



(19)中華民國智慧財產局

(12)發明說明書公告本 (11)證書號數：TW I579708 B

(45)公告日：中華民國 106 (2017) 年 04 月 21 日

(21)申請案號：101146891

(22)申請日：中華民國 101 (2012) 年 12 月 12 日

(51)Int. Cl. : G06F15/16 (2006.01) G06F21/00 (2013.01)

(30)優先權：2012/09/21 中國大陸 201210356956.6

(71)申請人：阿里巴巴集團服務有限公司 (香港地區) ALIBABA GROUP SERVICES LIMITED
(HK)

香港

(72)發明人：趙雪波 (CN)

(74)代理人：林志剛

(56)參考文獻：

TW 200939048A

TW 200950445A

TW 201212675A

US 6141778

US 7761558B1

審查人員：洪唯真

申請專利範圍項數：8 項 圖式數：4 共 30 頁

(54)名稱

交互用戶資料的方法和裝置

(57)摘要

本申請公開了一種交互用戶資料的方法和裝置，用於提高網站訪問的安全性。該方法從各訪問用戶提交的訪問請求中選擇至少一個報頭域資訊和網路位址標識，組成多個訪問集合，並確定每個訪問集合對應的資訊損失率，根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並據此交互各訪問用戶的用戶資料。由於上述過程中，網站無需向第三方伺服器申請各訪問用戶對應的統一用戶標識，因此減少了用戶資訊洩密機會，增加了網站訪問的安全性。

指定代表圖：

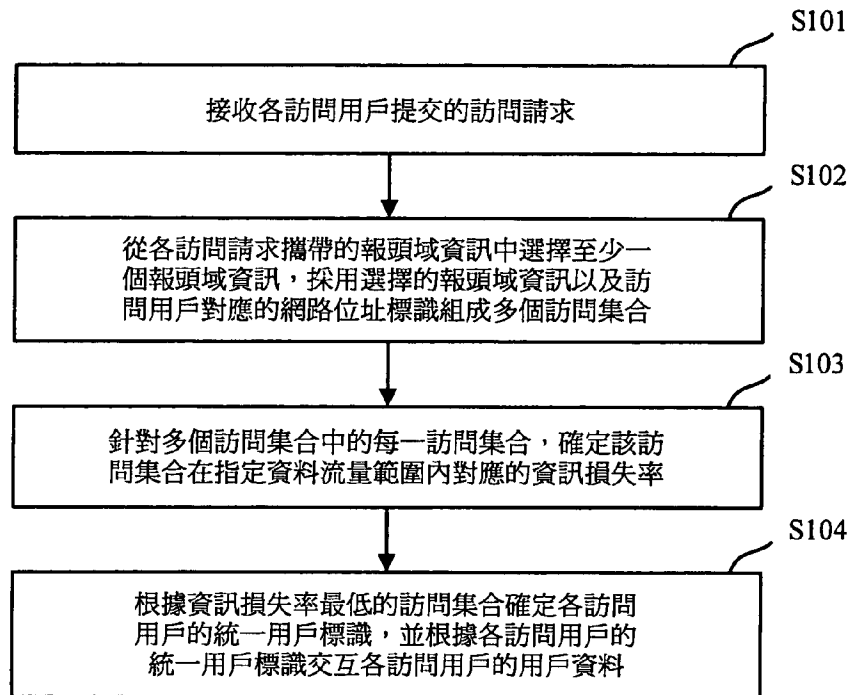


圖 1

發明專利說明書

(本申請書格式、順序，請勿任意更動，※記號部分請勿填寫)

※申請案號：101146891

※申請日：101 年 12 月 12 日

※IPC 分類：G06F 15/16 (2006.01)

G06F 21/00 (2013.01)

一、發明名稱：(中文／英文)

交互用戶資料的方法和裝置

二、中文發明摘要：

本申請公開了一種交互用戶資料的方法和裝置，用於提高網站訪問的安全性。該方法從各訪問用戶提交的訪問請求中選擇至少一個報頭域資訊和網路位址標識，組成多個訪問集合，並確定每個訪問集合對應的資訊損失率，根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並據此交互各訪問用戶的用戶資料。由於上述過程中，網站無需向第三方伺服器申請各訪問用戶對應的統一用戶標識，因此減少了用戶資訊洩密機會，增加了網站訪問的安全性。

三、英文發明摘要：

四、指定代表圖：

(一) 本案指定代表圖為：第(1)圖。

(二) 本代表圖之元件符號簡單說明：無

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：無

六、發明說明：

【發明所屬之技術領域】

本申請係關於網際網路技術領域，尤其關於一種交互用戶資料的方法和裝置。

【先前技術】

隨著網際網路技術的發展，各網際網路網站之間透過在一定程度上交互訪問用戶的用戶資料，從而能夠更全面地瞭解用戶的訪問資訊，使得各網際網路網站能夠根據用戶的訪問資訊，為用戶提供更加符合用戶需求的產品。

為了實現不同的網際網路網站之間交互同一用戶的用戶資料的目的，各網際網路網站需要確定一個唯一標識用於識別用戶，並在交互用戶資料時，根據該唯一標識交互該用戶的用戶資料。具體的，各網際網路網站在用戶訪問時，可向統一伺服器發送請求，以申請該用戶的唯一標識，統一伺服器根據網站發送的請求中攜帶的用戶客戶端特徵生成該用戶對應的唯一標識，並返回給各網站，各網站將該唯一標識寫入用戶客戶端 Cookie 中，各網際網路網站再根據該唯一標識交互該用戶的用戶資料。

上述過程中，由於各網站均需要向作為第三方的統一伺服器申請訪問用戶對應的唯一標識，並將申請到的唯一標識儲存於用戶客戶端 Cookie 中，這存在用戶資訊洩密風險，降低了網站訪問的安全性。

