9. The PrivaMix Demonstration System, as used in Iowa

In 2007, Privacert implemented a version of the PrivaMix Protocol (Section 8.2) for a real-world experiment; we term this software the "PrivaMix Demonstration System" (or merely "System"). Because there are numerous variations and many ways to implement the PrivaMix Protocol, this section describes the details of the PrivaMix Demonstration System specifically. Section 10 explains its use in the real-world experiment.

In the PrivaMix Demonstration System, each participating machine runs special software devoted to this task. Shelter machines run one edition of the software program ("the Shelter Edition"). The Planning Office machine runs a different edition ("the CoC Edition"). These editions differ because the responsibilities of Shelters and the Planning Office in the PrivaMix protocol are different.

Appendix A contains a copy of the Software User's Guide for the PrivaMix Demonstration System. It itemizes menu options and shows copies of screen shots at various points in operation. Overall, the operation is extremely simple. If Shelters and the Planning Office use default settings, then operation is as simple as loading the Client information and clicking one button.

A description of the following characteristics further describe an implementation of PrivaMix. Each appears as a subsection below.

9.1 Hardware and network assumptions

The PrivaMix Demonstration System has minimal machine requirements, which means almost any computer system sold today is sufficient for use. However, the machine must have access to the Internet. Below is more information about these requirements.

9.1.1. Using the Internet

Machines participating in the PrivaMix Demonstration System communicate using the Internet Protocol on the Internet through traditional means of accessing the Internet. The software encrypts all communications between Shelters and the Planning Office.

The PrivaMix Demonstration System works with all traditional forms of Internet connections, even wireless broadband access. In wireless broadband access, a special card fits into the computer. The card then communicates directly with a wireless mobile phone network to send and receive information over the Internet. Wireless broadband is usually slower than dial-up, where a machine uses a phone line to communicate over the Internet, and is usually slower than Cable Internet, where a network cable connects directly to the computer. Just as the PrivaMix Demonstration System works with wireless broadband, it also works with dial-up, and cable connections. In fact, participants can use a mixture of Internet connection methods.

In general, communication in the PrivaMix Demonstration System consists of transmitting information bundles between a Shelter and its Planning Office. For example, Shelters send Client visit information to the Planning Office as an information bundle. The Planning Office sends UIDs and mixes to Shelters for mixing as an information bundle of one or more UIDs at a time. And, Shelters return mix results to the Planning Office as an information bundle. Communication is therefore more episodic in nature than continuous. All these communications use the Internet Protocol.

In the Internet Protocol, each machine on the Internet has its own unique number; this is termed the machine's "IP address." We assume that the IP address of a machine in the PrivaMix Demonstration System is unknown at software start. To learn the IP addresses of participating machines, each participating machines accesses a special program running on a previously known server which gathers and then reports relevant IP addresses among participants. As a result of connecting to the server, each participating Shelter learns the IP address of the Planning Office and the Planning Office learns the IP addresses of all participating Shelters. No other communication is made with the server unless a Shelter machine restarts during the PrivaMix Protocol and is assigned a different IP address.

9.1.2. Machine requirements

The PrivaMix Demonstration System has minimal machine requirements given the typical configuration of today's machines. As described above, the machine must have an Internet connection. It must have enough hard drive space to store the Client visit information. Standard machine configurations for memory and processing power are sufficient. The software works with popular operating systems (Windows, Linux, Mac OS).

9.1.3. Machine and network security

The PrivaMix Demonstration System assumes the machine is able to function properly without viruses or other impediments. No additional security requirements exist beyond the accepted security practices for maintaining Client information.

9.2 The PrivaMix function

The PrivaMix Demonstration System uses a strong one-way function. For details about the function and its proofs of correctness and compliance to the six requirements of a PrivaMix function, see [32]. Here are two key features. The function includes the modulus operator, so it is not easily reversed. The modulus operator is embedded within the function so that it maintains the commutative property across Shelters. See [32] for details and descriptions of this and other possible PrivaMix functions.

