US 20140095320A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2014/0095320 A1

(43) Pub. Date: Apr. 3, 2014

Sivaramakrishnan et al.

(54) SYSTEM AND METHOD FOR DETERMINING RELATED DIGITAL IDENTITIES

(71) Applicant: **Drawbridge, Inc.**, San Mateo, CA (US)

(72) Inventors: **Kamakshi Sivaramakrishnan**, Sunnyvale, CA (US); **Devin Guan**, Monte Sereno, CA (US); **Tin H. Kyaw**, San Jose, CA (US); **Jerry Ye**, San Jose, CA (US); **Ravi Menon**, Austin, TX (US); **Yang Yu**, Redwood City, CA (US)

(73) Assignee: **Drawbridge, Inc.**, San Mateo, CA (US)

(21) Appl. No.: **13/891,764**

(22) Filed: **May 10, 2013**

Related U.S. Application Data

(60) Provisional application No. 61/645,549, filed on May 10, 2012.

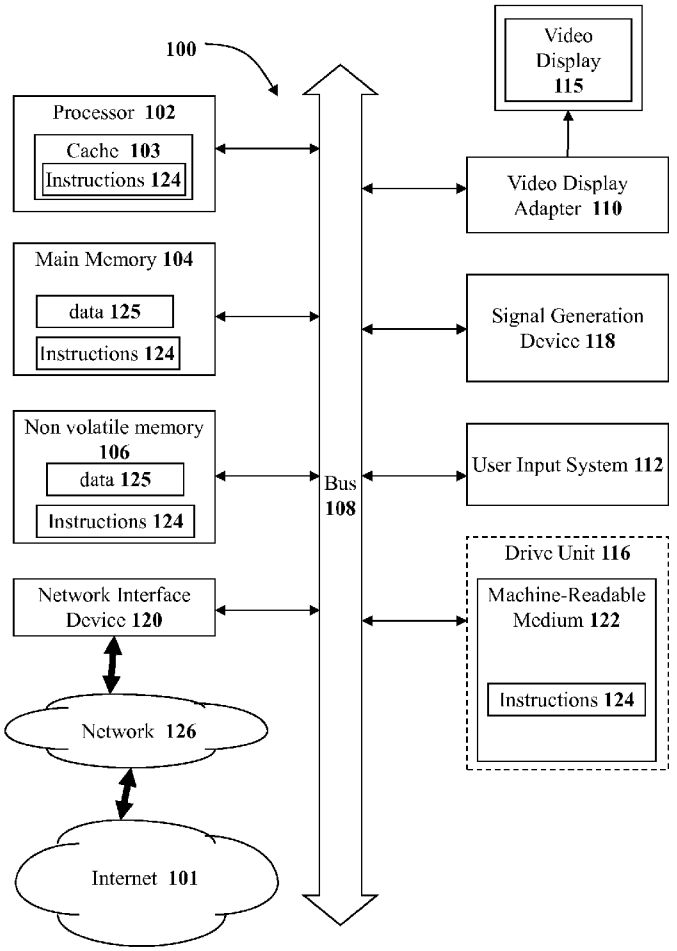
Publication Classification

(51) **Int. Cl.**
G06Q 30/02 (2006.01)

(52) **U.S. Cl.**
CPC ***G06Q 30/0269*** (2013.01)
USPC **705/14.66**

(57) **ABSTRACT**

Internet advertising to users of web browser personal computer systems is a very large and mature industry. However, many new digital devices such as cellular telephones, table computer systems, and video game console are now presenting an even larger internet advertising market. Although the techniques used for targeting advertisements to web browsers on personal computers are sophisticated, the techniques for accurately targeting internet advertisements to these new digital devices are limited. To improve the quality of targeting advertisements on new digital devices a set of techniques for accurately pairing digital identities is disclosed. Once various digital identities are linked, all of the accumulated digital profile information from these linked digital identities may be used to accurately select advertisements for all of the linked digital devices.



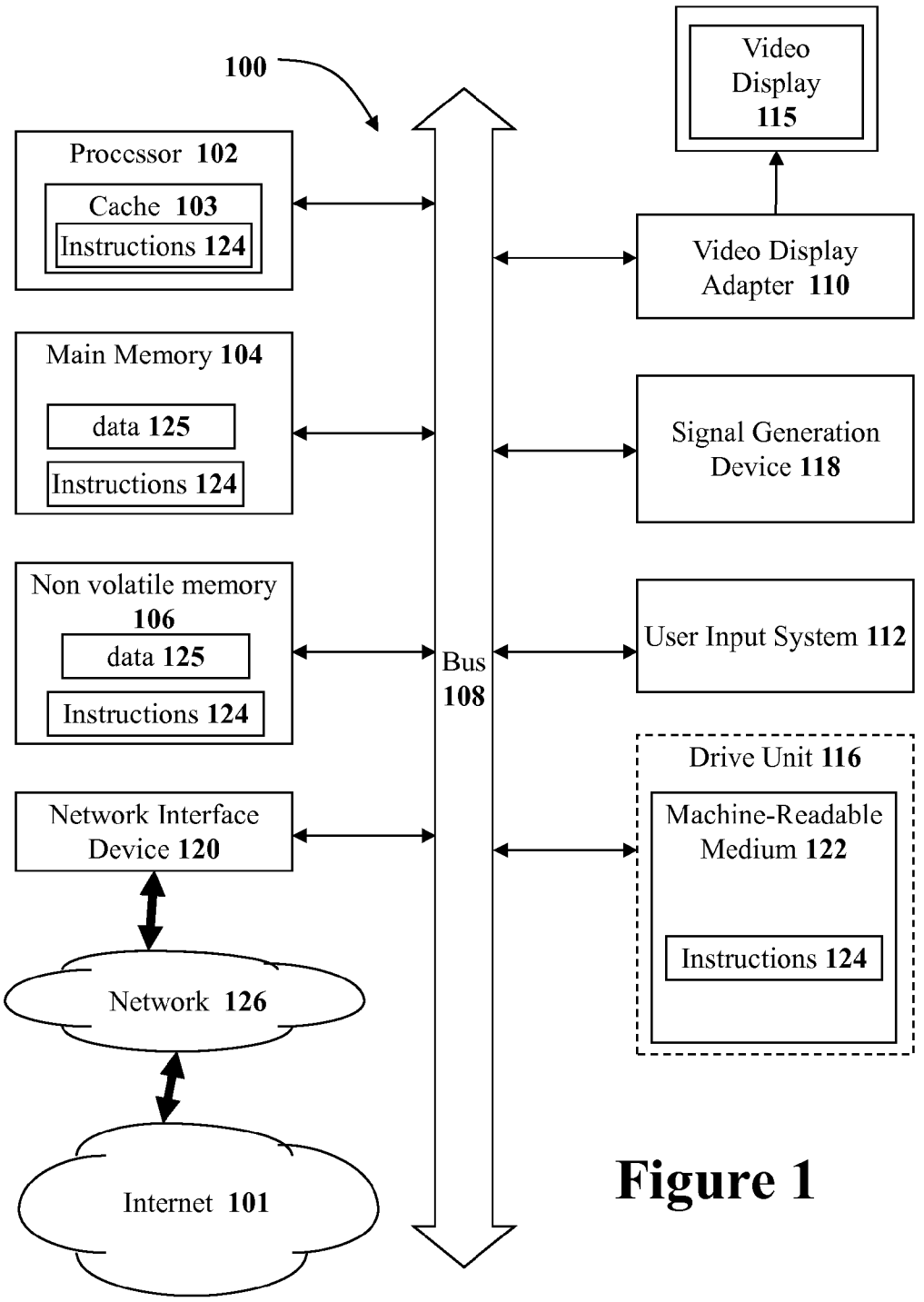

100
Video Display 115
Processor 102
Cache 103
Instructions 124
Video Display Adapter 110
Main Memory 104
data 125
Instructions 124
Signal Generation Device 118
Non volatile memory 106
data 125
Instructions 124
Bus 108
User Input System 112
Drive Unit 116
Network Interface Device 120
Machine-Readable Medium 122
Instructions 124
Network 126
Internet 101

Figure 1

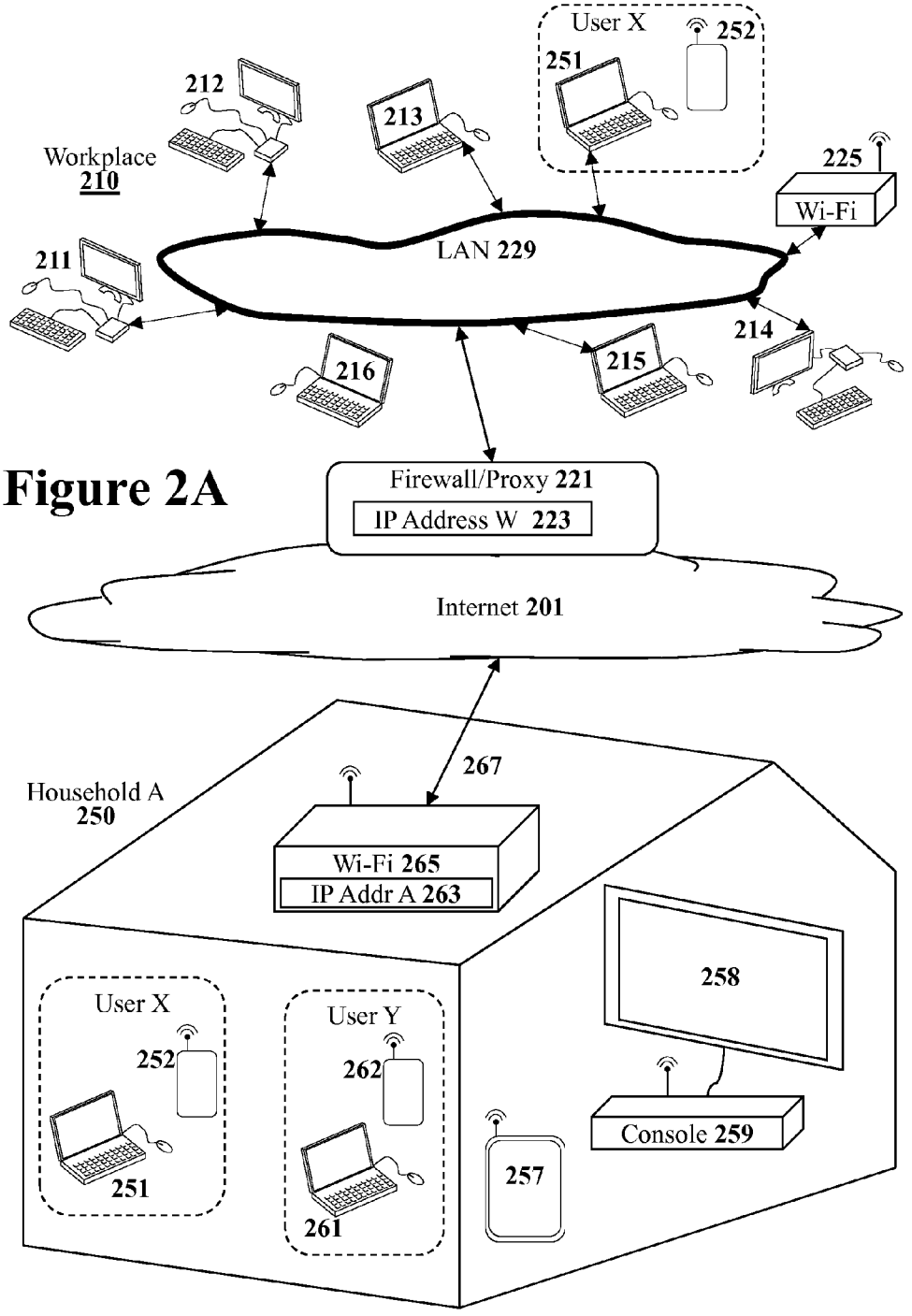

User X
252
251
212
213
Workplace
210
225
Wi-Fi
LAN 229
211
214
216
215
Firewall/Proxy 221
IP Address W 223
Internet 201
267
Household A
250
Wi-Fi 265
IP Addr A 263
258
User X
252
User Y
262
Console 259
257
251
261