【發明內容】

本申請提供一種交互用戶資料的方法和裝置，用於提高網站訪問的安全性。

本申請實施例提供一種交互用戶資料的方法，包括：

接收各訪問用戶提交的訪問請求，訪問請求中攜帶訪問用戶的初始用戶標識、至少一個報頭域資訊和網路位址標識；

從各訪問請求攜帶的報頭域資訊中選擇至少一個報頭域資訊，採用選擇的報頭域資訊以及訪問用戶對應的網路位址標識組成多個訪問集合；

針對多個訪問集合中的每一訪問集合，確定該訪問集合在指定資料流量範圍內對應的資訊損失率；

根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並根據各訪問用戶的統一用戶標識交互各訪問用戶的用戶資料。

本申請實施例提供一種交互用戶資料的裝置，包括：

接收單元，用於接收各訪問用戶提交的訪問請求，訪問請求中攜帶訪問用戶的初始用戶標識、至少一個報頭域資訊和網路位址標識；

選擇單元，用於從各訪問請求攜帶的報頭域資訊中選擇至少一個報頭域資訊以及訪問用戶對應的網路位址標識組成多個訪問集合；

確定單元，用於針對每一訪問集合，確定該訪問集合在指定資料流量範圍內對應的資訊損失率；

資料交互單元，用於根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並根據各訪問用戶的統一用戶標識交互各訪問用戶的用戶資料。

本申請實施例提供一種交互用戶資料的方法及裝置，該方法從各訪問用戶提交的訪問請求中選擇至少一個報頭域資訊和網路位址標識，組成多個訪問集合，並確定每個訪問集合對應的資訊損失率，根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並據此交互各訪問用戶的用戶資料。由於上述過程中，網站無需向第三方伺服器申請各訪問用戶對應的統一用戶標識，因此減少了用戶資訊洩密機會，增加了網站訪問的安全性。

【實施方式】

由於現有技術中各網站之間交互用戶資料時，對於每一個訪問用戶，需要利用該訪問用戶的一些相關資訊向第三方伺服器申請該訪問用戶對應的統一用戶標識，因此，在此過程中存在一定的洩密風險，降低了網站訪問的安全性。有鑒於此，本申請實施例利用各訪問用戶提交的訪問請求中攜帶的報頭域資訊確定統一用戶標識，而無需向第三方伺服器申請統一用戶標識，以達到降低用戶資訊洩漏風險、提高網站訪問的安全性的目的。

以下結合說明書附圖對本申請的較佳實施例進行說明，應當理解，此處所描述的較佳實施例僅用於說明和解釋本申請，並不用於限定本申請，並且在不衝突的情況

下，本申請中的實施例及實施例中的特徵可以相互組合。

圖 1 為本申請實施例提供的交互用戶資料的過程，包括以下步驟：

S101：接收各訪問用戶提交的訪問請求。

其中，一個訪問用戶提交的訪問請求中攜帶該訪問用戶的初始用戶標識、至少一個報頭域資訊和網路位址標識。網路位址標識可以為訪問用戶對應的網際網路協定（Internet Protocol，IP）位址；用戶的初始用戶標識可以為用戶 cookie 標識；當訪問請求為超文件傳輸協定（hypertext transport protocol，http）訪問請求時，該訪問請求攜帶的報頭域資訊包括以下資訊中的至少一種：Accept-Language（用戶接受的語言）資訊、Accept-Encoding（用戶接受的編碼方式）資訊、User-Agent（用戶代理）資訊以及用戶客戶端資訊，用戶客戶端資訊包括：UA-Pixels（用戶客戶端圖元）資訊、UA-Color（用戶客戶端色彩）資訊、UA-OS（用戶客戶端作業系統）資訊、UA-CPU（用戶客戶端中央處理單元）資訊等。另外，訪問請求也可以是其他協議的訪問請求，本申請中不具體限定。

S102：從各訪問請求攜帶的報頭域資訊中選擇至少一個報頭域資訊，採用選擇的報頭域資訊以及訪問用戶對應的網路位址標識組成多個訪問集合。

為了便於描述，分別以 A，B，C，D 表示 Accept-Language，Accept-Encoding，User-Agent 和用戶客戶端資

訊，例如，對於網路位址相同的訪問用戶來說，假設訪問用戶對應的網路位址標識均為 IP1，則根據該四種報頭域資訊之間的排列組合，可以組成 $15 (C_4^1 + C_4^2 + C_4^3 + C_4^4)$ 種訪問集合，分別為：[A，IP1]，[B，IP1]，[C，IP1]，[D，IP1]，[A，B，IP1]，[A，C，IP1]，[A，D，IP1]，[B，C，IP1]，[B，D，IP1]，[C，D，IP1]，[A，B，C，IP1]，[A，B，D，IP1]，[A，C，D，IP1]，[B，C，D，IP1]，[A，B，C，D，IP1]。

S103：針對多個訪問集合中的每一訪問集合，確定該訪問集合在指定資料流量範圍內對應的資訊損失率。

在本申請實施例中，首先可設置該指定資料流量範圍，具體可以根據實際需要設置該資料流量範圍。例如，該指定資料流量範圍可以設置為指定時間段內（如 1 天內）的資料流量。又如，對於電子商務網站來說，其資料流量來源主要包括：登錄頁面與用戶交互的資料流量、商品瀏覽頁面與用戶交互的資料流量、商品交易頁面與用戶交互的資料流量，因此，也可以將該指定資料流量範圍設置為某一種或幾種類型的頁面與用戶交互的資料流量。

在設置了該指定資料流量範圍之後，則可以分別確定透過上述步驟 S102 組成的每個訪問集合在該指定資料流量範圍內的資訊損失率。

對於一個訪問集合來說，該訪問集合在該指定資料流量範圍內的資訊損失率可以表徵：採用該訪問集合確定各訪問用戶的統一用戶標識的準確性，或者說可信度。亦

即，該訪問集合在該指定資料流量範圍內的資訊損失率越低，採用該訪問集合確定各訪問用戶的統一用戶標識的準確性越高，或者說可信度越大，反之，該訪問集合在該指定資料流量範圍內的資訊損失率越高，採用該訪問集合確定各訪問用戶的統一用戶標識的準確性越低，或者說可信度越小。

