubicaciones Insertadas: La tecnología descrita en la presente descripción puede colocarse en la página de destino dentro de un iframe en la página desde la que se originó un clic a evaluarse, o fuera de un iframe en la página desde la que se originó un clic a evaluarse, que no solo aprovecha de los beneficios inherentes de cada tipo de ubicación, pero también permite monitorear el "ciclo de vida total del clic" o la secuencia de eventos que comienzan con la
25 presentación de un contenido específico como parte de la carga de una página web determinada, y continuar a través de la interacción de un usuario dado con ese contenido específico y hacer clic en él, a través de las páginas posteriores visitadas y los elementos de contenido con los que interactuó, y terminar con el abandono de la sesión de navegación o un evento de conversión.

30 Filtrado en Tiempo Real: Las invenciones establecidas en la presente descripción pueden usarse para proporcionar un determinado sitio web, anuncio, campaña publicitaria u otro usuario con filtrado en tiempo real, y para evitar de manera efectiva que agentes automatizados lleguen a sus destinos. Dicho filtrado en tiempo real puede ser tan rápido como 50 (cincuenta) milisegundos, aunque ciertas pruebas que se realizan por la presente invención ofrecen un resultado sólo después de que una página determinada está "completa". En el último caso, se usa una métrica de
35 "120 segundos desde la última vez que esa página envió algún dato al sistema". Además, la presente invención puede obligar a un código de cliente a dejar de enviar datos después de 120 segundos. Algunos robots no cumplen con el corte de 120 segundos y, por lo tanto, son fácilmente identificables.

40 Métricas del Proveedor de Servicios de Demanda: Los Proveedores de Servicios de Demanda de la industria de la publicidad generan ingresos al aprovechar las oportunidades de arbitraje con respecto a la colocación de anuncios en línea. Al usar la invención que se establece en la presente descripción para generar calidad en tiempo real, compromiso, el CPH u otras métricas relacionadas que se relacionan con dicha oportunidad, permitirá una evaluación más eficaz de dicha oportunidad.

45 Métricas de Compra de Anuncios en Tiempo Real: Específicamente, con respecto a lo anterior, es posible determinar en tiempo real si un anuncio determinado debe colocarse o mostrarse para una IP determinada, lo que hace posible no solo detectar, sino prevenir proactivamente clics fraudulentos o no deseados.

50 Validación del Navegador: Un agente de usuario de un navegador web (es decir, el tipo de navegador web que se usa actualmente) puede tergiversarse o "falsificarse" tanto por su fuente HTTP como por el contenido del DOM de JavaScript. Las invenciones que se establecen en la presente descripción pueden usarse para detectar dicha falsificación mediante el uso de métricas específicas de la versión del navegador.

55 Validación de Carga: Por motivos de eficiencia, es posible que los agentes automatizados no carguen algunos contenidos. Las invenciones descritas en la presente descripción pueden usarse para detectar tales cargas faltantes.

Detección de Proxy: Es posible alterar el comportamiento del servidor de evaluación con base en si existe un proxy. La forma en que se evalúan todas las demás métricas puede modificarse con base en el comportamiento de estos
60 nodos intermediarios.

La descripción de una modalidad preferida de la invención se ha presentado con fines ilustrativos y descriptivos. No pretende ser exhaustivo ni limitar la invención a las formas precisas descritas. Obviamente, muchas modificaciones y variaciones resultarán evidentes para los expertos en esta técnica. Se pretende que el alcance de la invención se defina por las siguientes reivindicaciones.