Figure 2A

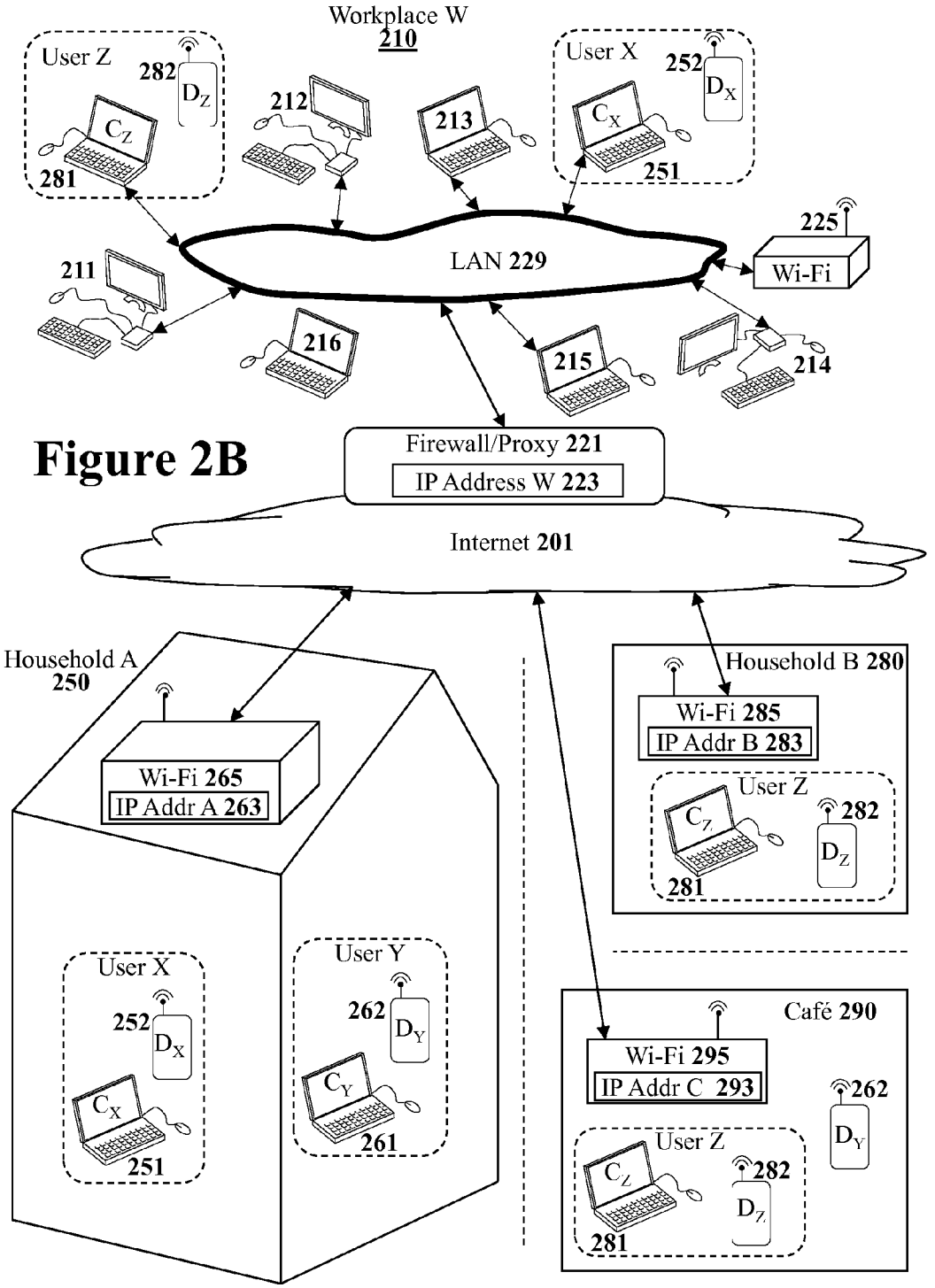

Workplace W
210
User Z
282
D_Z
C_Z
281
212
213
User X
252
D_X
C_X
251
225
Wi-Fi
LAN 229
211
216
215
214
Firewall/Proxy 221
IP Address W 223
Internet 201
Household A
250
Wi-Fi 265
IP Addr A 263
User X
252
D_X
C_X
251
User Y
262
D_Y
C_Y
261
Household B 280
Wi-Fi 285
IP Addr B 283
User Z
282
D_Z
C_Z
281
Café 290
Wi-Fi 295
IP Addr C 293
262
D_Y
User Z
282
D_Z
C_Z
281

Figure 2B

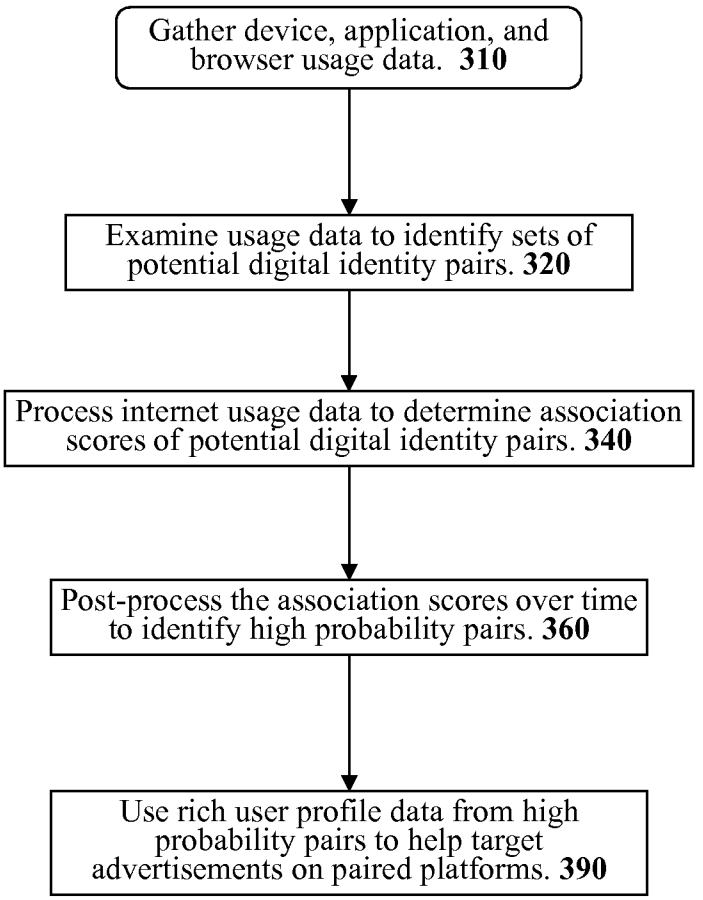

Gather device, application, and browser usage data. 310
Examine usage data to identify sets of potential digital identity pairs. 320
Process internet usage data to determine association scores of potential digital identity pairs. 340
Post-process the association scores over time to identify high probability pairs. 360
Use rich user profile data from high probability pairs to help target advertisements on paired platforms. 390

Figure 3

	Person **X**	Person **Y**	Person **Z**	Number of observations	Co-occurences
Home A IP A	$C_X=3$ $D_X=4$	$C_Y=2$ $D_Y=3$	$C_Z=0$ $D_Z=0$	12	$Co(C_X\&D_X)=3$ $Co(C_X\&D_Y)=3$ $Co(C_Y\&D_X)=2$ $Co(C_Y\&D_Y)=2$
Home B IP B	$C_X=0$ $D_X=0$	$C_Y=0$ $D_Y=0$	$C_Z=5$ $D_Z=8$	13	$Co(C_Z\&D_Z)=5$
Workplace IP W	$C_X=10$ $D_X=8$	$C_Y=0$ $D_Y=0$	$C_Z=20$ $D_Z=10$	48	$Co(C_X\&D_Z)=10$ $Co(C_X\&D_X)=8$ $Co(C_Z\&D_X)=8$ $Co(C_Z\&D_Z)=10$
Café IP C	$C_X=0$ $D_X=0$	$C_Y=7$ $D_Y=9$	$C_Z=4$ $D_Z=4$	24	$Co(C_Z\&D_Y)=4$ $Co(C_Z\&D_Z)=4$ $Co(C_Y\&D_Z)=4$ $Co(C_Y\&D_Y)=7$
				97 Total	

Figure 4

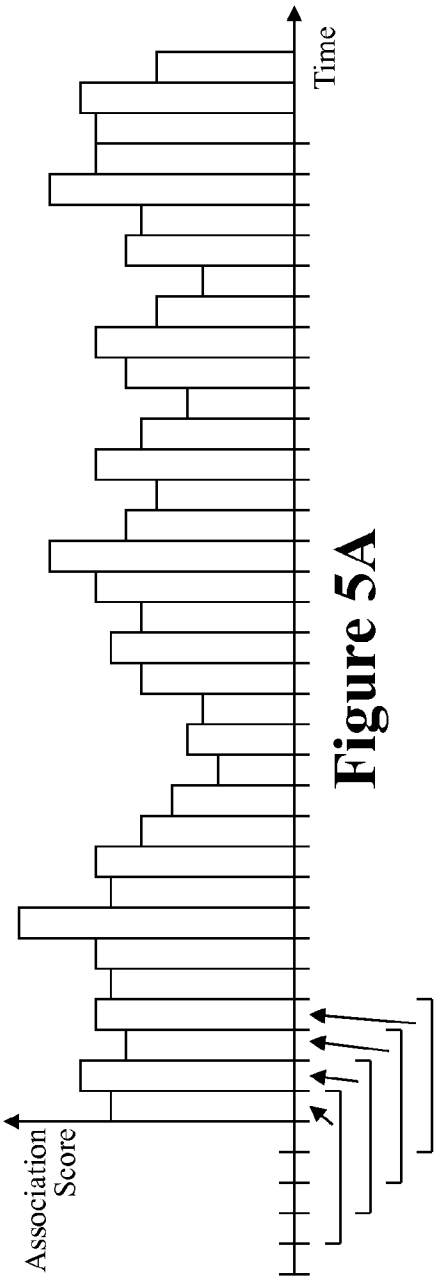


Figure 5A

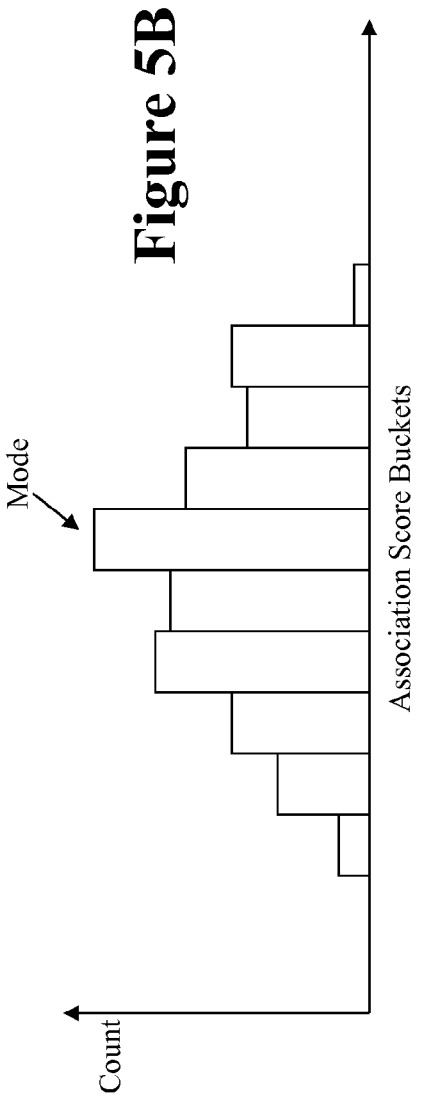


Figure 5B

	Person **X**	Person **Y**	Person **Z**	Co-occurences
Home A IP A	$C_X=1$ $D_X=1$	$C_Y=1$ $D_Y=1$	$C_Z=0$ $D_Z=0$	$Co(C_X\&D_X)=1$ $Co(C_X\&D_Y)=1$ $Co(C_Y\&D_X)=1$ $Co(C_Y\&D_Y)=1$
Home B IP B	$C_X=0$ $D_X=0$	$C_Y=0$ $D_Y=0$	$C_Z=1$ $D_Z=1$	$Co(C_Z\&D_Z)=1$
Workplace IP W	$C_X=1$ $D_X=1$	$C_Y=0$ $D_Y=0$	$C_Z=1$ $D_Z=1$	$Co(C_X\&D_Z)=1$ $Co(C_X\&D_X)=1$ $Co(C_Z\&D_X)=1$ $Co(C_Z\&D_Z)=1$
Café IP C	$C_X=0$ $D_X=0$	$C_Y=0$ $D_Y=1$	$C_Z=1$ $D_Z=1$	$Co(C_Z\&D_Y)=1$ $Co(C_Z\&D_Z)=1$