S104：根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並根據各訪問用戶的統一用戶標識交互各訪問用戶的用戶資料。

由於一個訪問集合在指定資料流量範圍內的資訊損失率越低，採用該訪問集合確定各訪問用戶的統一用戶標識的準確性越高，因此，在透過步驟 S102 組成的各訪問集合中，選擇出在指定資料流量範圍內的資訊損失率最低的訪問集合，並根據選擇出的該訪問集合確定各訪問用戶的統一用戶標識，可以儘量的提高確定的統一用戶標識的準確性。後續則可以根據確定的各訪問用戶的統一用戶標識交互各訪問用戶的用戶資料。

具體的，在上述圖 1 所示的步驟 S103 中，針對每一訪問集合，確定該訪問集合在指定資料流量範圍內對應的資訊損失率的方法如圖 2 所示。

圖 2 為本申請實施例提供的確定一個訪問集合在指定流量範圍內對應的資訊損失率的過程，具體包括以下步驟：

S1031：根據該訪問集合包含的報頭域資訊的內容，

將該訪問集合重新組合為多個訪問子集合。

其中，每一訪問子集合的元素為報頭域資訊的內容和網路位址標識。

例如，假設有 3 個用戶提交了訪問請求，這 3 個用戶的初始用戶標識分別為 B_1 、 B_2 、 B_3 ，這 3 個訪問用戶的網路位址標識均為 IP1，指定資料流量範圍內的資料流量的數量為 50 個，現在要確定包含的元素為 Accept-Language 資訊、Accept-Coding 資訊以及 IP1 的訪問集合 [Accept-Language 資訊，Accept-Coding 資訊，IP1] 在指定流量範圍內的資訊損失率。

如果這 3 個用戶提交的訪問請求中，Accept-Language 資訊的具體內容均為中文，而 Accept-Coding 資訊的具體內容為編碼方式 1 和編碼方式 2 共兩種，那麼，對於網路位址標識均為 IP1 的這 3 個訪問用戶，針對訪問集合 [Accept-Language 資訊，Accept-Coding 資訊，IP1] 重組的訪問子集合即為：[中文，編碼方式 1，IP1] 和 [中文，編碼方式 2，IP1]。

為了便於描述，下面以 A_1 表示中文和英文，以 C_1 和 C_2 分別表示編碼方式 1 和編碼方式 2 進行說明。

亦即重組的訪問子集合為：[A_1 ， C_1 ，IP1] 和 [A_1 ， C_2 ，IP1]。

S1032：在該指定資料流量範圍內，統計每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量。

由於訪問請求中攜帶的報頭域資訊全部來自於標準協

定，因此，對於不同訪問用戶來說，其提交的訪問請求中攜帶的請求報頭域資訊的內容可能相同，也可能不同，亦即，對於一個訪問子集合來說，其對應的訪問用戶可能不只一個。另外，對於同一訪問用戶來說，其提交的不同的訪問請求中所攜帶的報頭域資訊可能相同，也可能不同。

繼續沿用上例，假設：

在初始用戶標識 B_1 對應的訪問用戶提交的訪問請求中，存在採用中文、編碼方式 1、IP1 提交的訪問請求，也存在採用中文、編碼方式 2、IP1 提交的訪問請求；

在初始用戶標識 B_2 對應的訪問用戶提交的訪問請求中，存在採用中文、編碼方式 1、IP1 提交的訪問請求，也存在採用中文、編碼方式 2、IP1 提交的訪問請求；

在初始用戶標識 B_3 對應的訪問用戶提交的訪問請求中，只存在採用中文、編碼方式 1、IP1 提交的訪問請求。

則：訪問子集合 $[A_1, C_1, IP1]$ 和 $[A_1, C_2, IP1]$ ，與用戶初始標識 B_1 、 B_2 、 B_3 的對應關係如圖 3 所示。圖 3 為本申請實施例提供的訪問子集合與初始用戶標識之間的對應關係示意圖，在圖 3 中，訪問子集合 $[A_1, C_1, IP1]$ 與初始用戶標識 B_1 、 B_2 、 B_3 對應，訪問子集合 $[A_1, C_2, IP1]$ 與初始用戶標識 B_1 、 B_2 對應。

因此，在指定資料流量範圍內，亦即在 50 個資料流量中，統計 $[A_1, C_1, IP1]$ 在對應的 B_1 、 B_2 、 B_3 上承載的資料流量，統計 $[A_1, C_2, IP1]$ 在對應的 B_1 、 B_2 上承載

的資料流量。

具體的，統計一個訪問子集合在其對應的所有初始用戶標識上承載的資料流量的方法為：在指定資料流量範圍內，統計該訪問子集合承載的資料流量、該訪問子集合對應的初始用戶標識數量，以及該訪問子集合對應的每一初始用戶標識對應的訪問子集合的數量；針對該訪問子集合對應的每一初始用戶標識，確定該訪問子集合在該初始用戶標識承載的資料流量；將該訪問子集合在其對應的每一初始用戶標識上承載的資料流量之和，確定為該訪問子集合在其對應的所有初始用戶標識上承載的資料流量。

其中，可以採用公式 $\frac{N}{i*j}$ 確定一個訪問子集合在其對應的一個初始用戶標識承載的資料流量， N 為該訪問子集合承載的資料流量， i 為該訪問子集合對應的初始用戶標識數量， j 為該初始用戶標識對應的訪問子集合的數量。

繼續沿用上例，假設在這指定的 50 個資料流量中，有 30 個承載在 $[A_1, C_1, IP1]$ 上，有 20 個承載在 $[A_1, C_2, IP1]$ 上，則：