9.3 Selection and size of Shelter private values

The PrivaMix Demonstration System automatically selects a random 64-bit value as the Shelter's private value. The System selects the value after the Shelter provides Client source information. At that point, the System must generates UIDs, which requires its use. The System never reveals the Shelter's private value to the Shelter or the Planning Office. The value resides only in the Shelter's RAM memory. It is not stored or shared. If the machine has to restart during use, the System will select another value for the Shelter's private value and the network will start mixing again.

While the System uses 64 bit values, this is an internal setting. The PrivaMix Demonstration System supports 32, 64, 128, and 256 bit values.

9.4 Selection and size of Client source information

The PrivaMix Demonstration System does not prescribe which fields to use as Client source information. Given a set of fields identified for use, the System will compute a 64-bit UID for the Client. Just as with private values for Shelters, the PrivaMix Demonstration System uses 64 bit values for resulting UIDs. This is an internal setting. The PrivaMix Demonstration System also supports 32, 64, 128, and 256 bit values.

While the PrivaMix Demonstration System does not dictate which Client fields to use as source information, precautions are needed. Here are two important precautions.

(7) Care must be taken that sufficient variability exists in the fields so that resulting UIDs have a sufficiently wide range of possible values.

(8) Care must also be taken to make sure that different Clients are not likely to have to the same set of values appearing in the source information.

Further details about selecting Client source information to best work with a PrivaMix function appears in [32].

9.5 Transfer of Universal Data Elements

In the PrivaMix Demonstration System, Shelters transfer a comma-delimited text file to the Planning Office as encrypted content over an Internet connection. Each line contains a Client's visit information. The leftmost field on the line is the Client's UID. The remaining fields on the line are fields associated with the Client's visit to the Shelter, presumably the Universal Data Elements associated with that Client.. Below are more details about the file.

A comma-delimited text file is a simple text file that stores a table of information as follows. Each row of the table is a line in the text file. Columns in the table appear in order, left to right, with values separated by commas. The values themselves may be enclosed in quotation marks. Figure 68 shows an example of a comma-delimited text file. The original table shown in Figure 68(a) appears as comma-delimited file in Figure 68(b). A comma-delimited text file can be composed in a word processor (e.g., Microsoft Word), in a text editor (e.g., Notepad), or converted from a spreadsheet (e.g., Excel) or database program (e.g. Access).

In the PrivaMix Demonstration System, two versions of a comma-delimited text file exist. The Shelter provides an initial comma-delimited text file for processing; see Figure 68(b). This text file has the fields that comprise the Client's source information appearing as the leftmost fields. The fields that are to be sent to the Planning Office associated with that Client appear in the remaining fields. After a Shelter machine computes UIDs for each of its Clients using the Client source information, it produces a comma-delimited file replacing the leftmost fields with the Client UID; see Figure 68(c). Copies of UID information appear in temporary memory only. No copies appear on the hard drive.[21] Processing only involves the fields of the Client source information. The Planning Office receives the other fields with no processing or review. The Shelter machine merely forwards them "as is" to the Planning Office.

21 In the PrivaMix Demonstration System, copies of information containing UIDs appear only in the computer's memory (RAM). No copies appear on the machine's hard drive. However, if the size of the files warranted more storage than available in the computer's memory, encrypted copies could appear on the hard drive without posing security concern.

FirstName	DateOfBirth	YearOfBirth	Race	Gender	Veteran	Disability	Residence	Days	ZIP	EntryDate	ExitDate	ProviderID	GroupID	ProgramID
Dena	19650219	1965	Asian	Female	No	No	ownHome	730	32107	20060223	20060223	Shelter 185		TempShelter
Teresa	19580705	1958	White	Female	No	Yes	psychiatric	14	32109	20060401	20060425	Shelter 185		TempShelter
Roberta	19600115	1960	White	Female	No	No	friend	5	32108	20060401	20060425	Shelter 185		TempShelter
Britney	19500404	1950	White	Female	Yes	No	drugCtr	30	50123	20060401	20060425	Shelter 185		TempShelter
Christina	19690321	1969	White	Female	No	No	prison	90	32107	20060401	20060425	Shelter 185	16113	TempShelter
Arnold	20050505	2005	White	Male	No	No	fosterCare	90	32107	20060115	20060131	Shelter 185	16113	TempShelter