Figure 6

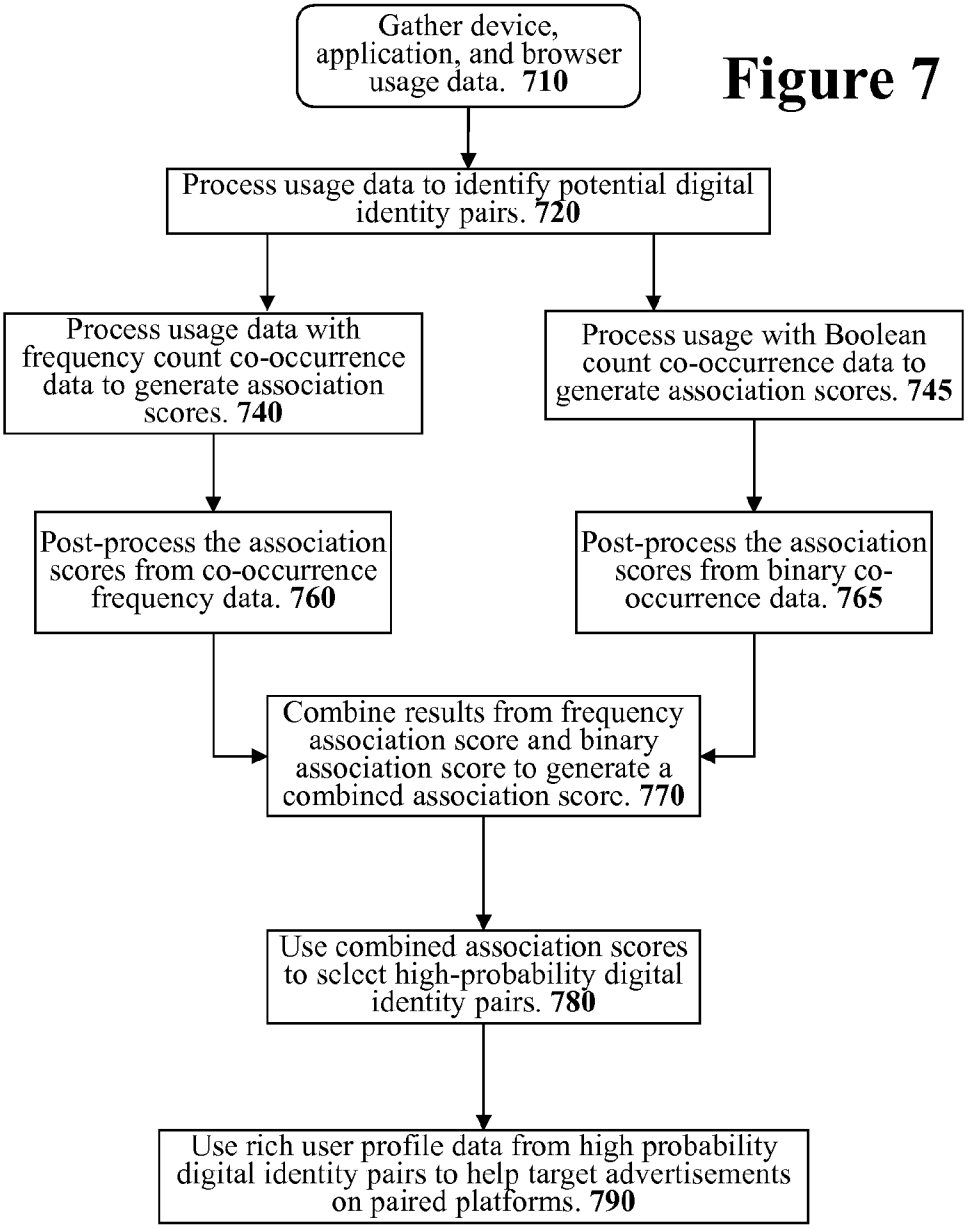

Gather device, application, and browser usage data. 710
Process usage data to identify potential digital identity pairs. 720
Process usage data with frequency count co-occurrence data to generate association scores. 740
Process usage with Boolean count co-occurrence data to generate association scores. 745
Post-process the association scores from co-occurrence frequency data. 760
Post-process the association scores from binary co-occurrence data. 765
Combine results from frequency association score and binary association score to generate a combined association score. 770
Use combined association scores to select high-probability digital identity pairs. 780
Use rich user profile data from high probability digital identity pairs to help target advertisements on paired platforms. 790

Figure 7

	Person **X**	Person **Y**	Number of observations	Co-occurences
ESPN.com	$C_X=5$ $D_X=6$	$C_Y=0$ $D_Y=0$	11	$Co(C_X\&D_X)=5$ $Co(C_X\&D_Y)=0$ $Co(C_Y\&D_X)=0$ $Co(C_Y\&D_Y)=0$
CNN.com	$C_X=0$ $D_X=2$	$C_Y=10$ $D_Y=9$	21	$Co(C_X\&D_X)=0$ $Co(C_X\&D_Y)=0$ $Co(C_Y\&D_X)=2$ $Co(C_Y\&D_Y)=9$
Finance.com	$C_X=2$ $D_X=0$	$C_Y=7$ $D_Y=5$	14	$Co(C_X\&D_X)=0$ $Co(C_X\&D_Y)=2$ $Co(C_Y\&D_X)=0$ $Co(C_Y\&D_Y)=5$
Fark.com	$C_X=4$ $D_X=6$	$C_Y=0$ $D_Y=0$	10	$Co(C_X\&D_X)=4$ $Co(C_X\&D_Y)=0$ $Co(C_Y\&D_X)=0$ $Co(C_Y\&D_Y)=0$
			56 Total	

Figure 8

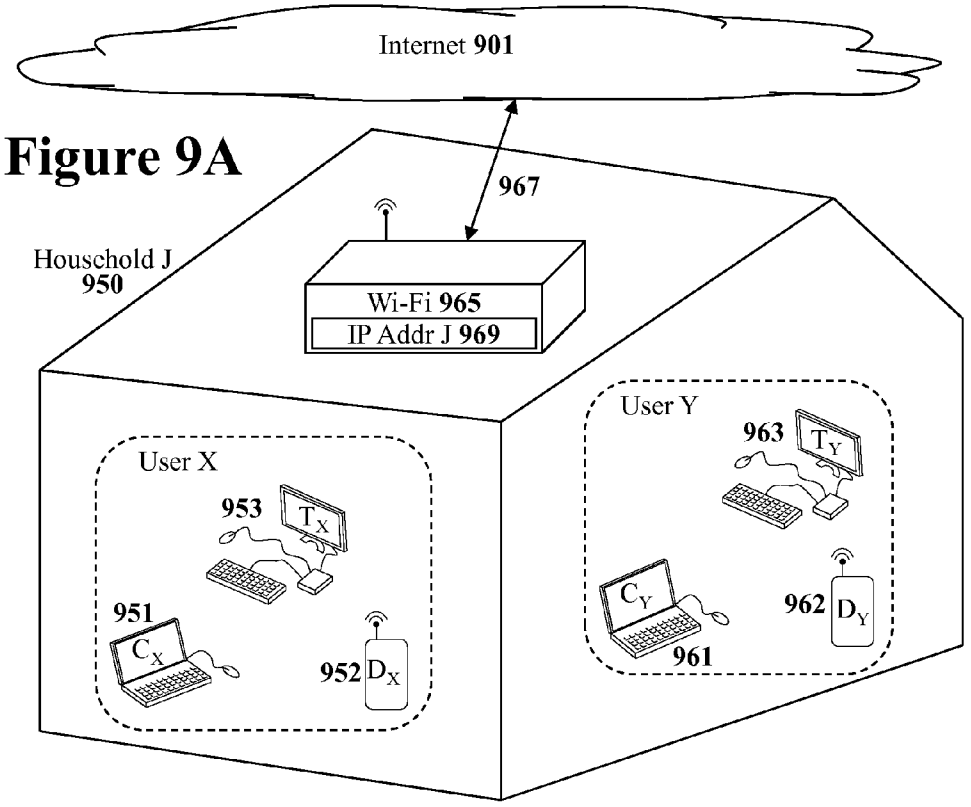


Figure 9A

Figure 9B

	Fashion	News	Finance	Other web
Person **X**	C_X= 40% T_X= 35% DI_X= 0%	C_X= 20% T_X= 25% DI_X= 20%	C_X= 10% T_X= 15% DI_X= 10%	C_X= 30% T_X= 25% DI_X= 10%
Person **Y**	C_Y= 0% T_Y= 0% DI_Y= 0%	C_Y= 30% T_Y= 25% DI_Y= 9%	C_Y= 30% T_Y= 35% DI_Y= 5%	C_Y= 40% T_Y= 40% DI_Y= 0%

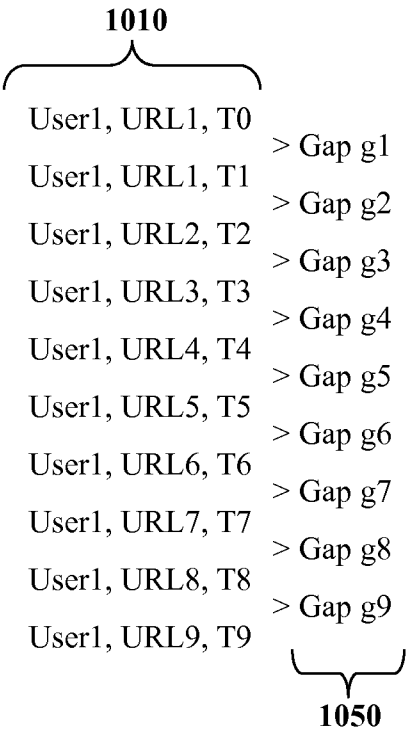

1010
User1, URL1, T0
> Gap g1
User1, URL1, T1
> Gap g2
User1, URL2, T2
> Gap g3
User1, URL3, T3
> Gap g4
User1, URL4, T4
> Gap g5
User1, URL5, T5
> Gap g6
User1, URL6, T6
> Gap g7
User1, URL7, T7
> Gap g8
User1, URL8, T8
> Gap g9
User1, URL9, T9
1050

Figure 10A

	Person **X**	Person **Y**	Number of observations	Co-occurences
Session1	$C_X=4$ $D_X=2$	$C_Y=0$ $D_Y=0$	6	$Co(C_X\&D_X)=2$ $Co(C_X\&D_Y)=0$ $Co(C_Y\&D_X)=0$ $Co(C_Y\&D_Y)=0$
Session2	$C_X=0$ $D_X=0$	$C_Y=15$ $D_Y=12$	27	$Co(C_X\&D_X)=0$ $Co(C_X\&D_Y)=0$ $Co(C_Y\&D_X)=0$ $Co(C_Y\&D_Y)=12$
Session3	$C_X=1$ $D_X=0$	$C_Y=9$ $D_Y=7$	17	$Co(C_X\&D_X)=0$ $Co(C_X\&D_Y)=1$ $Co(C_Y\&D_X)=0$ $Co(C_Y\&D_Y)=7$
Session4	$C_X=14$ $D_X=9$	$C_Y=0$ $D_Y=1$	24	$Co(C_X\&D_X)=4$ $Co(C_X\&D_Y)=1$ $Co(C_Y\&D_X)=0$ $Co(C_Y\&D_Y)=0$
			74 Total	

Figure 10B

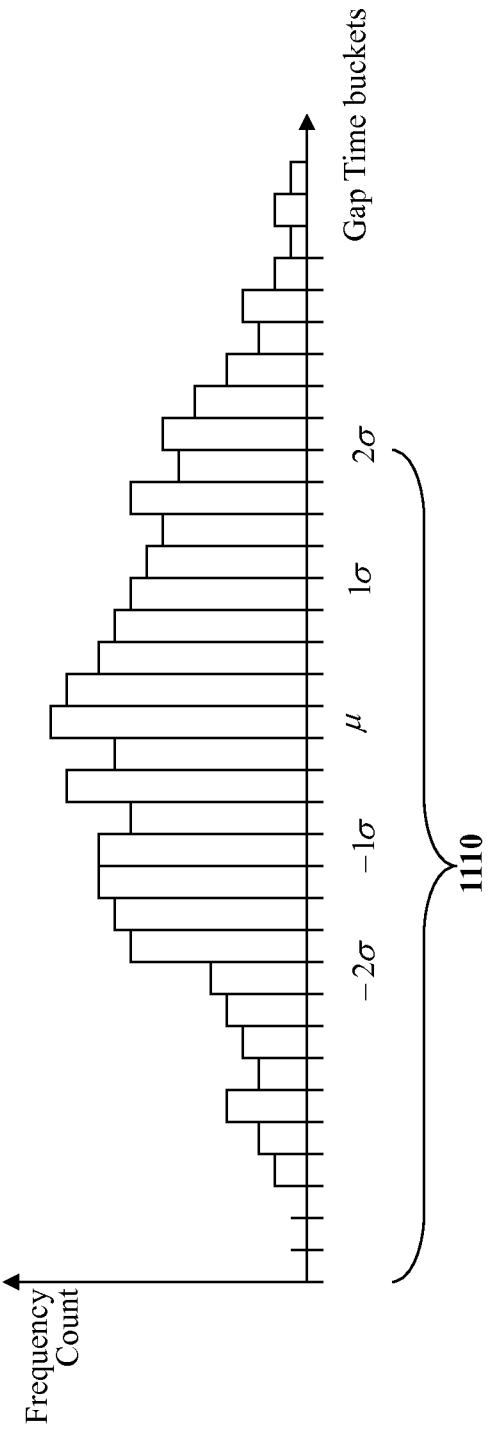

Frequency Count
Gap Time buckets
-2σ
-1σ
μ
1σ
2σ
1110

Figure 11

SYSTEM AND METHOD FOR DETERMINING RELATED DIGITAL IDENTITIES

RELATED APPLICATIONS

[0001] The present Nonprovisional U.S. Patent Application claims the benefit of the previous U.S. Provisional Patent Application entitled "System and Method for Determining Related Digital Identities" filed with the U.S. Patent Office on May 10, 2012 and having Ser. No. 61/645,549.