針對 $[A_1, C_1, IP1]$ ， $N=30$ ，其對應的初始用戶標識分別為 B_1 、 B_2 、 B_3 ， $i=3$ ，初始用戶標識 B_1 對應的訪問子集合分別為 $[A_1, C_1, IP1]$ 和 $[A_1, C_2, IP1]$ ， $j=2$ ，因此， $[A_1, C_1, IP1]$ 承載在 B_1 上的資料流量為 $((30/3)/2) = 5$ 個；

類似的， $[A_1, C_1, IP1]$ 承載在 B_2 上的資料流量為 $((30/3)/2) = 5$ 個， $[A_1, C_1, IP1]$ 承載在 B_3 上的資料流

量為 $(30/3) = 10$ 個，從而，訪問子集合 $[A_1, C_1, IP1]$ 在其對應的所有初始用戶標識上承載的資料流量為 $(5+5+10) = 20$ 個。

相應的， $[A_1, C_2, IP1]$ 承載在 B_1 上的資料流量為 $((20/2)/2) = 5$ 個， $[A_1, C_2, IP1]$ 承載在 B_2 上的資料流量為 $((20/2)/2) = 5$ 個，因此訪問子集合 $[A_1, C_2, IP1]$ 在其對應的所有初始用戶標識上承載的資料流量為 $(5+5) = 10$ 個。

S1033：根據統計的每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量，確定該訪問集合在指定資料流量範圍內對應的資訊損失率。

具體的，可以採用公式 $1 - \frac{R}{W}$ 確定該訪問集合在指定資料流量範圍內對應的資訊損失率， R 表示該訪問集合包含的每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量之和， W 表示指定資料流量範圍內包含的所有資料流量之和。

繼續沿用上例，由於 $[A_1, C_1, IP1]$ 在其對應的所有初始用戶標識上承載的資料流量為 20 個，訪問子集合 $[A_1, C_2, IP1]$ 在其對應的所有初始用戶標識上承載的資料流量為 10 個，因此 $R=20+10=30$ ， $W=50$ ，從而，訪問集合 $[Accept-Language \text{ 資訊}, Accept-Coding \text{ 資訊}, IP1]$ 在這 50 個資料流量內對應的資訊損失率為： $1-30/50=0.4$ 。

以上是本申請實施例提供的確定一個訪問集合在指定資料流量範圍內的資訊損失率的方法。在後續過程中，則

可以選擇出資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並根據各訪問用戶的統一用戶標識交互各訪問用戶的用戶資料。

其中，根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識的方法具體為：分別以資訊損失率最低的訪問集合中包含的不同訪問子集合作為不同訪問用戶的統一用戶標識。

繼續沿用上例，如果訪問集合[Accept-Language 資訊，Accept-Coding 資訊，IP1]的資訊損失率最低，則以該訪問集合[Accept-Language 資訊，Accept-Coding 資訊，IP1]中包含的訪問子集合 $[A_1, C_1, IP1]$ 作為一個訪問用戶的同一用戶標識，以訪問子集合 $[A_1, C_2, IP1]$ 作為另一個訪問用戶的同一用戶標識。

需要說明的是，在上例中，初始用戶標識有 3 個，亦即有 3 個訪問用戶，而最終確定出的統一用戶標識只有 2 個，用這 2 個統一用戶標識只能標識出 2 個訪問用戶，所以，採用一個訪問集合中包含的每個訪問子集合標識訪問用戶時，其標識的準確性會下降。而其準確性下降的程度，就是透過上述計算的資訊損失率來表徵的，因此，本申請實施例中要根據資訊損失率最小的訪問集合來確定統一用戶標識。

進一步的，理想的情況是：找到一種訪問集合，這種訪問集合中包含的每個訪問子集合與每個初始用戶標識是完全一一對應的關係。這樣，用該訪問集合中包含的不同

訪問子集合來標識不同訪問用戶的準確率是 100%。但是，實際應用中幾乎找不到這樣的訪問集合，亦即，對於一個訪問集合來說，如果其包含的每個訪問子集合與每個初始用戶標識並不是一一對應的關係，則用該訪問集合中的訪問子集合標識訪問用戶時，就會造成準確率的下降。造成準確率下降的情況有以下兩種：

情況一、不同的訪問用戶採用同一個訪問子集合承載資料流量時，該同一個訪問子集合並未標識出不同的用戶。例如，上例中的初始用戶標識為 B_1 、 B_2 、 B_3 的 3 個訪問用戶，這 3 個訪問用戶都採用訪問子集合 $[A_1, C_1, IP1]$ 承載了資料流量，但是訪問子集合 $[A_1, C_1, IP1]$ 只能標識一個訪問用戶，亦即，同樣採用 $[A_1, C_1, IP1]$ 承載資料流量的 3 個訪問用戶並未被區分出來。

情況二、同一個訪問用戶採用不同的訪問子集合承載資料流量時，該同一個用戶承載的資料流量會被標識為不同的用戶承載的資料流量。例如，上例中的初始用戶標識為 B_1 的訪問用戶，該訪問用戶採用訪問子集合 $[A_1, C_1, IP1]$ 和 $[A_1, C_2, IP1]$ 承載了資料流量，但是， $[A_1, C_1, IP1]$ 和 $[A_1, C_2, IP1]$ 分別標識了不同的用戶，亦即，同樣是初始用戶標識為 B_1 的訪問用戶承載的資料流量卻被標識成了兩個不同的用戶承載的資料流量。