65 Además, las palabras "ejemplo" o "ilustrativo" se usan en la presente descripción con el significado de servir como

ejemplo, instancia o ilustración. Cualquier aspecto o diseño descrito en la presente descripción como "ilustrativo" no debe interpretarse necesariamente como preferido o ventajoso sobre otros aspectos o diseños. Más bien, el uso de las palabras "ejemplo" o "ilustrativo" pretende presentar conceptos de una manera concreta. Como se usa en esta solicitud, el término "o" pretende significar un "o" inclusivo en lugar de un "o" exclusivo. Es decir, a menos que se especifique lo contrario, o sea claro por el contexto, "X emplea A o B" pretende significar cualquiera de las permutaciones inclusivas naturales. Es decir, si X emplea A; X emplea B; o X emplea tanto A como B, entonces "X emplea A o B" se satisface en cualquiera de los casos anteriores. Además, los artículos "un" y "una", tal como se usan en esta solicitud, y las reivindicaciones adjuntas deben interpretarse generalmente como "uno o más" a menos que se especifique lo contrario o claramente del contexto para que se dirijan a una forma singular.

10

REIVINDICACIONES

1. Un método para detectar e informar sobre la actividad del agente de navegación automatizado, que comprende: emplear unos medios para detectar información del usuario para obtener una métrica, medir una diferencia con base en características de patrones para humanos y características de patrones para agentes navegadores automatizados, transmitir, a través de publicaciones HTTP asincrónicas, dicha información de usuario a un servidor, en donde dicho servidor registra un hallazgo con base en dicha información del usuario y dicha diferencia, y repetir dicha detección, medición y transmisión, compilando así un informe sobre la actividad del navegador humano contra el agente automatizado con base en una evaluación cualitativa de las métricas obtenidas, el método **caracterizado porque** la diferencia se basa en una comparación de dicha métrica a un primer grupo de control de características de patrones para humanos, una comparación de dicha métrica a un segundo grupo de control de características de patrones para agentes navegadores automatizados, y una comparación de dicha métrica a un tercer grupo de control de características que no están claras en cuanto a si se realizan por un humano o un agente navegador automatizado, en donde dichos medios para detectar además comprenden: insertar un fragmento de código en un código HTML de una página antes de enviar una página al navegador de un usuario y enviar dicha página al navegador de un usuario, en donde dicho fragmento de código provoca la recopilación de datos de la información del usuario una vez que el usuario ha cargado la página.
2. El método de la reivindicación 1, en donde dicha información de usuario comprende además: contenido que está presente que debería estar presente, contenido que está presente que debería estar ausente, contenido que está ausente que debería estar presente y contenido que está ausente que debería estar ausente.
3. El método de la reivindicación 1, en donde dicho fragmento de código se inyecta como una tecnología de secuencia de comandos activa, o en donde dicho fragmento de código se inyecta como JavaScript o como Flash.
4. El método de la reivindicación 1, en donde dicha información de usuario comprende además: una interacción con elementos invisibles de una página, propiedades faltantes de una interacción, una discrepancia entre los eventos del ratón, comportamiento atípico de la interfaz, una propiedad de elemento de página incorrecta, canales de comunicación que no coinciden, una tasa de actualización de Flash, sincronización de etapas de Flash, una tasa de actualización gráfica, elementos de JavaScript, información de manejo de errores, cumplimiento de estándares HTML5, configuraciones inyectadas específicas de robot, actividad del teclado, datos del acelerómetro, eventos de desplazamiento, tiempo promedio de lectura y visita, tasa de actualización de la página y protocolos de red y estándares web compatibles, o además comprende: información, generada a lo largo del tiempo, con relación a la cantidad de tiempo que tarda una operación de navegación determinada en expresar un resultado.
5. El método de la reivindicación 1, en donde dicho informe comprende además, simultáneamente, información con relación a al menos dos de: evaluación de ubicación, evaluación de la temporización entre clics, evaluación de la temporización entre clics del escritorio remoto y la VPN, detección de agentes automatizados móviles relacionados con el movimiento y el estado, detección de agentes automatizados móviles relacionados con el movimiento y el estado, detección de agentes automatizados móviles relacionados con el IP y la geolocalización, detección de agentes automatizados móviles relacionados con el IP y la geolocalización basada en el tiempo, ocultación y separación de datos, evaluación diferencial de la prestación, evaluación de inestabilidad, análisis de distribución de tiempo de la VM y validación de caché.
6. El método de la reivindicación 1, que comprende además: registrar un manejador y un oyente para un evento de navegación dado, en donde dicho manejador recibe información de usuario asociada con dicho evento de navegación y dicho oyente permite la recuperación de datos que de otro modo no serían identificables.
7. El método de la reivindicación 1, en donde dicha recopilación, comparación e informe de datos se implementan mediante procesamiento por lotes o se implementan mediante procesamiento de flujo.
8. El método de la reivindicación 1, en donde dicho informe se usa, simultáneamente, para al menos dos de: evaluación del compromiso, huella de robots, evaluación de errores de navegación, evaluación AB, evaluación de firmas estocásticas, evaluación en términos de costo por humano, evaluación de firmas de mapas de calor, correlación de firmas de mapas de calor, visibilidad global, incrustación de páginas de origen, ubicaciones de incrustaciones, filtrado en tiempo real, demandar métricas del proveedor de servicios, evaluación de métricas de compra de anuncios en tiempo real, validación del navegador, validación de carga, detección de proxy, tecnología financiera antifraude y un auditor de registro previo a CAPTCHA.
9. El método de la reivindicación 4, que comprende además una prueba repetida para la amplificación de pequeñas diferencias de temporización de agentes automatizados avanzados.