TECHNICAL FIELD

[0002] The present invention relates to the field of digital identities and digital advertising. In particular, but not by way of limitation, the present disclosure teaches techniques for determining related digital identities.

BACKGROUND

[0003] The internet has changed the mass media landscape forever. Before the internet became a mainstream mass media system, advertisers were generally limited to communicating with potential customers using television, radio, and print media (newspapers and magazines) advertising. With the popularization of the global internet, advertisers can now advertise to billions of computer users as those computer users browse the World Wide Web on the internet.

[0004] Internet advertising has become a very large industry. Two of the most commonly used advertising channels on the World Wide Web are internet search advertising and banner advertisements. Internet search advertising operates by allowing users to enter search keywords into an internet search service and then interspersing advertisements (generally related to the search keywords) within the results of the internet search. Banner advertisements are defined areas of a web page that contain advertisements in the same manner that traditional magazines and newspapers use newsprint area for advertising. Both internet search result advertisements and internet banner advertisements have a significant advantage over prior advertising systems since the recipient of an internet advertisement may click on the internet advertisement to obtain more information or directly proceed to an internet retailer for an immediate sale.

[0005] The internet advertising industry for advertising on personal computer systems on the internet has matured and become very sophisticated. The internet advertisers and internet advertising services use several techniques of obtaining information about internet users such that the most appropriate advertisements may be selected for each internet user. For example, internet advertisement services may track the web browsing history from particular personal computer to determine the interests of that user and thus create a demographic profile of that internet user. Furthermore, the contents of a web page that is being delivered to personal computer may be analyzed to help select an appropriate banner advertisement that is closely related to the contents of the web page.

[0006] Although internet advertising to personal computer users that are browsing the World Wide Web has become relatively sophisticated, the overall internet advertising industry is still in its infancy. There are now many new digital electronic devices that use the internet and can be used to deliver advertising to their users. For example, cellular telephones, video game consoles, set-top video streaming boxes, internet radio devices, and tablet computer systems can all be used to deliver internet advertisements to their respective users. The techniques used to select and deliver advertisements to the users of these emerging internet platforms are relatively primitive. Thus, it would be desirable to provide tools that provide an improved ability to select appropriate internet advertisements to users of these new internet-connected digital electronic devices.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] In the drawings, which are not necessarily drawn to scale, like numerals describe substantially similar components throughout the several views. Like numerals having different letter suffixes represent different instances of substantially similar components. The drawings illustrate generally, by way of example, but not by way of limitation, various embodiments discussed in the present document.

[0008] FIG. **1** illustrates a diagrammatic representation of machine in the example form of a computer system within which a set of instructions, for causing the machine to perform any one or more of the methodologies discussed herein, may be executed.

[0009] FIG. **2**A conceptually illustrates a household with several digital devices that access the internet which are potential digital identity pairs.

[0010] FIG. **2**B conceptually illustrates the digital devices of three different people used at four different locations: household A, household B, a workplace W, and a cyber café.

[0011] FIG. **3** illustrates a high-level flow diagram that describes one embodiment of a digital identity pairing system.

[0012] FIG. **4** illustrates an example set of observation samples of the various digital identities (C_X, D_X, C_Y, D_Y, C_Z, D_Z) at four different source locations (household A, household B, workplace W, and a cyber café) depicted in FIG. **2**B.

[0013] FIG. **5**A graphically illustrates an example of a series of daily association scores for a particular potential digital identity pair calculated over time.

[0014] FIG. **5**B graphically illustrates the daily association scores of FIG. **5**A collected in a set of association score range buckets to identify a statistical mode.

[0015] FIG. **6** illustrates the observation data from FIG. **4** but in Boolean form for use in calculating a Boolean system association score.

[0016] FIG. **7** illustrates a flow diagram describing how a digital pairing system may use a weighted score of frequency count association scores and Boolean system association scores to identify high-probability digital identity pairs.

[0017] FIG. **8** illustrates a set of destination usage observations for four different digital devices illustrated in FIG. **2**B.

[0018] FIG. **9**A conceptually illustrates a household with two users that each use desktop computer, a laptop computer, and a cellular smartphone.

[0019] FIG. **9**B conceptually illustrates website usage patterns for the computers in FIG. **9**A and inferred website usage patterns for the cellular smartphones in FIG. **9**A.

[0020] FIG. **10**A illustrates a log of usage triads wherein each triad contains a client identifier, destination, and an associated timestamp.

[0021] FIG. **10**B illustrates a set of sessions for the laptop computers C_X and C_Y in household A of FIG. **2**B.

[0022] FIG. **11** graphically illustrates all of the observed gap times are placed into gap time buckets as illustrated in.

DETAILED DESCRIPTION

[0023] The following detailed description includes references to the accompanying drawings, which form a part of the detailed description. The drawings show illustrations in accordance with example embodiments. These embodiments, which are also referred to herein as "examples," are described in enough detail to enable those skilled in the art to practice the invention. It will be apparent to one skilled in the art that specific details in the example embodiments are not required in order to practice the present invention. For example, although some of the embodiments are mainly disclosed with reference to cellular telephones, the techniques disclosed in this document may be used with other types of digital electronic devices such as tablet computer systems and video game systems. The example embodiments may be combined, other embodiments may be utilized, or structural, logical and electrical changes may be made without departing from the scope of what is claimed. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope is defined by the appended claims and their equivalents.

[0024] In this document, the terms "a" or "an" are used, as is common in patent documents, to include one or more than one. In this document, the term "or" is used to refer to a nonexclusive or, such that "A or B" includes "A but not B," "B but not A," and "A and B," unless otherwise indicated. Furthermore, all publications, patents, and patent documents referred to in this document are incorporated by reference herein in their entirety, as though individually incorporated by reference. In the event of inconsistent usages between this document and those documents so incorporated by reference, the usage in the incorporated reference(s) should be considered supplementary to that of this document; for irreconcilable inconsistencies, the usage in this document controls.

[0025] Computer Systems

[0026] The present disclosure concerns digital computer systems. FIG. **1** illustrates a diagrammatic representation of a machine in the example form of a computer system **100** that may be used to implement portions of the present disclosure. Within computer system **100** of FIG. **1**, there are a set of instructions **124** that may be executed for causing the machine to perform any one or more of the methodologies discussed within this document.

[0027] In a networked deployment, the machine of FIG. **1** may operate in the capacity of a server machine or a client machine in a client-server network environment, or as a peer machine in a peer-to-peer (or distributed) network environment. The machine may be a personal computer (PC), a tablet computer, a set-top box (STB), a Personal Digital Assistant (PDA), a cellular telephone, a web appliance, a server, a network router, a network switch, a network bridge, a video game console, or any machine capable of executing a set of computer instructions (sequential or otherwise) that specify actions to be taken by that machine. Furthermore, while only a single machine is illustrated, the term "machine" shall also be taken to include any collection of machines that individually or jointly execute a set (or multiple sets) of instructions to perform any one or more of the methodologies discussed herein.

[0028] The example computer system **100** of FIG. **1** includes a processor **102** (e.g., a central processing unit (CPU), a graphics processing unit (GPU) or both), a main memory **104**, and a non volatile memory **106**, which communicate with each other via a bus **108**. The non volatile memory **106** may comprise flash memory and may be used either as computer system memory, as a file storage unit, or both. Both the main memory **104** and a non volatile memory **106** may store instructions **124** and data **125** that are processed by the processor **102**.

[0029] The computer system **100** may include a video display adapter **110** that drives a video display system **115** such as a Liquid Crystal Display (LCD) in order to display visual output to a user. The computer system **100** may also include other output systems such as signal generation device **118** that drives an audio speaker.

[0030] Computer system **100** includes a user input system **112** for accepting input from a human user. The user input system **112** may include an alphanumeric input device such as a keyboard, a cursor control device (e.g., a mouse or trackball), touch sensitive pad (that may be overlaid on top of video display **115**), a microphone, or any other device for accepting input from a human user.

[0031] The computer system **100** may include a disk drive unit **116** for storing data. The disk drive unit **116** includes a machine-readable medium **122** on which is stored one or more sets of computer instructions and data structures (e.g., instructions **124** also known as 'software') embodying or utilized by any one or more of the methodologies or functions described herein. The instructions **124** may also reside, completely or at least partially, within the main memory **104** and/or within a cache memory **103** associated with the processor **102**. The main memory **104** and the non-volatile memory **106** associated with the processor **102** also constitute machine-readable media. The non-volatile memory **106** may comprise a removable flash memory device.

[0032] The computer system **100** may include one more network interface devices **120** for transmitting and receiving data on one or more networks **126**. For example wired or wireless network interfaces **120** may couple to a local area network **126**. Similarly, a cellular telephone network interface **120** may be used to couple to a cellular telephone network **126**. The various different networks **126** are often coupled directly or indirectly to the global internet **101**. The instructions **124** and data **125** used by computer system **100** may be transmitted or received over network **126** via the network interface device **120**. Such transmissions may occur utilizing any one of a number of well-known transfer protocols such as the well known File Transport Protocol (FTP).

[0033] Note that not all of the parts illustrated within FIG. **1** will be present in all embodiments. For example, a computer server system may not have a video display adapter **110** or video display system **115** if that server is controlled through the network interface device **120**. Similarly, a tablet computer or cellular telephone will generally not have a disk drive unit **116** and instead use flash memory or another form of long-term storage.

[0034] While the machine-readable medium **122** is shown in an example embodiment to be a single medium, the term "machine-readable medium" should be taken to include a single medium or multiple media (e.g., a centralized or distributed database, and/or associated caches and servers) that store the one or more sets of instructions. The term "machine-readable medium" shall also be taken to include any medium that is capable of storing, encoding or carrying a set of instructions for execution by the machine and that cause the machine to perform any one or more of the methodologies described herein, or that is capable of storing, encoding or carrying data structures utilized by or associated with such a set of instruc-

tions. The term "machine-readable medium" shall accordingly be taken to include, but not be limited to, solid-state memories, optical media, battery-backed RAM, and magnetic media.

[0035] For the purposes of this specification, the term "module" includes an identifiable portion of code, computational or executable instructions, data, or computational object to achieve a particular function, operation, processing, or procedure. A module need not be implemented in software; a module may be implemented in software, hardware/circuitry, or a combination of software and hardware.

[0036] Internet Advertising

[0037] The global internet has become a mass medium that connects publishers of information with consumers of information. Some of the publishing on the internet is done on a subscription or paid-for basis wherein a consumer pays for access to specific information. For example, a news publisher may create news web site that provides specific premium content only to customers that pay a subscription fee. However, a very large portion of the informational content available on the global internet is freely available. Web site content, videos, podcasts, and games are all freely available on the internet. To fund much of the freely available content on the global internet, internet publishers rely on advertisers that pay to have their advertisements displayed alongside and embedded within an internet publisher's content.

[0038] Although internet advertising started with advertisements to personal computer users browsing the World Wide Web with web browser programs, the internet advertising market has grown significantly beyond that early stage. FIG. **2**A conceptually illustrates a household A **250** that is coupled to the global internet **201** through an internet connection **267**. Household A **250** uses a wireless network access point **265** to provide internet access to many different digital electronic devices within household A **250**. An internet user may use a laptop computer **251** or **261** for traditional personal computer based browsing of internet web sites that are supported by internet advertising. However, all of the other internet-connected devices in household A **250** may also be used to display internet advertisements. Cellular telephones **252** and **262** may display banner advertisements in web pages displayed by a micro browser or within advertisement-supported application programs such as games downloaded onto the cellular phones. Tablet computer system **257** may also display internet advertisements in a web browser or in other application programs that are supported by advertising. A video game console **259** may display advertisements within a navigator application or even within the video games that users play on the video game console **259**.

[0039] Thus, as illustrated in FIG. **2**A, many different digital electronic devices within household A **250** may be served advertisements from internet advertisers. In the future, many new internet-connected digital devices will be introduced and each new internet-connected digital device may be used to display internet advertisements to its user. Note that all of the internet-connected digital devices within household A **250** will generally share a single Internet Protocol (IP) address **263** that is assigned to the Wi-Fi router device **265** that couples household A **250** to the global internet **201**.