因此，上述計算的一個訪問集合在指定資料流量訪問內的資訊損失量，就是在計算採用這個訪問集合中的每個訪問子集合標識每個訪問用戶時，由於上述情況一和情況

二所帶來的準確率的下降。

從而，針對上例中的計算方法也可以這樣理解：

如圖 3 所示，針對訪問子集合 $[A_1, C_1, IP1]$ ，由於 3 個訪問用戶都採用了 $[A_1, C_1, IP1]$ 承載資料流量，因此基於情況一，無論採用 $[A_1, C_1, IP1]$ 標識哪個訪問用戶，其準確率也只有 $\frac{1}{3}$ 。進一步的，在基於情況一的基礎上，當採用 $[A_1, C_1, IP1]$ 標識 B_1 對應的訪問用戶時，基於情況二， B_1 對應的訪問用戶採用 $[A_1, C_2, IP1]$ 承載的資料流量也會被標識為其他用戶承載的資料流量，因此採用 $[A_1, C_1, IP1]$ 標識 B_1 對應的訪問用戶的準確率為 $\frac{1}{3} \times \frac{1}{2}$ 。同樣的，採用 $[A_1, C_1, IP1]$ 標識 B_2 對應的訪問用戶的準確率為 $\frac{1}{3} \times \frac{1}{2}$ ，採用 $[A_1, C_1, IP1]$ 標識 B_3 對應的訪問用戶的準確率為 $\frac{1}{3} \times 1$ 。因此，採用訪問子集合 $[A_1, C_1, IP1]$ 標識一個訪問用戶的準確率為 $\frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 1$ 。

類似的，針對訪問子集合 $[A_1, C_2, IP1]$ ，由於兩個訪問用戶（ B_1 對應的訪問用戶和 B_2 對應的訪問用戶）採用了 $[A_1, C_2, IP1]$ 承載資料流量，因此基於情況一，無論採用 $[A_1, C_2, IP1]$ 標識這兩個中的哪個訪問用戶，其準確率也只有 $\frac{1}{2}$ 。進一步的，在基於情況一的基礎上，當採用 $[A_1, C_2, IP1]$ 標識 B_1 對應的訪問用戶時，基於情況二， B_1 對應的訪問用戶採用 $[A_1, C_1, IP1]$ 承載的資料流量也會被標識為其他用戶承載的資料流量，因此採用 $[A_1, C_2, IP1]$ 標識 B_1 對應的訪問用戶的準確率為 $\frac{1}{2} \times \frac{1}{2}$ 。同樣的，採用 $[A_1, C_2, IP1]$ 標識 B_2 對應的訪問用戶的準確率

為 $\frac{1}{2} \times \frac{1}{2}$ 。因此，採用訪問子集合 $[A_1, C_2, IP1]$ 標識一個訪問用戶的準確率為 $\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2}$ 。

再考慮到 $[A_1, C_1, IP1]$ 承載的資料流量在總流量中所占的比例 $\frac{30}{50}$ 和 $[A_1, C_2, IP1]$ 承載的資料流量在總流量中所占的比例 $\frac{20}{50}$ ，則在這 50 個資料流量中，採用訪問集合 $[Accept-Language \text{ 信息}, Accept-Coding \text{ 信息}, IP1]$ 中包含的兩個訪問子集合分別標識不同的訪問用戶的準確率即為： $\frac{30}{50} \left(\frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 1 \right) + \frac{20}{50} \left(\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} \right) = 0.6$ 。因此，其準確率下降了 $1 - 0.6 = 0.4$ ，亦即，資訊損失率是 0.4。

從而，本申請實施例實際上是尋找一種包含的每個訪問子集合盡可能與每個初始用戶標識接近於一一對應的訪問集合，亦即，採用資訊損失率最低的訪問集合確定訪問用戶的統一用戶標識。

較佳地，為了保證採用訪問集合對訪問用戶進行標識的可用性，本申請實施例中，可以預先設置資訊損失率閾值，當資訊損失率最低的訪問集合對應的資訊損失率不超過預設閾值時，則資訊損失率最低的訪問集合中包含的訪問子集合可以用於標識訪問用戶，當資訊損失率最低的訪問集合對應的資訊損失率超過預設閾值時，說明採用該訪問集合中包含的訪問子集合標識訪問用戶的準確率過低，因此該訪問集合中包含的訪問子集合無法用於標識訪問用戶。

在本申請實施例中，由於同一用戶在不同網站有不同的用戶資料，例如在某搜尋網站有對關鍵字搜尋的資料，

在入口網站有對各頻道頁面的訪問資料，在電子商務網站有商品瀏覽、收藏以及交易等行為資料，因此在每個網站可以分別保存同一用戶的不同資料。在確定出統一用戶標識後，就可以根據該用戶的統一用戶標識對該用戶在不同網站的不同資料進行拼接，使得用戶資料更加完整，後續各網站根據自身需要可以用於最佳化搜尋、個性化頁面推送、廣告推送最佳化等應用中。

基於同一申請構思，本申請實施例中還提供了一種交互用戶資料的裝置，由於該裝置解決問題的原理與網站訪問用戶標識確定方法相似，因此該裝置的實施可以參見方法的實施，重複之處不再贅述。

圖 4 為本申請實施例提供的交互用戶資料的裝置結構示意圖，具體包括：

接收單元 401，用於接收各訪問用戶提交的訪問請求，訪問請求中攜帶訪問用戶的初始用戶標識、至少一個報頭域資訊和網路位址標識；

選擇單元 402，用於從各訪問請求攜帶的報頭域資訊中選擇至少一個報頭域資訊以及訪問用戶對應的網路位址標識組成多個訪問集合；

確定單元 403，用於針對每一訪問集合，確定該訪問集合在指定資料流量範圍內對應的資訊損失率；

資料交互單元 404，用於根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並根據各訪問用戶的統一用戶標識交互各訪問用戶的用戶資料。

該確定單元 403，包括：

組合子單元 4031，用於針對每一訪問集合，根據該訪問集合包含的報頭域資訊的內容，將該訪問集合重新組合為多個訪問子集合，每一訪問子集合的元素為報頭域資訊的內容和網路位址標識；

統計子單元 4032，用於在該指定資料流量範圍內，統計每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量；

確定子單元 4033，用於根據統計的每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量，確定該訪問集合在該指定資料流量範圍內對應的資訊損失率。

該確定子單元 4033 具體用於，按照以下公式確定該訪問集合在該指定資料流量範圍內對應的資訊損失率：

$1 - \frac{R}{W}$ ，其中： R 表示該訪問集合包含的每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量之和； W 表示指定資料流量範圍內包含的所有資料流量之和。