10. Un sistema informático para la detección de robots, que comprende:

5 una primera etapa de identificación diferencial, que comprende determinar la actividad de navegación con base en el origen y el tipo de usuario, siendo usuarios humanos contra un usuario automatizado,
 una segunda etapa de recopilación de métricas de rendimiento, que comprende el envío de una página que contiene un fragmento de código preinsertado para el registro de información particular del usuario, en la carga de la página y después de la carga de la página, transmitir posteriormente dicha métrica de rendimiento a un primer servidor,
 10 una tercera etapa de evaluación de dicha métrica de rendimiento dentro de dicho primer servidor, que comprende comparar dicha métrica de rendimiento con un primer grupo de control de características de patrones para humanos, un segundo grupo de control de características de patrones para agentes navegadores automatizados, y un tercer grupo de control de características que no están claras en cuanto a si se realizan por un humano o un agente de navegación automatizado, creando así una unidad de datos de usuario, que luego transmite, a través de una publicación HTTP asincrónica, dicha unidad de datos de usuario en un segundo servidor,
 15 y una cuarta etapa de informes dentro dicho segundo servidor, que comprende registrar un hallazgo con base en dicha unidad de datos de usuario,
 en donde dichas etapas se repiten, compilando así un informe sobre la actividad humana contra a la del robot con base en métricas de rendimiento recopiladas.

11. El sistema de la reivindicación 10, en donde dichas métricas de rendimiento comprenden además: contenido que está presente que debería estar presente, contenido que está presente que debería estar ausente, contenido que está ausente que debería estar presente, contenido que está ausente que debería estar ausente e información, generada a lo largo del tiempo, con relación a la cantidad de tiempo que una determinada operación de navegación tarda en expresar un resultado.

12. El sistema de la reivindicación 10, en donde dichas unidades de datos de usuario comprenden además: una interacción con elementos invisibles de una página, propiedades faltantes de una interacción, una discrepancia entre los eventos del ratón, comportamiento atípico de la interfaz, una propiedad de elemento de página incorrecta, canales de comunicación que no coinciden, una tasa de actualización de Flash, sincronización de etapas de Flash, una tasa de actualización gráfica, elementos de JavaScript, información de manejo de errores, cumplimiento de estándares HTML5, configuraciones inyectadas específicas de robot, actividad del teclado, datos del acelerómetro, eventos de desplazamiento, tiempo promedio de lectura y visita, tasa de actualización de la página y protocolos de red y estándares web compatibles.

13. El sistema de la reivindicación 10, en donde dicho informe sobre actividad humana contra a la del robot comprende además, simultáneamente, información con relación a al menos dos de: evaluación de ubicación, evaluación de la temporización entre clics, evaluación de la temporización entre clics del escritorio remoto y la VPN, detección de agentes automatizados móviles relacionados con el movimiento y el estado, detección de agentes automatizados móviles relacionados con el movimiento y el estado, detección de agentes automatizados móviles relacionados con el IP y la geolocalización, detección de agentes automatizados móviles relacionados con el IP y la geolocalización basada en el tiempo, ocultación y separación de datos, evaluación diferencial de la prestación, evaluación de inestabilidad, análisis de distribución de tiempo de la VM y validación de caché.