[0040] In order to really provide substantial value to internet advertisers, internet advertising must be well targeted. Advertising suntan lotion during the winter to a person living in Minnesota does not provide much value to the suntan lotion manufacturer. Thus, internet advertising services attempt to learn as much as possible about their audience in order to select the most appropriate internet advertisements. As set forth in the background, internet web site publishers that advertise to users running web browsers on personal computers have developed many different techniques for learning about their users. These techniques may include:

[0041] Generating user profiles based on web browsing histories

[0042] Using search terms entered into search engines to target advertisements

[0043] Using the content of web pages to select matching advertisements

[0044] Creating user registration systems where users divulge demographic data

[0045] Using information on the PC to help select advertisements

[0046] However, these advertising targeting techniques developed for personal computer based internet advertising often cannot be used with other internet-connected digital devices. For example, there may be no web browsing history on the video game console **259** of household A **250**. Even though the cellular telephones **252** and **262** and the tablet computer system **257** may have web browsing histories to draw from, the users may only rarely use those devices for web browsing such that their very limited browsing history does not provide an accurate demographic profile of the user to guide advertisement selection. It is therefore often much more difficult to provide well-targeted advertising to internet-connected digital devices other than personal computer systems.

[0047] Digital Identity Pairing Overview

[0048] Referring to FIG. **2**A, the people that live at household A **250** use several mobile internet-connected digital devices to communicate with the internet **201**. Specifically, the laptop computer systems (**251** and **261**), the cellular phones (**252** and **262**), and the tablet computer system **257** are all mobile internet-connected digital devices. These mobile devices can be and often are taken with the user to other locations that provide wireless (or wired) internet access. For example, user X in the household may use laptop computer system **251** and cellular phone **252** both at household **250** and at a workplace W **210**.

[0049] When user X brings laptop computer system **251** and cellular phone **252** to workplace W **210** those two devices obviously can no longer use Wi-Fi router **265** to access the internet **201**. Instead, user X connects laptop computer system **251** to the local area network **229** at work and configures cellular phone **252** to use a local Wi-Fi network provided by wireless access point **225**. With these two internet connections, user X will then be able to access the internet **201** using laptop computer system **251** and cellular phone **252** through the firewall/proxy server **221** at the workplace **210**.

[0050] When user X is at home A **250**, the laptop computer system **251** and the cellular phone **252** will both use a single IP address A **263** that is on Wi-Fi router **265**. Similarly, when that same user X is at workplace **210**, laptop computer system **251** and cellular phone **252** will both use the single IP address W **223** that is on firewall/proxy server **221**. Given that specific internet usage pattern data, an astute observer that knows nothing about user X could make the rational inference that laptop computer system **251** and cellular phone **252** are very likely used by the same person since laptop computer system **251** and cellular phone **252** are used together at both household **250** and at workplace **210**. After having made such an

inference, an advertiser may link together a digital identifier associated with laptop computer system **251** and a digital identifier associated with cellular phone **252** for advertising purposes. Such a pairing of distinct digital identities to a single user is referred to as digital identity pairing.

[0051] Assuming that a digital identity pairing has been performed accurately (the two paired digital devices are actually used by the same person) then this digital identity pairing may be used to greatly improve the targeting of internet advertising to both of the linked platforms. For example, if an advertising service correctly deduces that laptop computer system **251** and cellular phone **252** are used by the same person then the advertiser can advertise cellular telephone accessories for the specific brand of cellular phone **252** when that person uses their laptop computer system **251** to browse the World Wide Web. More importantly, when that person (user X) uses an advertising supported application on cellular phone **252**, advertisers can leverage the longer and more detailed user profile data that has been collected from laptop computer system **251** to accurately select targeted advertisements for display on that cellular phone **252**. Thus, digital identity pairing can greatly improve the quality of internet advertising on digital devices that have been correctly identified as belonging to the same user.

[0052] Digital Identity Pairing Basics

[0053] Referring to FIG. **2**A, there many different internet-connected devices (**251**, **252**, **261**, **262**, **257**, and **259**) at household A **250** that may access the internet **201**. All of these internet-connected devices at household A **250** will use the same IP address A **263** that is on Wi-Fi router **265**. Similarly, when the user of laptop computer system **251** and cellular phone **252** is at workplace **210** those two devices will have the same IP address W **223** as all of the other internet connected devices (**211**, **212**, **213**, **214**, **215**, and **225**) that share the connection to the internet **201** through firewall/proxy device **221**. Therefore, the task of specifically identifying laptop computer system **251** and cellular phone **252** as a pair of related digital devices used by the same human user is a difficult task. To make the association, sophisticated analysis is performed on the server logs of internet web site servers and internet advertisement servers that collect information on every internet-connected device that makes any type of request to the server.

[0054] FIG. **3** illustrates a high-level flow diagram that describes how a digital identity pairing system may operate in one embodiment. Initially, at stage **310**, internet usage data is collected for many different digital identities. In one embodiment, the basic internet usage data that is collected for analysis are data triads consisting of a client identifier, a common source/destination identifier, and a timestamp. Each part of the data triad will be individually described.

[0055] The client identifier is used to identify a specific client device, a client program, a user identity, or any other type of digital identity. Examples of client identifiers include web browser cookies, cellular telephone device identifiers, MAC addresses, userids, and any other similar identifier that is linked to a specific client device, client program, or user. The teachings of the present disclosure may be used with a wide variety of different client identifiers. In example digital identity pairings that will be disclosed with reference to FIGS. **2**A and **2**B, web browser cookies on laptop computer systems and device identifiers on cellular phone devices are used as client identifiers. However, the disclosed techniques may be used with any other suitable client identifiers that can be used to identify specific client devices, web browsers, users, or other digital identities.

[0056] The common source/destination identifier is the identity of some source or destination that client devices (as identified by their client identifiers) will likely have in common if the two client devices are related. In the situation depicted in FIG. **2**A, the Internet Protocol (IP) addresses are the common source/destination identifier that may be used to link related client devices. Specifically, the digital identity pairing system may use the fact that laptop computer system **251** and cellular phone **252** both share the IP address W **223** when at work and the IP address A **263** when at home to deduce that laptop computer system **251** and cellular phone **252** are related client devices.

[0057] The timestamps in each data triad may be used to ensure that the data used is relevant. The ownership of internet connected devices may change over time such that very old internet usage data should not be used. Furthermore, many Internet Protocol addresses are "dynamic addresses" that may be used by different entities at different times. Thus, internet usage data observations should have relatively close temporal relations in order to provide accurate digital identity pairing results. In addition to ensuring that internet usage observations are temporally proximate, certain embodiments of the disclosed system use the timestamps of internet usage data triads in a more sophisticated manner as will be disclosed in a later section of this document.

[0058] The triads of internet usage data (client identifier, common source/destination identifier, and timestamp) may be collected by internet servers that track each internet server request received. In particular, internet advertisement services and internet web publishers that track each advertisement or web page served are excellent sources of internet usage data information. Individual application programs (such as games, media aggregators, utilities, etc.) that run on client devices and report usage information to servers on the internet are also excellent sources of usage data.

[0059] Referring back to FIG. **3**, after collecting internet usage data the next step in digital identity pairing is to determine a set of potential digital identity pairs. There are many millions of different digital identities involved in internet activities every day. Attempting to analyze every possible permutation of digital identities as a potential digital identity pair would be an extremely difficult and probably futile task. Thus, to reduce the size of the digital identity pairing problem, the gathered internet usage data is analyzed at stage **320** to identify a much smaller number of potential digital identity pairs that have a decent probability of being related.

[0060] In embodiments that use IP addresses as common source/destination identifiers, two different techniques have been used to select potential digital identity pairs for further analysis. A first strategy is to examine the number of different digital identities known to use the same IP address. Specifically, if less than a threshold number of digital identities are known to use a specific IP address then all of the different logical pairings of digital identities from that single IP address may be viewed as potential digital identity pairs. The reasoning is that if there are just a few different digital identities related to a single common IP then there is a good probability that some of those different digital identities are associated with the same person and that one may be able to statistically link the digital identities belonging to that same person. For example, a family household that shares a single

internet account will likely have family members that use more than one digital identity that can be statistically linked.

[0061] In one embodiment the threshold value is set to six such that if there are six or less digital identities seen at a particular IP address then various logical combinations of those six or less digital identities may be considered potential digital identity pairs. For example, in FIG. **2**A household A **250** has only six different digital devices (**251**, **252**, **261**, **262**, **257**, and **259**) that couple to the internet **201** though a single IP address A **263** on Wi-Fi router **265** such that the various digital devices in household A **250** may be considered as potential digital identity pairs. In contrast, a very large number of digital devices couple to the internet **201** though the single IP address W **223** at workplace W **210** such that the digital identity pairing system does not immediately consider all of the combinations of digital devices at workplace **210** as potential digital identity pairs. In effect, the system identifies family households (which often have less than six internet connected devices) and then attempts to pair up digital devices from the family household that are used by the same user.

[0062] In another embodiment, the digital identity pairing system considers the specific IP address origin and determines if that IP address is an address where paired digital identities are likely to be found (such as household residences as set forth above). All of the static IP addresses on the internet are allocated by the Internet Corporation for Assigned Names and Numbers (ICANN). By examining who owns a particular IP address, one may make a determination as to whether it will be easy to identify related digital identities may be located at that IP address. Thus, for example, IP addresses that are used by an internet service provider (ISP) to provide residential internet service may be good IP addresses to use when identifying potential digital identity pairs.

[0063] Various other systems of identifying residential household IP addresses may also be used. In addition, other techniques of identifying likely digital identity pairs may also be used in step **330** in addition to or instead of the systems for identifying residential households.

[0064] After selecting sets of potential digital identity pairs at stage **330**, the digital identity pairing system then processes the gathered internet usage data at stage **340** to determine association scores for the potential digital identity pairings. Those digital identity pairings with the most favorable association scores will be deemed most likely to be associated with the same human user. Many different techniques may be used to calculate association scores for the potential digital identity pairings. A detailed explanation of one particular method of calculating association scores (and other related metrics that may also be used) is presented in a later section of this document but various other different scoring systems may be used.

[0065] Since the observed internet usage data will vary over time and certain chance activities may cause false digital identity associations to be detected, the association scores may be post-processed to remove noise. For example, association scores may be smoothed out over time using various techniques. Thus, at stage **360**, the association score data that has been generated over time may be post-processed such that outlier data points are largely filtered out. The end result of stage **360** is a set of high probability digital identity pairings.

[0066] Finally, at stage **390**, the identified high probability digital identity pairs are used to improve the targeting of internet advertising to both of the paired digital identities. The accumulated profile data from the two separate digital identities may be combined in a synergistic manner to provide a detailed profile of the human user associated with the paired digital identities. This detailed profile may then be used to select the best internet advertisements for that human user when either of the digital identities requires an internet advertisement.

[0067] A Basic Digital Identity Pairing Method

[0068] As set forth in the flow diagram of FIG. **3**, one embodiment of the digital identity pairing system first identifies pairs of digital identities that have a good probability of being a related pair (in stage **330**) and then processes the collected internet usage data for those two digital identities to generate an 'association score' for the potential pair (in stage **340**). The calculated association score may be used to determine if the two digital identities are deemed to be associated with the same human user. To describe how association scores are calculated in one specific embodiment, a detailed example is hereby set forth with reference to FIG. **2**B and FIG. **4**.

[0069] FIG. **2**B illustrates the household A **250** and the workplace W **210** of FIG. **2**A with an additional household B **280** and a cyber café C **290**. In household A **250** there are two digital device users: User X and User Y. User X regular uses laptop computer system **251** and cellular phone **252** such that the digital identities of laptop computer system **251** and cellular phone **252** should be identified as paired digital identities. User X's laptop computer system **251** is digitally identified using a cookie (C) on a web browser such that it is labelled C_X and user X's cellular phone **252** is digitally identified with a deviceID (D) such that it is labelled D_X. This digital device labelling nomenclature will be used with the other laptop and cellular phones as well. User Y of household A **250** regularly uses laptop computer system C_Y **261** and cellular phone D_Y **262** such that C_Y **261** and D_Y **262** should also be identified as a related pair of digital identities. Recall that all of the digital devices (C_X **251**, D_X **252**, C_Y **261**, and D_Y **262**) in household A **250** will use the same IP address A **263** that is assigned to Wi-Fi router **265** when those digital devices are used at household A **250**. Note that although this embodiment is based upon using web browser cookies and mobile device identifiers, any other similar identifiers that can be associated with the digital devices may be used.

[0070] User Z resides at household B **280** regularly uses laptop computer system C_Z **281** and cellular phone D_Z **282**. While at household B **280**, both C_Z **281** and D_Z **282** will use IP address B **283** that is assigned to Wi-Fi router **285** in use at household B **280**. Both user X and user Z work together at workplace W **210** such that C_X **251**, D_X **252**, C_Z **281**, and D_Z **282** are regularly used at workplace W **210**. While at workplace W **210** those digital devices will all use IP address W **223** that is assigned to firewall/proxy **221** at workplace W **210**. Many other digital devices (**211**, **212**, **213**, **214**, **215**, and **216**) will also use IP address W **223** at workplace W **210**.