該統計子單元 4032，包括：

統計模組 40321，用於針對每一訪問子集合，在該指定資料流量範圍內，統計該訪問子集合承載的資料流量、該訪問子集合對應的初始用戶標識數量，以及該訪問子集合對應的每一初始用戶標識對應的訪問子集合的數量；

第一確定模組 40322，用於針對該訪問子集合對應的每一初始用戶標識，確定該訪問子集合在該初始用戶標識承載的資料流量；

第二確定模組 40323，用於將該訪問子集合在其對應的每一初始用戶標識上承載的資料流量之和，確定為該訪問子集合在對應的所有初始用戶標識上承載的資料流量。

該第一確定模組 40322 具體用於，按照以下公式確定該訪問子集合在該初始用戶標識承載的資料流量： $\frac{N}{i*j}$ ，其中： N 為該訪問子集合承載的資料流量； i 為該訪問子集合對應的初始用戶標識數量； j 為該初始用戶標識對應的訪問子集合的數量。

該資料交互單元 404 還用於，在根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識之前，確定資訊損失率最低的訪問集合對應的資訊損失率不超過預設閾值。

該資料交互單元 404 具體用於，分別以資訊損失率最低的訪問集合中包含的不同訪問子集合作為不同訪問用戶的統一用戶標識。

具體的上述交互用戶資料的裝置可以位於網站的伺服器中。

本申請實施例提供一種交互用戶資料的方法及裝置，該方法從各訪問用戶提交的訪問請求中選擇至少一個報頭域資訊和網路位址標識，組成多個訪問集合，並確定每個訪問集合對應的資訊損失率，根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並據此交互各訪問用戶的用戶資料。由於上述過程中，網站無需向第三方伺服器申請各訪問用戶對應的統一用戶標識，因此減少了用

戶資訊洩密機會，增加了網站訪問的安全性。

本領域內的技術人員應明白，本申請的實施例可提供為方法、系統、或電腦程式產品。因此，本申請可採用完全硬體實施例、完全軟體實施例、或結合軟體和硬體方面的實施例的形式。而且，本申請可採用在一個或多個其中包含有電腦可用程式碼的電腦可用儲存媒體（包括但不限於磁碟記憶體、CD-ROM、光學記憶體等）上實施的電腦程式產品的形式。

本申請是參照根據本申請實施例的方法、設備（系統）、和電腦程式產品的流程圖和／或方框圖來描述的。應理解可由電腦程式指令實現流程圖和／或方框圖中的每一流程和／或方框、以及流程圖和／或方框圖中的流程和／或方框的結合。可提供這些電腦程式指令到通用電腦、專用電腦、嵌入式處理機或其他可編程資料處理設備的處理器以產生一個機器，使得透過電腦或其他可編程資料處理設備的處理器執行的指令產生用於實現在流程圖一個流程或多個流程和／或方框圖一個方框或多個方框中指定的功能的裝置。

這些電腦程式指令也可儲存在能引導電腦或其他可編程資料處理設備以特定方式工作的電腦可讀記憶體中，使得儲存在該電腦可讀記憶體中的指令產生包括指令裝置的製造品，該指令裝置實現在流程圖一個流程或多個流程和／或方框圖一個方框或多個方框中指定的功能。

這些電腦程式指令也可裝載到電腦或其他可編程資料

處理設備上，使得在電腦或其他可編程設備上執行一系列操作步驟以產生電腦實現的處理，從而在電腦或其他可編程設備上執行的指令提供用於實現在流程圖一個流程或多個流程和／或方框圖一個方框或多個方框中指定的功能的步驟。

儘管已描述了本申請的較佳實施例，但本領域內的技術人員一旦得知了基本創造性概念，則可對這些實施例做出另外的變更和修改。所以，所附申請專利範圍意欲解釋為包括較佳實施例以及落入本申請範圍的所有變更和修改。

顯然，本領域的技術人員可以對本申請進行各種改動和變型而不脫離本申請的精神和範圍。這樣，倘若本申請的這些修改和變型屬於本申請之申請專利範圍及其等同技術的範圍之內，則本申請也意圖包含這些改動和變型在內。

【圖式簡單說明】

圖 1 為本申請實施例提供的交互用戶資料的過程；

圖 2 為本申請實施例提供的確定一個訪問集合在指定流量範圍內對應的資訊損失率的過程；

圖 3 為本申請實施例提供的訪問子集合與初始用戶標識之間的對應關係示意圖；

圖 4 為本申請實施例提供的交互用戶資料的裝置結構示意圖。

【 主 要 元 件 符 號 說 明 】

401：接收單元

402：選擇單元

403：確定單元

404：資料交互單元

4031：組合子單元

4032：統計子單元

4033：確定子單元

40321：統計模組

40322：第一確定模組

40323：第二確定模組

七、申請專利範圍：

1. 一種交互用戶資料的方法，其特徵在於，包括：

接收各訪問用戶提交的訪問請求，訪問請求中攜帶訪問用戶的初始用戶標識、至少一個報頭域資訊和網路位址標識；

從各訪問請求攜帶的報頭域資訊中選擇至少一個報頭域資訊，採用選擇的報頭域資訊以及訪問用戶對應的網路位址標識組成多個訪問集合；

針對多個訪問集合中的每一訪問集合，確定該訪問集合在指定資料流量範圍內對應的資訊損失率；

根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並根據各訪問用戶的統一用戶標識交互各訪問用戶的用戶資料，