14. El sistema de la reivindicación 10, en donde dichas métricas de rendimiento, dicha evaluación o dichos informes se usan, simultáneamente, para al menos dos de: evaluación del compromiso, huella de robots, evaluación de errores de navegación, evaluación AB, evaluación de firmas estocásticas, evaluación en términos de costo por humano, evaluación de firmas de mapas de calor, correlación de firmas de mapas de calor, visibilidad global, incrustación de páginas de origen, ubicaciones de incrustaciones, filtrado en tiempo real, demandar métricas del proveedor de servicios, evaluación de métricas de compra de anuncios en tiempo real, validación del navegador, validación de carga y detección de proxy.

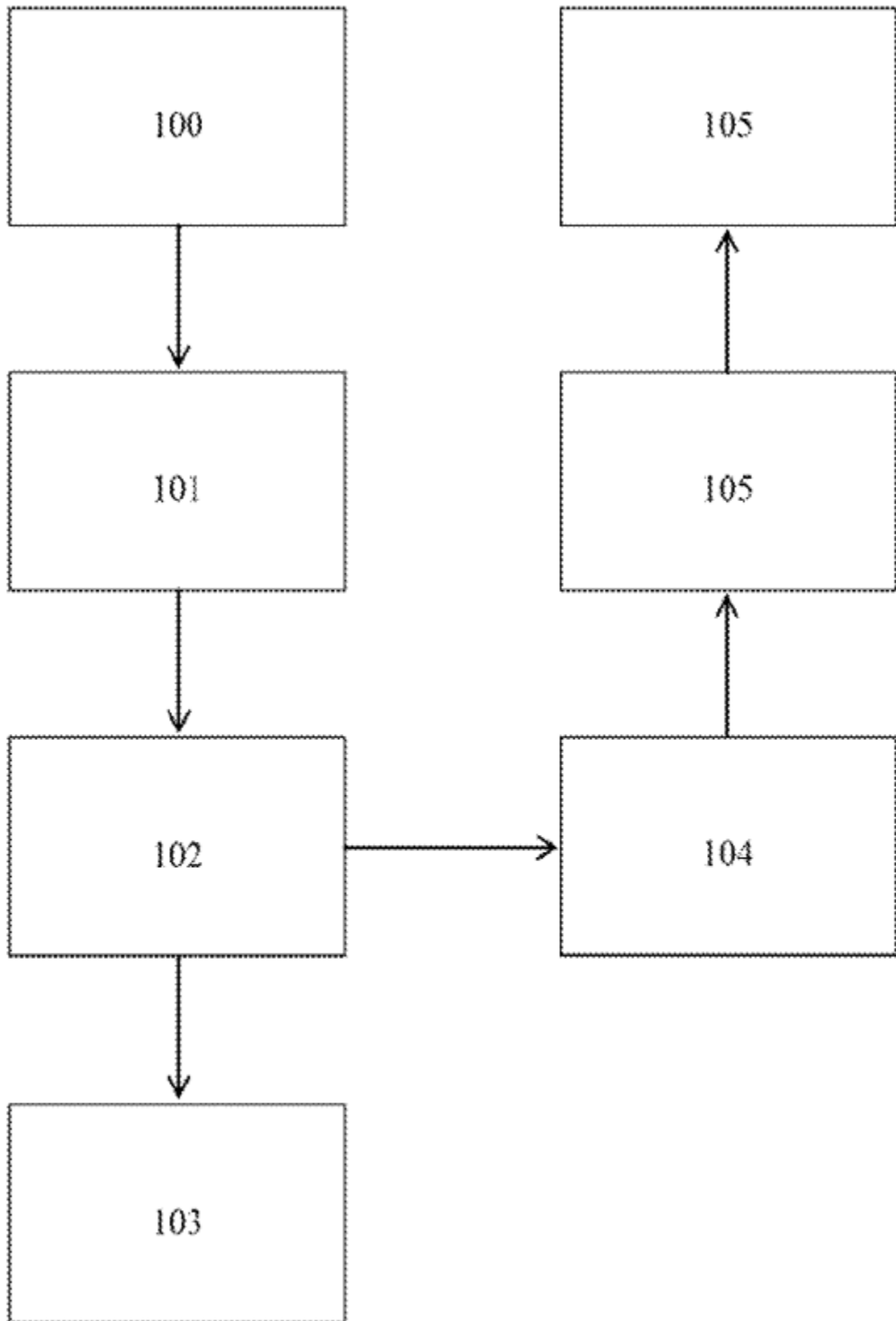


Figura 1

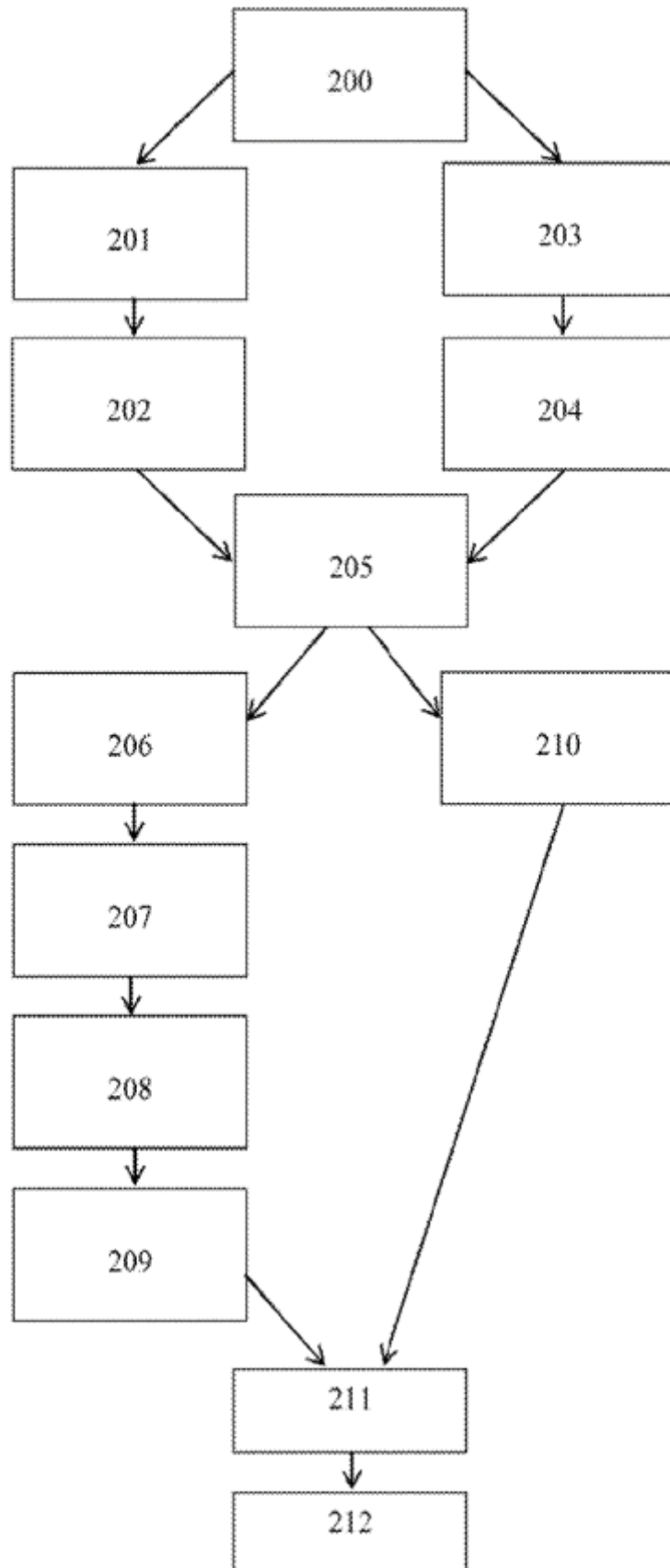


Figura 2

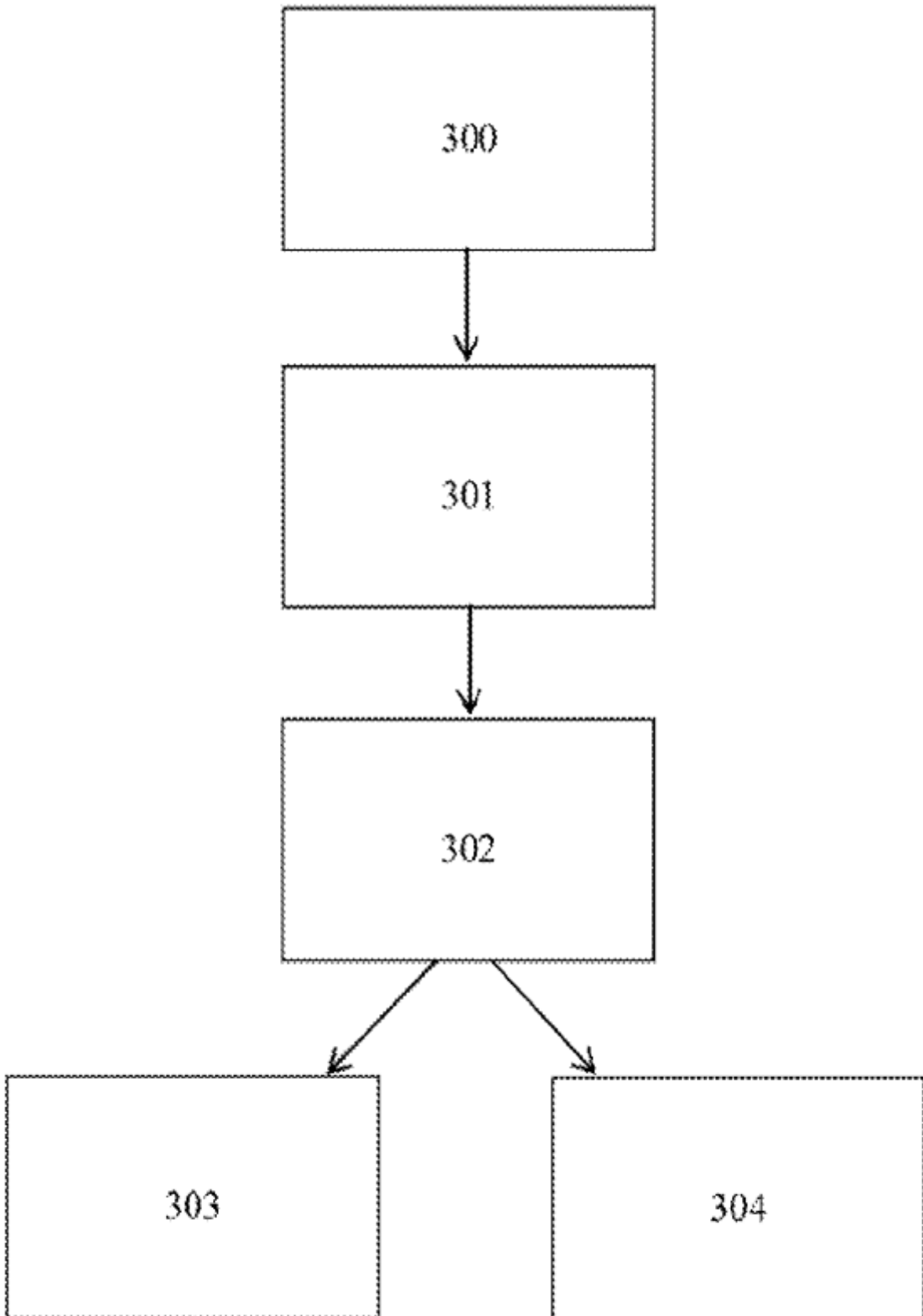


Figura 3