[0071] Finally, FIG. **2**B also illustrates a cyber café **290** that offers free Wi-Fi service to customers of cyber café **290**. User Z and user Y frequent cyber café C **290** such that C_Z **281**, D_Z **282**, and D_Y **262** are illustrated at cyber café **290** where IP address C **293** is used on Wi-Fi router **295**. Note that many other visitors (not shown) will also frequent cyber café **290**. However, the various digital devices only seen together at cyber café **290** will not be considered potential digital identity pairs since there are too many digital identity pairings seen together at cyber café **290**.

[0072] After collecting internet usage data (as set forth in stage **310**), the next step in identifying digital identity pairs is to select a set of potential digital identity pairs as set forth in stage **320** of FIG. **3**. As set forth in the previous section, the various combinations of digital identities that are associated with IP addresses having six or less digital identities may be selected as potential digital identity pairings. In the example of FIG. **2**B, the various combinations of the digital devices at households A **250** and B **280** are therefore considered potential digital identity pairs. (The workplace W **210** and cyber café C **290** have too many different digital identities associated with them and thus do not provide good candidates for digital identity pairs.) For simplicity, only the possible digital identity pairings (C_X, D_X), (C_X, D_Y), (C_Y, D_X), and (C_Y, D_Y) from household A **210** will be analyzed in this example.

[0073] After identifying a set of potential digital identity pairs, the digital identity pairing system then calculates association scores for all of the potential digital identity pairs as set forth in stage **340** of FIG. **3**. In one embodiment, the pairing system attempts to pair each digital device with only one web browser cookie. (However, more than one different digital device may be paired with the same browser cookie.) Thus, among two competing cookies attempting to be paired with the same digital device, the digital identity pairing with the higher association score may be selected. For example, between potential pairings (C_X, D_X) and (C_Y, D_X), the pair with the higher association score may be selected as the most likely digital identity pair. Similarly, between potential digital identity pairings (C_X, D_Y) and (C_Y, D_Y) the pair with the higher association score may be selected.

[0074] In one particular embodiment, the digital identity pairing system uses a variation of Bayesian probability analysis to calculate an association score for each of the potential cookie and deviceID digital identity pairs. In addition a "support" score and "confidence" score may also be calculated. The support, confidence, and association scores may be defined as follows:

$$\text{Support}=P(\text{cookie, deviceID})$$

$$\text{Confidence}=P(\text{cookie}|\text{deviceID})$$

$$\text{Association}(\text{cookie}\rightarrow\text{deviceID})=P(\text{cookie}|\text{deviceID})/P(\text{cookie})$$

[0075] These three scores may be used to identify digital identity pairings and to rate the confidence in a digital identity pairing that has been made. The support score gives an indication of how much data support there is for the analysis of this particular cookie and deviceID pair. The confidence score gives an indication of how much confidence there is in the association score. The association score provide rating of how closely the cookie and deviceID are associated.

[0076] In the present disclosure, the support, confidence, and association scores are calculated using the set of internet usage observations on the various digital identities being considered that were collected in stage **310** of FIG. **3**. The following equations describe how the internet usage observations may be used to calculate the support, confidence, and association scores.

$$\text{co-occurrences}(\text{cookie, deviceID})=\text{number of times both cookie and deviceID at the same location (same source identifier IP address).}$$

$$P(\text{cookie, deviceID})=\text{co-occurrences}(\text{cookie, deviceID})/\text{total sample size}$$

$$P(\text{cookie}|\text{deviceID})=\text{co-occurrences}(\text{cookie, deviceID})/\text{occurrences}(\text{deviceID})$$

$$P(\text{cookie})=\text{number of occurrence (cookie)/total sample size}$$

[0077] To best illustrate the manner in which support, confidence, and association scores may be calculated in one embodiment, the association scores for some potential digital identity pairings in FIG. **2**B will be calculated using a set of digital identity internet usage observations depicted in FIG. **4**. The internet usage observation samples of FIG. **4** are a set of observations of the various digital identities (C_X, D_X, C_Y, D_Y, C_Z, D_Z) at various different source locations (household A, household B, workplace W, and the cyber café C) that are all within a specified time window to keep the data fresh. These internet usage observations come from collected internet usage data triads (client identifier, common source/destination identifier, and timestamp) wherein the client identifiers are the PC browser cookies and cellular phone deviceIDs, the common source/destination identifiers are the IP address of the location where a client identifier was observed, and the timestamps are all within a defined window.

[0078] To illustrate how the digital identity pairing system works, an example is hereby presented wherein the digital identity pairing system attempts to pair cellular phone D_X **252** from household A **250** with one of the laptop computers C_X **251** or C_Y **261** used at household A **250**. To calculate an association score for D_X and C_Y the following information from the table in FIG. **4** is needed:

[0079] 1) C_Y was observed 2 times at home A

[0080] 2) D_X was observed 4 times at home A and 8 times at workplace W

[0081] 3) C_Y and D_X co-occurred 2 times at home A

[0082] 4) There are 90 total observations

[0083] The above observations are then used to calculate Association($C_Y\rightarrow D_X$) as follows:

$$P(C_Y)=\#\text{ of occurrence }(C_Y)/\text{total sample size}=(2+7)/97=11/97$$

$$P(C_Y|D_X)=\text{co-occurrences}(C_Y, D_X)/\text{occurrences}(D_X)=2/(4+8)=1/6$$

$$\text{Association}(C_Y\rightarrow D_X)=P(C_Y|D_X)/P(C_Y)=(1/6)/(11/97)=1.47$$

[0084] The other potential pairing for cellular phone D_X **252** is the pairing of cellular phone D_X **252** with laptop computers C_X **251**. To calculate an association score for the pair of C_X and D_X the following information from the table in FIG. **4** is needed:

[0085] 1) C_X was observed 3 times at home A and 10 times at workplace W

[0086] 2) D_X was observed 4 times at home A and 8 times at workplace W

[0087] 3) C_X and D_X co-occurred 3 times at home A and 8 times at workplace W

[0088] 4) There are 97 total observations

[0089] The above observations are then used to calculate Association($C_X\rightarrow D_X$) as follows:

$$P(C_X)=\#\text{ of occurrence }(C_X)/\text{total sample size}=(3+10)/90=13/97$$

$$P(C_X|D_X)=\text{co-occurrences}(C_X, D_X)/\text{occurrences}(D_X)=(3+8)/(4+8)=11/12$$

Association($C_X \rightarrow D_X$)=$P(C_X|D_X)/P(C_X)$=(11/12)/(13/97)=6.34

[0090] When comparing the two association scores, the higher association score is selected. In this case, the Association($C_X \rightarrow D_X$)=6.34 score is much higher than the Association($C_Y \rightarrow D_X$)=1.47 score such that the pairing of laptop computer system C_X **251** and cellular phone D_X **252** are deemed to be a high-probability digital identity pair. The support and confidence scores for this pairing are as follows:

Support=$P(C_X, D_X)$=co-occurrences(C_X, D_X)/total sample size

Support=(3+8)/97=11/97

Confidence=$P(C_X|D_X)$=co-occurrences(C_X, D_X)/occurrences(D_X)

Confidence=(3+8)/(4+8)=11/12

[0091] The Support metric may be used to filter out observations that do not have enough statistical significance. In one embodiment the minimum Support metric is calibrated based on a desired Precision/Recall measurement score. The Confidence metric is a value calculated as part of the Association score.

[0092] In household B **280**, the only possible digital identity pairing is of laptop computer system C_Z **281** and cellular phone D_Z **282**. To calculate an association score for C_Z and D_Z the following information from the table in FIG. **4** is needed:

[0093] 1) C_Z observations: 5 at home B, 20 at workplace W, and 4 at café C.

[0094] 2) D_Z observations: 8 at home B, 10 at workplace W, and 4 at café C.

[0095] 3) C_Z & D_Z co-occurrence: 5 at home B, 10 at workplace W, and 4 at café.

[0096] 4) There are 97 total observations

[0097] The above observations are then used to calculate Association($C_Z \rightarrow D_Z$) as follows:

$P(C_Z)$=# of occurrence (C_X)/total sample size=(5+20+4)/97=29/97

$$P(C_z \mid D_z) = \text{co-occurrences}(C_z, D_z) / \text{occurrences}(D_z)$$
$$= (5 + 10 + 4) / (8 + 10 + 4)$$
$$= 19 / 22$$

Association($C_Z \rightarrow D_Z$)=$P(C_Z|D_Z)/P(C_Z)$=(19/22)/(29/97)=2.89

Support=$P(C_Z, D_Z)$=co-occurrences(C_Z, D_Z)/total sample size

Support=(5+10+4)/97=19/97

Confidence=$P(C_Z|D_Z)$=co-occurrences(C_Z, D_Z)/occurrences(D_Z)

Confidence=(5+10+4)/(8+10+4)=19/22

[0098] Since C_Z and D_Z is the only possible digital identity pairing of digital identities at household B **280** there are no other association scores to directly compare it against. Thus, some threshold value may be used to determine whether that association score is large enough to determine that the two devices should be paired. The 19/97 support score and 19/22 confidence score may also be used to help determine if C_Z and D_Z will be considered as a digital identity pair.

[0099] Post-Processing Association Scores

[0100] The association scores calculated for a potential digital identity pair will vary over time depending on the specific digital identity usage observations that are being used. To eliminate this 'noise' in the data, various techniques may be used to smooth out the data and provide more consistent results. Thus, as set forth in stage **360** of FIG. **3**, a digital identity pairing system may post-process association scores to improve the overall accuracy of the system.

[0101] One simple means of post-processing association scores to improve the results is to discard association scores that fall below a particular designated threshold level. Association scores that fall below a designated threshold level may simply be irrelevant noise.

[0102] FIG. **5**A graphically illustrates an example of a series of daily association scores for a particular potential digital identity pair calculated over time. Each association score is actually calculated using a moving window of recent digital identity usage observations as conceptually illustrated in the lower left of the graph. In the example of FIG. **5**A, each association score is calculated using digital identity usage observations from the last five days. Using a moving window filters out some of variability in the data but the data is still somewhat variable as illustrated in FIG. **5**A. Note that a five-day moving window is just an arbitrary selection and the digital identity pairing system may user other time scales. Selection of the data window size will be made by measuring the precision and recall of different sizes and selecting the window size that provides the best result.

[0103] To further reduce the noise in the association score data, a digital identity pairing system may post-process association scores to eliminate some of the outlying data samples. For example, if a person goes on a vacation then that person's digital device usage pattern may vary dramatically and thus eliminate a digital identity pair that was discovered. Proper post-processing of the association scores may prevent a temporary change in usage patterns from eliminating an accurately made digital identity pairing.

[0104] In one embodiment, the digital identity pairing system collects a number of association scores and calculates a statistical mode of the association scores that have been gathered over time. The statistical modes of different association scores are then compared against each other to determine the high-probability pairs (instead of comparing recently calculated association scores directly). Using statistical modes may effectively reduce the noise in the sampled digital identity usage data.

[0105] One method of calculating a statistical mode involves first creating a set of different association score range buckets. The width of the association score range buckets will vary depending on the data density. Then, the collected set of association scores (such as the scores from FIG. **5**A) are divided into the various association scores buckets thereby creating a count of the number of association scores that fall within each association score range bucket. An example of this is conceptually illustrated in FIG. **5**B. The association score range bucket with the greatest number of association scores in that bucket is deemed to be the statistical mode in one embodiment and may be used to compare against other association score statistical modes calculated in the same manner.

[0106] Various other post-processing methods may also be used to smooth out the association scores. For example, instead of using a statistical mode, other implementations may use a median, a mean, or another method of smoothing out the association scores.

[0107] Boolean Association Score System

[0108] As set forth in the previous sections, one embodiment of digital identity pairing system operates by counting the frequency of observations of a digital identity (such as a cookie or a device identifier) at a particular source/destination address (such as an IP address). The internet usage frequency counts are then used to identify a set of potential digital identity pairs that are then processed to determine a set of association scores for each of the potential digital identity pairs.

[0109] The internet usage frequency data can be heavily biased by the number of times that a digital identity is linked to a particular source/destination identifier. For example, a user that spends a very large amount of time using a laptop computer system at work but only rarely uses that laptop computer system while at home will have results that are heavily biased by the large number of data samples collected from the user's work location. To reduce this heavy biasing that may occur; some embodiments employ a Boolean based association score system that does not operate by counting the number of times (frequency) a digital identity is observed at a source/destination in a specified time period. Instead, the Boolean association score system only counts whether a particular digital identity was observed or not at a source/destination identifier during the relevant time period.