其中，針對多個訪問集合中的每一訪問集合，確定該訪問集合在指定資料流量範圍內對應的資訊損失率，具體包括：

針對每一訪問集合，根據該訪問集合包含的報頭域資訊的內容，將該訪問集合重新組合為多個訪問子集合，每一訪問子集合的元素為報頭域資訊的內容和網路位址標識；

在該指定資料流量範圍內，統計每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量；

根據統計的每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量，確定該訪問集合在該指定資料

流量範圍內對應的資訊損失率。

2.如申請專利範圍第 1 項所述的方法，其中，該確定該訪問集合在指定資料流量範圍內對應的資訊損失率，根據以下公式確定：

$$1 - \frac{R}{W},$$

其中，

R 表示該訪問集合包含的每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量之和；

W 表示指定資料流量範圍內包含的所有資料流量之和。

3.如申請專利範圍第 2 項所述的方法，其中，統計一個訪問子集合在其對應的所有初始用戶標識上承載的資料流量，具體包括：

在該指定資料流量範圍內，統計該訪問子集合承載的資料流量、該訪問子集合對應的初始用戶標識數量，以及該訪問子集合對應的每一初始用戶標識對應的訪問子集合的數量；

針對該訪問子集合對應的每一初始用戶標識，確定該訪問子集合在該初始用戶標識承載的資料流量；

將該訪問子集合在其對應的每一初始用戶標識上承載的資料流量之和，確定為該訪問子集合在其對應的所有初始用戶標識上承載的資料流量。

4.如申請專利範圍第 3 項所述的方法，其中，按照以下公式確定該訪問子集合在該初始用戶標識承載的資料流

量：

$$\frac{N}{i*j},$$

其中：

N 為該訪問子集合承載的資料流量；

i 為該訪問子集合對應的初始用戶標識數量；

j 為該初始用戶標識對應的訪問子集合的數量。

5. 如申請專利範圍第 1 項所述的方法，其中，根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識之前，還包括：

確定資訊損失率最低的訪問集合對應的資訊損失率不超過預設閾值。

6. 如申請專利範圍第 1 項所述的方法，其中，當訪問請求為超文件傳輸協定 http 訪問請求時，該訪問請求攜帶的報頭域資訊包括以下資訊中的至少一種：用戶接受的語言 Accept-Language 資訊、用戶接受的編碼方式 Accept-Encoding 資訊、用戶代理 User-Agent 資訊以及用戶客戶端資訊。

7. 如申請專利範圍第 1 項所述的方法，其中，根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，具體包括：

分別以資訊損失率最低的訪問集合中包含的不同訪問子集合作為不同訪問用戶的統一用戶標識。

8. 一種交互用戶資料的裝置，其特徵在於，包括：

接收單元，用於接收各訪問用戶提交的訪問請求，訪

問請求中攜帶訪問用戶的初始用戶標識、至少一個報頭域資訊和網路位址標識；

選擇單元，用於從各訪問請求攜帶的報頭域資訊中選擇至少一個報頭域資訊以及訪問用戶對應的網路位址標識組成多個訪問集合；

確定單元，用於針對每一訪問集合，確定該訪問集合在指定資料流量範圍內對應的資訊損失率；

資料交互單元，用於根據資訊損失率最低的訪問集合確定各訪問用戶的統一用戶標識，並根據各訪問用戶的統一用戶標識交互各訪問用戶的用戶資料，

其中，該確定單元，包括：

組合子單元，用於針對每一訪問集合，根據該訪問集合包含的報頭域資訊的內容，將該訪問集合重新組合為多個訪問子集合，每一訪問子集合的元素為報頭域資訊的內容和網路位址標識；