(a)

```
FirstName,DateOfBirth,YearOfBirth,Race,Gender,Veteran,Disability,Residence,Days,ZIP,EntryDate,ExitDate,ProviderID,GroupID,ProgramID
Dena,19650219,1965,Asian,Female,No,No,ownHome,730,32107,20060223,20060223,Shelter 185,,TempShelter
Teresa,19580705,1958,White,Female,No,Yes,psychiatric,14,32109,20060401,20060425,Shelter 185,,TempShelter
Roberta,19600115,1960,White,Female,No,No,friend,5,32108,20060401,20060425,Shelter 185,,TempShelter
Britney,19500404,1950,White,Female,Yes,No,drugCtr,30,50123,20060401,20060425,Shelter 185,,TempShelter
Christina,19690321,1969,White,Female,No,No,prison,90,32107,20060401,20060425,Shelter 185,16113,TempShelter
Arnold,20050505,2005,White,Male,No,No,fosterCare,90,32107,20060115,20060131,Shelter 185,16113,TempShelter
```

(b)

```
UID,YearOfBirth,Race,Gender,Veteran,Disability,Residence,Days,ZIP,EntryDate,ExitDate,ProviderID,GroupID,ProgramID
092n5jw09fu05j23450s5,1965,Asian,Female,No,No,ownHome,730,32107,20060223,20060223,Shelter 185,,TempShelter
mj0jt309usm5jd93kjskp6,1958,White,Female,No,Yes,psychiatric,14,32109,20060401,20060425,Shelter 185,,TempShelter
wopynps962kmnsi062mg,1960,White,Female,No,No,friend,5,32108,20060401,20060425,Shelter 185,,TempShelter
297sn0y92750276nklso2,1950,White,Female,Yes,No,drugCtr,30,50123,20060401,20060425,Shelter 185,,TempShelter
3908me5pwwntig85zxzz,1969,White,Female,No,No,prison,90,32107,20060401,20060425,Shelter 185,16113,TempShelter
87327qopwroinaptnlksfjh,2005,White,Male,No,No,fosterCare,90,32107,20060115,20060131,Shelter 185,16113,TempShelter
```

(c)

Figure 68. Comma-delimited text file having Client visit information. Original table (a) having 6 records and 14 fields appears as an equivalent comma-delimited file (b). The first line of the file includes the list of field names. The two leftmost fields, FirstName and DateOfBirth, are Client source information. The remaining fields provide Client visit information (see Figure 5 for descriptions). In the comma-delimited file in (c), the Client source information fields are replaced with a UID field. All other values remain the same. Dates appear as year, month, day (yyyymmdd). ProvideID is the Shelter's ID number. GroupID identifies Clients belonging to the same household. ProgramID identifies the kind of service provided.

9.6 UID validation

While Section 8.2.3 describes a variation of PrivaMix in which Shelters validate the number of values the Planning Office asks them to mix, the PrivaMix Demonstration System makes no such check. Shelter machines automatically mix values provided by the Planning Office without counting how many values that may be. This leaves a vulnerability: if the Planning Office pads the UIDs with known values, the Planning Office could learn Client source information (see Section 8.4.5). A simple remedy appears in Section 8.2.3, but in the interest of available resources, the PrivaMix Demonstration did not implement this variation.

9.7 De-duplication network

In the PrivaMix Demonstration System, the Planning Office orchestrates mixing as described in the generic PrivaMix Protocol (Section 8.2). The Planning Office sends values to each Shelter, one Shelter at a time, to mix, such that each Shelter mixes each UID once and each Shelter mixes all UIDs. Each Shelter only responds to mixing requests from the Planning Office's machine.