[0110] FIG. **6** illustrates the observed usage data from FIG. **4** but in Boolean form instead of frequency count form. Thus, if a particular digital identity was observed at a particular location during a specified time period then a '1' is assigned and when there is no observation a "0" is assigned. Once the digital identity observation data has been converted to Boolean form, the Boolean observation data can be processed into Boolean-based association scores using slightly modified techniques. One important difference is that instead of using the total number of observations as a sample size, the number of possible source/destination location identifiers is considered as the sample size. Therefore, the following equation has changed:

$$P(\text{cookie})=\text{number of occurrences(cookie)/possible locations}$$

[0111] To fully explain the Boolean association score calculation methodology, as example is presented with reference to the Boolean usage data disclosed in FIG. **6**. Specifically, to calculate a Boolean system association score of C_X and D_Y the following information from the table in FIG. **6** is used:

[0112] 1) C_X was observed at home A and at workplace W

[0113] 2) D_Y was observed at home A and at the café.

[0114] 3) C_X and D_Y co-occurred at home A

[0115] 4) There are 4 places where observations were made.

[0116] The above observations are then used to calculate Association($C_X \rightarrow D_Y$) as follows:

$$P(C_X)=\#\text{ of occurrence }(C_X)/\text{total sample size}=(1+1)/4=1/2$$

$$P(C_X|D_Y)=\text{co-occurrences}(C_X, D_Y)/\text{occurrences}(D_Y)=1/2$$

$$\text{Association}(C_X \rightarrow D_Y)=P(C_X|D_Y)/P(C_X)=(1/2)/(1/2)=1$$

[0117] The other potential pairing for C_X is the pairing of C_X and D_X. To calculate a Boolean system association score of C_X and D_X the following information from the table in FIG. **6** is needed:

[0118] 1) C_X was observed at home A and at workplace W

[0119] 2) D_X was observed at home A and at workplace W

[0120] 3) C_X and D_X co-occurred at home A and at workplace W

[0121] 4) There are 4 places where observations were made.

[0122] The above observations are then used to calculate Association($C_X \rightarrow D_X$) as follows:

$$P(C_X)=\#\text{ of occurrence }(C_X)/\text{total sample size}=(1+1)/4=1/2$$

$$P(C_X|D_X)=\text{co-occurrences}(C_X, D_X)/\text{occurrences}(D_X)=(1+1)/(2)=1$$

$$\text{Association}(C_X \rightarrow D_X)=P(C_X|D_X)/P(C_X)=(1)/(1/2)=2$$

[0123] As with the frequency count based system, when comparing the two Boolean system association scores, the higher association score may be selected. In this case, the Association($C_X \rightarrow D_X$)=2 score is higher than the Association ($C_X \rightarrow D_Y$)=1 score such that the pairing of laptop computer system (C_X) **251** and cellular phone (D_X) **252** are deemed to be a high-probability digital identity pair (not the pairing C_X and D_Y).

[0124] Combined Frequency and Boolean Counting Score

[0125] Both the frequency counting system and the Boolean counting system have their own advantages. To benefit from the advantages in both approaches, some embodiments of the digital pairing system employ a weighted average of the frequency counted association score and the Boolean counted association score. FIG. **7** illustrates a flow diagram describing how such a two score based digital identity pairing system may operate. Note that although FIG. **7** and this section describes a system that uses two different score, other embodiments may consider 3, 4, or more different types of association scores.

[0126] Initially, at stage **710**, the digital identity internet usage data is collected in the same manner as previously described. Next, at stage **720**, the digital identity internet usage data is analyzed to select a set of potential digital identity pairings that will further be analyzed. After this point, the internet usage data is then analyzed with the two different scoring systems: a frequency counting system and a Boolean counting system.

[0127] Along a first path on the left side of FIG. **7**, the digital identity internet usage data is analyzed with a frequency counting based scoring system. At stage **740**, the internet usage data is processed with frequency counting to generate a first set of association scores. One possible method of implementing a frequency counting system for calculating association scores is described in the earlier "A Basic Digital Identity Pairing Method" section of this document. The frequency count based association scores may then be post-processed at stage **760** to reduce the noise in the data. The association score post-processing may performed as set forth in the earlier "Post-Processing Association Scores" section of this document.

[0128] Along the second path on the right side of FIG. **7** after stage **720**, the digital identity internet usage data is analyzed with Boolean counting. At stage **745**, the internet usage data is processed with Boolean counting to calculate

association scores as described in the "Boolean Counting System" section of this document. Next, at stage **765**, the Boolean count association scores may then be processed with the same post-processing techniques used to processing the frequency count association scores as set forth in the earlier "Post-Processing Association Scores" section of this document.

[0129] At stage **770**, the processed association scores from the two different association scoring systems are then combined. Various different methods of combining the two different association scores may be used. In one embodiment, the frequency counting association scores and the Boolean counting association scores are combined in a weighted manner with the following basic equation:

$$Score_{comb}=\alpha Score_{Freq}+\beta Score_{Bool}$$

[0130] Various different methods may be used to determine the best α and β weighting factors that are used to combine the two different association scores. In one embodiment, class validation was used to calculate the best α and β weighting factors. Specifically, digital identity usage data was collected for several pairs of digital devices where each pair of devices was known to belong to a single user. Since those pairs of digital devices were known to be actually associated with a single user, the values of the α and β weighting factors were selected to maximize the combined association scores those known paired devices. Again, it must be emphasized that additional association scores may also be considered such that a weighted score may be created from several different association scores.

[0131] In some embodiments linear regression analysis may be used to determine how to combine the two association scores. Specifically, with both frequency counting association scores and the Boolean counting association scores there are two different predictors. Thus, using a set of known accurate digital identity pairings, linear regression may be used to determine how to combine the two different predictors in a manner that provides accurate results.

[0132] After creating a combined association score at stage **770**, the digital pairing system may use the combined association score to select high-probability digital identity pairs at stage **780**. The high-probability digital identity pairs may be selected from competing potential digital identity pairs. Finally, at stage **790**, the combined profile information from the two digital identities in a digital identity pair may be used to accurately select targeted advertisements for both digital identities in the digital identity pair.

[0133] Destination Identifiers

[0134] In the preceding examples, the common source/destination identifier used was a source IP address that a digital device was using to access the internet. However, as the name implies the common source/destination identifier may identify a destination that is accessed by a particular client device. An example of the destination identifier is presented with reference to FIGS. **2**B and **8**.

[0135] As illustrated in FIG. **2**B household A **250** has four different digital devices (C_X**251**, D_X**252**, C_Y**261**, and D_Y**262**) that access the internet **201**. Since there are only four digital devices that access the internet **201** from household A **250**, various combinations of these four different digital devices may be considered potential digital identity pairings. So a set of digital device usage data associated with destinations visited by these four different digital devices may be used to determine high-probability digital identity pairings.

[0136] FIG. **8** illustrates a set of destination based usage observations for the four different digital devices (C_X**251**, D_X**252**, C_Y**261**, and D_Y**262**). Specifically, FIG. **8** lists the counts of visits to the web sites ESPN.com, CNN.com, Finance.com, and Fark.com for the four different digital devices (C_X **251**, D_X**252**, C_Y**261**, and D_Y**262**). The observed visits to each web site from each of the four different digital devices all occurring within a defined time period such that all the observations are current.

[0137] The same techniques disclosed in the previous sections can be used to calculate association scores for the destination addresses. Once again, the pairing of laptop computer C_X **251** to one of the two cellular phones D_X **252** or D_Y **262** will be performed. To calculate an association score of C_X and D_Y the following information from the table in FIG. **4** is needed:

[0138] 1) C_X visited ESPN.com 5 times, Finance.com 2 times, fark.com 4 times

[0139] 2) D_Y visited CNN.com 9 times and Finance.com 5 times

[0140] 3) C_X and D_Y co-occurred 2 times at Finance.com

[0141] 4) There are 56 total observations

[0142] The above observations are then used to calculate Association($C_X \rightarrow D_Y$) as follows:

$$P(C_X)=\#\text{ of occurrence }(C_X)/\text{total sample size}=(5+2+4))/56=11/56$$

$$P(C_X|D_Y)=\text{co-occurrences}(C_X, D_Y)/\text{occurrences}(D_Y)=2/(9+5)=1/7$$

$$\text{Association}(C_X \rightarrow D_Y)=P(C_X|D_Y)/P(C_X)=(1/7)/(11/56)=0.73$$

[0143] The other potential digital identity pairing for laptop computer C_X **251** is the pairing of C_X and D_X. To calculate an association score for the pair of C_X and D_X the following information from the table in FIG. **4** is needed:

[0144] 1) C_X visited ESPN.com 5 times, Finance.com 2 times, fark.com 4 times

[0145] 2) D_X visited ESPN.com 6 times, CNN.com 2 times, fark.com 6 times

[0146] 3) C_X and D_X co-occurred 5 times at ESPN.com and 4 times at fark.com

[0147] 4) There are 56 total observations

[0148] The above observations are then used to calculate Association($C_X \rightarrow D_X$) as follows:

$$P(C_X)=\#\text{ of occurrence }(C_X)/\text{total sample size}=(5+2+4))/56=11/56$$

$$P(C_X|D_X)=\text{co-occurrences}(C_X D_X)/\text{occurrences}(D_X)=(5+4)/(16)=9/14$$

$$\text{Association}(C_X \rightarrow D_X)=P(C_X|D_X)/P(C_X)=(9/14)/(11/56)=3.27$$

[0149] Again, the higher association score may be selected among competing digital identity pairings. In this case, the Association($C_X \rightarrow D_X$)=3.27 score is much higher than the Association($C_X \rightarrow D_Y$)=0.73 score such that the pairing of laptop computer system C_X **251** and cellular phone D_X **252** are deemed to be a high-probability digital identity pair. Note that the Boolean Association score may be calculated in the same manner with regard to destination identifiers.

[0150] Matching Accuracy Measurement System

[0151] User destination information can also be used to help measure the accuracy of digital identity pairings made with common source identifiers. As set forth in earlier sec-

people often do not use their cellular phones for much web browsing due to the small display screen and limited user interface. Thus, there may not be enough web browsing history for a useful user entropy calculation.

[0166] Instead of analyzing the web browsing history on a cellular phone, a system may instead analyze the usage of application programs on the cellular smartphones. Apple iOS, Android, RIM Blackberry, and other cellular smartphone and tablet systems have thousands of small application programs for accomplishing a wide variety of tasks. These different application programs can be categorized based on what features and tools the application programs provide. However, the usage of smartphone and table application programs cannot be directly compared to web browsing histories. To handle this, a large user history of web browsing and smartphone/tablet application program usage has been analyzed. There is a correlation between smartphone/table application program usage and web browsing patterns such that one can infer a likely web browsing pattern from a known smartphone/tablet application program usage history. The inferred web browsing pattern can then be used in user entropy comparisons.

[0167] For example, the user of cellular phone D_X **952** is a light mobile application user that has the following application program usage pattern: 30% on casual games, 30% on life style, 20% on news, 10% on finance, and 10% on other application programs. The user of cellular phone D_Y **962** has the following application program usage pattern: 50% on action games, 20% on news, 20% on finance and 10% on other apps. With a large amount of historical information web browsing and application usage on known users, a system for inferring web browsing patterns application usage patterns. For example, from the application usage pattern of cellular phone D_X **952** an inference system may determine that such a user may allocate their web browsing as follows: 20% on fashion, 25% on news, 15% on finance websites, and 40% on others. In the same manner, the user of cellular phone D_Y **962** may web browse 25% on news, 20% on finance, 20% on games and 35% on others.

[0168] From these inferred web browsing patterns, the user entropy metrics may be calculated. The inferred user entropy values would be:

$$H_{DIX}=(0.20*\log 0.20+0.25*\log 0.25+0.15*\log 0.15+0.4*\log 0.4)=5.81$$

$$H_{DIY}=(0.25*\log 0.25+0.20*\log 0.20+0.20*\log 0.20+0.35*\log 0.35)=5.65$$

[0169] Assuming that laptop C_X **951** and desktop T_X **953** were correctly paired then one can compare the inferred web browsing history of cellular phone D_X **952** with the group laptop C_X **951** and desktop $T_{X\,1\,953}$.

$$H_{DIX,CX,TX}=(0.2875*\log 0.2875+0.2375*\log 0.2375+0.1375*\log 0.1375+0.3375*\log 0.3375)=5.75$$

[0170] The entropy gain may then be calculated as follows:

$$H_\Delta = H_{DIX,CX,TX} - 0.5*H_{DIX} - 0.5*H_{CX,TX} = 5.75 - (0.5*5.81) - (0.5*5.84) = -0.07$$

[0171] Such a relatively small entropy gain would provide evidence supporting that the devices are used by the same user. If one attempts to match cellular phone D_X **952** with laptop C_Y **961** and desktop T_Y **963** then the entropy gain is 1.42 ($H_\Delta=H_{DIX,CY,TY}-0.5*H_{DIX}-0.5*H_{CY,TY}=1.42$). Note that the web browsing inference process may not always be accurate such that higher entropy gains may appear. Thus, inference data may be better used for helping select from a set of possible matches. For example, the entropy gain of matching cellular phone D_Y **962** with laptop C_X **951** and desktop T_X **953** is 2.64 whereas the entropy gain of matching cellular phone D_Y **962** with laptop C_Y **961** and desktop T_Y **963** is 1.46 such that it would be better to match cellular phone D_Y **962** with laptop C_Y **961** and desktop T_Y **963** instead of with laptop C_X **951** and desktop T_X **953**. Thus, even with an imperfect inferred website access pattern (which is quite different from an actual recorded website pattern) the system can still determine that cellular phone D_X **952** matches laptop C_X **951** and desktop T_X **953** better than cellular phone D_Y **962**; while cellular phone D_Y **962** matches laptop C_Y **961** and desktop T_Y **963** better than cellular phone D_X **952**.