統計子單元，用於在該指定資料流量範圍內，統計每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量；

確定子單元，用於根據統計的每一訪問子集合在各自對應的所有初始用戶標識上承載的資料流量，確定該訪問集合在該指定資料流量範圍內對應的資訊損失率。

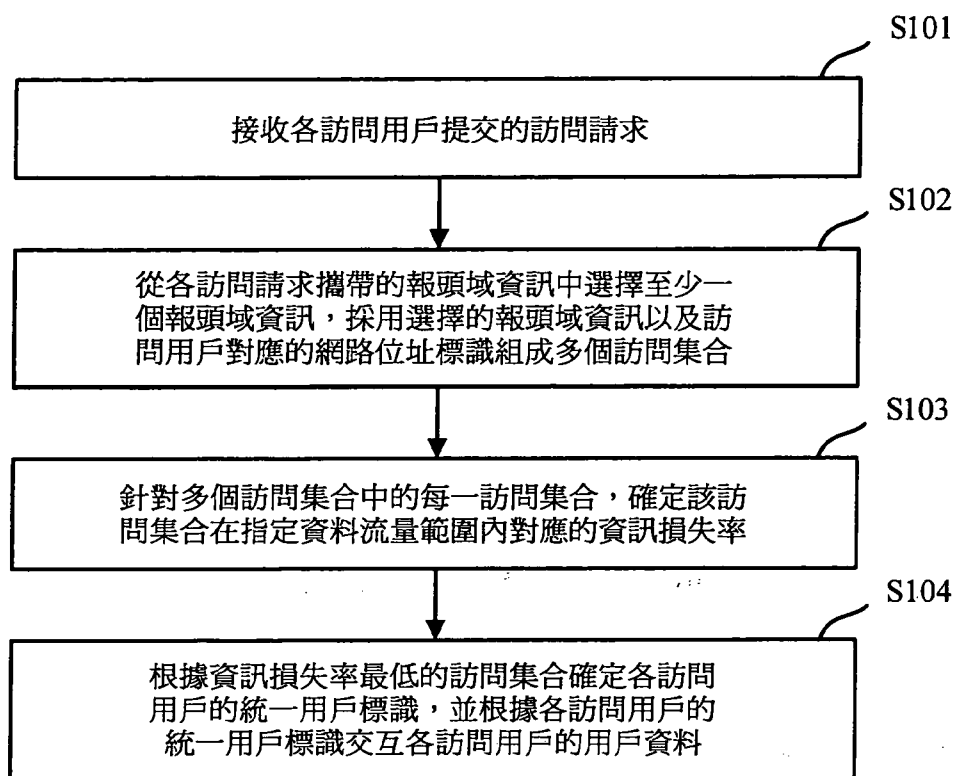


圖 1

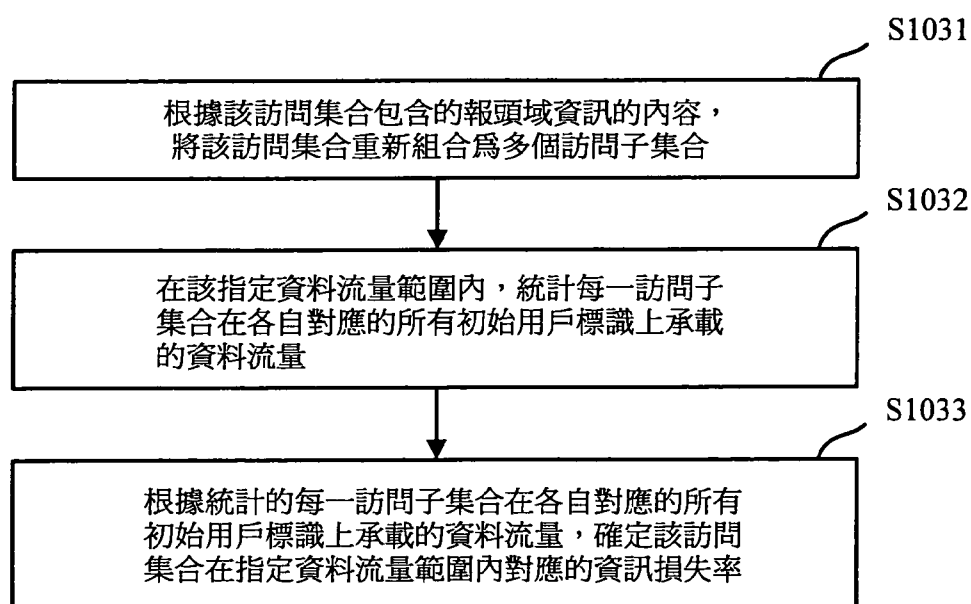


圖 2

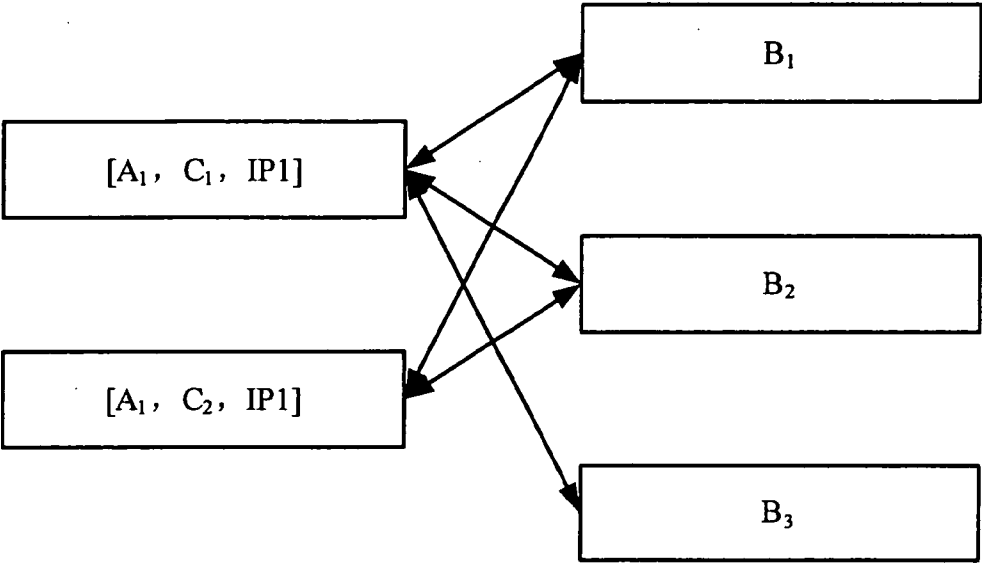


圖3

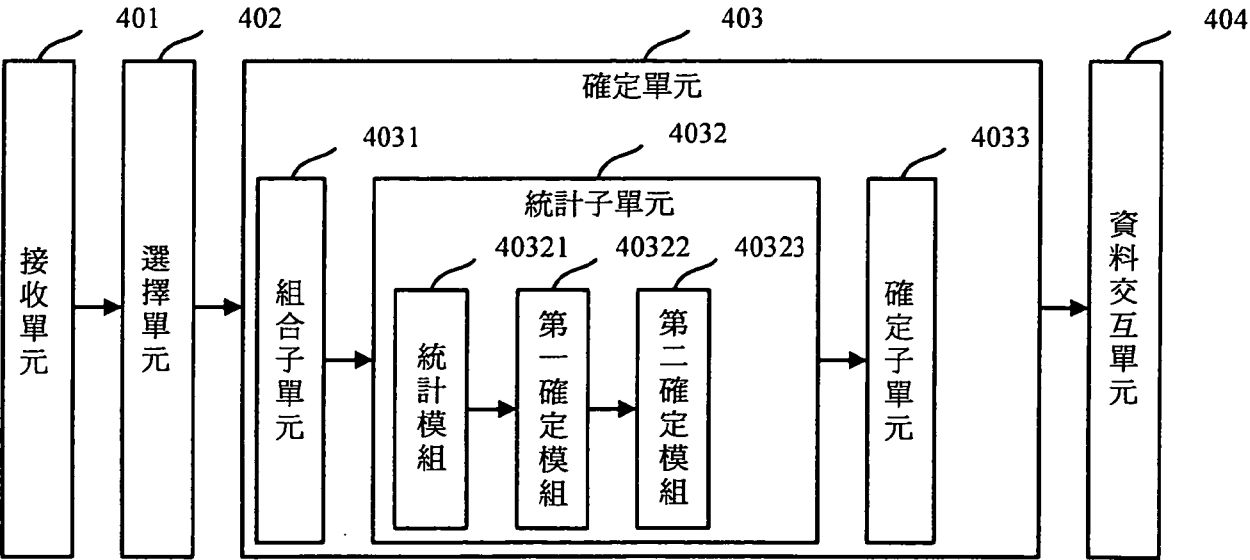


圖4