[0172] Occasionally, this system may encounter two different users have similar website access patterns, and the entropy gain metric will erroneously consider digital identities from the two different users to be from the same user. However, even when this mismatch occurs this mismatch will not decrease the power of targeting advertisement because these two users have the same interests.

[0173] User Session Level Sampling

[0174] In the previous sections, the source/destination identifier was the main distinguishing factor used to identify high-probability pairs. However, the timestamp can also be used to help identify digital identity pairs. Specifically, different digital devices that are often used at the same location around the same time may have a higher probability of being related to the same user. Thus, some embodiments of a digital pairing system use the timestamps in the triads of observed usage data to help calculate association scores.

[0175] FIG. **10**A illustrates a log of usage triads **1010** wherein a client identifier of User**1** has been observed visiting a series of destination identifiers (URL**1** to URL**9**) wherein each observation has an associated timestamp (T**0** to T**9**). To determine if different digital identities are used around the same time, the concept of a user "session" may be used. A session is defined as a set usage observations that are never more than a session threshold amount of time between successive observations. If more than a session threshold amount of time passes between two consecutive digital identity observations then those two digital identity observations are deemed to be from two separate user sessions.

[0176] Various different methods may be used to calculate the session threshold amount of time. In one embodiment, an analysis of time gaps between digital identity usage observations is performed to calculate a session threshold amount of time. Referring to FIG. **10**A, the amount of time between each digital identity observation timestamp (T**0** to T**9**) is calculated as a gap time (g**1** to g**9**) **1050**. Once a large collection of gap times has been calculated, these gap times are then analyzed with standard normal distribution statistics analysis. Initially, all of the observed gap times are placed into gap time buckets as illustrated in FIG. **11**. Then, the mean (μ) and standard deviation (σ) for the gap times are calculated as illustrated in FIG. **11**. Anything that is less than 2 standard deviations (σ) from the mean is considered to be within the same session. Thus, all of the gap times **1110** are deemed to be within the same user session. Note that this is just one possible embodiment and that many other methods may be used to determine

what constitutes a user session. User sessions may be determined using other criteria such that user session time periods may vary in length.

[0177] After determining a method of splitting observations into temporally distinct sessions, the session data may be processed with the same association score methodology set forth in the previous sections. For example, FIG. **10**B illustrates a set of sessions for the laptop computers C_X **251** and C_Y **261** in household A **250** of FIG. **2**B. For each session, the usage observations of the four different digital devices (C_X **251**, D_X **252**, C_Y **261**, and D_Y **262**) are listed. Since the different users tend to use their laptop computers and cellular phones at different times, the two digital pairings of the two different devices may be distinguished despite only having observations from a single location (household A **250**).

[0178] The techniques disclosed in the previous sections may be used to calculate association scores for the four different sessions depicted in FIG. **10**B. Once again, the pairing of laptop computer C_X **251** to one of the two cellular phones D_X **252** or D_Y **262** in household A **250** will be performed. To calculate an association score of C_X and D_Y the following information from the table in FIG. **10**B is needed:

[0179] 1) C_X observations: 4 in session 1, 1 in session 3, 14 in session 4

[0180] 2) D_Y observations: 12 in session 2, 7 in session 3, 1 in session 4

[0181] 3) C_X and D_Y co-occurred once in session3 and once in session 4

[0182] 4) There are 74 total observations

[0183] The above internet usage observations are then used to calculate Association($C_X \rightarrow D_Y$) as follows:

$P(C_X)$=# of occurrence (C_X)/total sample size=(4+1+14))/74=19/74

$P(C_X|D_Y)$=co-occurrences(C_X, D_Y)/occurrences(D_Y)=2/(12+7+1)=1/10

Association($C_X \rightarrow D_Y$)=$P(C_X|D_Y)/P(C_X)$=(1/10)/ (19/74)=0.39

[0184] The other potential digital identity pairing for laptop computer C_X **251** is the pairing of C_X and D_X. To calculate an association score for the pair of C_X and D_X the following information from the table in FIG. **10**B is needed:

[0185] 1) C_X observations: 4 in session 1, 1 in session 3, 14 in session 4

[0186] 2) D_X observations: 2 in session 1, 9 in session 4

[0187] 3) C_X and D_X co-occurred twice in session 1 and 9 times in session 4

[0188] 4) There are 74 total observations

[0189] The above observations are then used to calculate Association($C_X \rightarrow D_X$) as follows:

$P(C_X)$=# of occurrence (C_X)/total sample size=(5+2+4))/56=11/56

$P(C_X|D_X)$=co-occurrences($C_X D_X$)/occurrences(D_X)=(2+9)/(11)=1

Association($C_X \rightarrow D_X$)=$P(C_X|D_X)/P(C_X)$=1/(11/56)=5.09

[0190] With data depicted in FIG. **10**B, the Association ($C_X \rightarrow D_X$)=5.09 score is much higher than the Association ($C_X \rightarrow D_Y$)=0.39 score such that the pairing of laptop computer system C_X **251** and cellular phone D_X **252** are deemed to be a high-probability digital identity pair. Note that the Boolean Association score may be calculated in the same manner for these session-generated association scores.

[0191] Digital Identity Chaining

[0192] In the preceding sections of this document various techniques have been disclosed for identifying digital identities that are likely to belong to the same human user. When such digital identity pairs are determined, the advertising to both digital identities can be synergistically improved by combining digital profile information collected on both digital identities. However, the process does not have to end with simple pairings of digital identities. The technique can be extended to combine multiple digital identities together thereby further improving the accuracy of advertising targeting. The technique of combining together the digital profiles collected from multiple digital identities may be referred to as 'digital identity chaining'.

[0193] Referring back to FIG. **2**A, household A **250** included laptop computers **251** and **261**, cellular telephones **252** and **262**, tablet computer system **257**, and video game console **259**. All of these digital electronic devices are coupled to the internet **201** and capable of receiving and displaying internet advertisements. In the preceding sections of this document, techniques disclosed how the usage patterns of the laptop computers **251** and **261** and the cellular telephones **252** and **262** could be used to identify laptop computer **251** and cellular telephone **252** as belonging to a single user X. User X may proceed to install an application onto cellular telephone **252** that is linked to a user account on video game console **259**. Thus, user X may self-identify cellular telephone **252** and video game console **259** as a pair of linked identities. Since cellular telephone **252** was already linked to laptop computer **251** that means that laptop computer **251** is linked to video game console **259** by the transitive property such that digital profile information from laptop computer **251** may be used to selected targeted advertisements to video game console **259**.

[0194] Digital identity chaining can synergistically improve the targeting of internet advertisements. For example, the accumulated information collected from laptop computer **251**, cellular telephone **252**, and video game console **259** may be used to create a very detailed digital profile of user X. This detailed digital profile of user X may then be used whenever an internet advertisement must be selected for laptop computer **251**, cellular telephone **252**, or video game console **259**.

[0195] Note that in the above example, the cellular telephone **252** and video game console **259** were linked together using a specific application program installed onto cellular telephone **252**. However, the two devices could have been pair together using the association score systems described in earlier sections. Thus, digital identity chaining may use many different methods of linking together different digital devices.

[0196] The preceding technical disclosure is intended to be illustrative, and not restrictive. For example, the above-described embodiments (or one or more aspects thereof) may be used in combination with each other. Other embodiments will be apparent to those of skill in the art upon reviewing the above description. The scope of the claims should, therefore, be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. In the appended claims, the terms "including" and "in which" are used as the plain-English equivalents of the respective terms "comprising" and "wherein." Also, in the

following claims, the terms "including" and "comprising" are open-ended, that is, a system, device, article, or process that includes elements in addition to those listed after such a term in a claim is still deemed to fall within the scope of that claim. Moreover, in the following claims, the terms "first," "second," and "third," etc. are used merely as labels, and are not intended to impose numerical requirements on their objects.

[0197] The Abstract is provided to comply with 37 C.F.R. §1.72(b), which requires that it allow the reader to quickly ascertain the nature of the technical disclosure. The abstract is submitted with the understanding that it will not be used to interpret or limit the scope or meaning of the claims. Also, in the above Detailed Description, various features may be grouped together to streamline the disclosure. This should not be interpreted as intending that an unclaimed disclosed feature is essential to any claim. Rather, inventive subject matter may lie in less than all features of a particular disclosed embodiment. Thus, the following claims are hereby incorporated into the Detailed Description, with each claim standing on its own as a separate embodiment.

We claim:

1. A method of identifying linked digital identities, said method comprising:

collecting a set of digital identity observations, said digital identity observations comprising a digital identity, source/destination identifier, and a timestamp;

identifying a set of potential digital identity pairings from said digital identity observations;

calculating association scores for said potential digital identity pairings using said digital identity observations; and

selecting potential digital identity pairings having the most favorable association scores as high probability digital identity pairs.

2. The method of identifying linked digital identities as set forth in claim **1** wherein identifying a set of potential digital identity pairings from said digital identity observations selecting permutations of digital identities associated with a single source/destination identifier and wherein fewer than a maximum threshold of digital identities are associated with said single source/destination identifier.

3. The method of identifying linked digital identities as set forth in claim **1** wherein identifying a set of potential digital identity pairings from said digital identity observations selecting permutations of digital identities associated with a residential source/destination identifier.

4. The method of identifying linked digital identities as set forth in claim **3** wherein said residential source/destination identifier comprises an IP address given to a residential customer.

5. The method of identifying linked digital identities as set forth in claim **1** wherein calculating association scores for said potential digital identity pairings are calculated with Bayesian inference.

6. The method of identifying linked digital identities as set forth in claim **5** wherein said Bayesian inference is based upon frequency counts of digital identities associated with source/destination identifier.

7. The method of identifying linked digital identities as set forth in claim **5** wherein said Bayesian inference is based upon Boolean associations between digital identities and source/destination identifiers.

8. The method of identifying linked digital identities as set forth in claim **1** wherein said digital identities comprise web browser cookies.

9. The method of identifying linked digital identities as set forth in claim **1** wherein said source/destination identifiers comprise Internet Protocol addresses.

10. The method of identifying linked digital identities as set forth in claim **1** wherein said selecting potential digital identity pairings having the most favorable association scores comprises selecting a digital identity pairing having a more favorable association score than competing digital identity pairings.

11. A machine-readable medium, said machine-readable medium storing a set of computer instructions for performing a set of actions comprising:

collecting a set of digital identity observations, said digital identity observations comprising a digital identity, source/destination identifier, and a timestamp;

identifying a set of potential digital identity pairings from said digital identity observations;

calculating association scores for said potential digital identity pairings using said digital identity observations; and

selecting potential digital identity pairings having the most favorable association scores as high probability digital identity pairs.

12. The machine-readable medium as set forth in claim **11** wherein identifying a set of potential digital identity pairings from said digital identity observations selecting permutations of digital identities associated with a single source/destination identifier and wherein fewer than a maximum threshold of digital identities are associated with said single source/destination identifier.

13. The machine-readable medium as set forth in claim **11** wherein identifying a set of potential digital identity pairings from said digital identity observations selecting permutations of digital identities associated with a residential source/destination identifier.

14. The machine-readable medium as set forth in claim **13** wherein said residential source/destination identifier comprises an IP address given to a residential customer.

15. The machine-readable medium as set forth in claim **11** wherein calculating association scores for said potential digital identity pairings are calculated with Bayesian inference.

16. The machine-readable medium as set forth in claim **15** wherein said Bayesian inference is based upon frequency counts of digital identities associated with source/destination identifier.

17. The machine-readable medium as set forth in claim **15** wherein said Bayesian inference is based upon Boolean associations between digital identities and source/destination identifiers.

18. The machine-readable medium as set forth in claim **11** wherein said digital identities comprise web browser cookies.

19. The machine-readable medium as set forth in claim **11** wherein said source/destination identifiers comprise Internet Protocol addresses.

20. The machine-readable medium as set forth in claim **11** wherein said selecting potential digital identity pairings having the most favorable association scores comprises selecting a digital identity pairing having a more favorable association score than competing digital identity pairings.

* * * * *