After mixing completes, the PrivaMix Demonstration System performs de-duplication on the Planning Office machine matching complete mixes across Shelter data. All values are held in the computer's memory. No information appears on the hard drive.

9.8 Post processing

Before making final de-duplicated results available to the Planning Office, the PrivaMix Demonstration System removes all UIDs, replacing them with numbers from 1 to the total number of distinct Clients and can do similar processing on PINs and Household IDs. The Planning Office does not receive a copy of the UIDs or complete mixes, only the results of de-duplication. Figure 68 shows the kinds of results made available to the Planning Office.

The PrivaMix Demonstration System automatically replaces UIDs with serialized numbers. An option exists by which other fields can be identified for serial renumbering. For example, in Figure 69, the Group ID, which identifies persons belonging to the same households, is serially renumbered.

CompleteMix	UID	YearOfBirth	Race	Gender	Veteran	Disability	Residence	Days	ZIP	EntryDate	ExitDate	ProviderID	GroupID	ProgramID
07ihweiy025j2r9u0a97a	092n5jw09fu05j23450s5	1965	Asian	Female	No	No	ownHome	730	32107	20060223	20060223	Shelter 185		TempShelter
jsdouf99kaiyaaitjbauqa1	mj0jt309usm5jd93kjskp6	1958	White	Female	No	Yes	psychiatric	14	32109	20060401	20060425	Shelter 185		TempShelter
hohho98y651890oabfa	wopynps962kmnsi062mg	1960	White	Female	No	No	friend	5	32108	20060401	20060425	Shelter 185		TempShelter
liyauoyathaeut295898y	297sn0y92750276nklso2	1950	White	Female	Yes	No	drugCtr	30	50123	20060401	20060425	Shelter 185		TempShelter
nflhtp094759hsgpohart	3908me5pwwntig85zxzz	1969	White	Female	No	No	prison	90	32107	20060401	20060425	Shelter 185	16113	TempShelter
skhiy906skjblsjgp25hpg	87327qopwroinaptnlksfjh	2005	White	Male	No	No	fosterCare	90	32107	20060115	20060131	Shelter 185	16113	TempShelter
rou7ou69079jojljpsouu	iihss949jpjp[ojzHUUY239	1969	Black	Female	No	No			32108	20060302	20060320	Center		Meal
jsdouf99kaiyaaitjbauqa1	mj0jt309usm5jd93kjskp6	1958	White	Female	No	Yes	rental	60	32109	20060317	20060331	Psychiatric		In-house
hohho98y651890oabfa	wopynps962kmnsi062mg	1960	White	Female	No	No			32108	20060401	20060425	Center		Meal
yojenisiuth2n596khslkh	9807hsdh0w850whsf022	1973	Black	Female	No	No			32108	20060302	20060320	Shelter 132		TempShelter
j6390rkhj6978970942	pjspdj092759700jpsjh09	1959	Black	Female	No	No			32108	20060401	20060405	Shelter 132		TempShelter
hohho98y651890oabfa	kjihg0sy090knkh2papiu96	1960	White	Female	No	No	TempShelter	26	32108	20060426	20060525	Shelter 132		TempShelter

(a)

UID	YearOfBirth	Race	Gender	Veteran	Disability	Residence	Days	ZIP	EntryDate	ExitDate	ProviderID	GroupID	ProgramID
1	1965	Asian	Female	No	No	ownHome	730	32107	20060223	20060223	Shelter 185		TempShelter
2	1958	White	Female	No	Yes	psychiatric	14	32109	20060401	20060425	Shelter 185		TempShelter
3	1960	White	Female	No	No	friend	5	32108	20060401	20060425	Shelter 185		TempShelter
4	1950	White	Female	Yes	No	drugCtr	30	50123	20060401	20060425	Shelter 185		TempShelter
5	1969	White	Female	No	No	prison	90	32107	20060401	20060425	Shelter 185	1	TempShelter
6	2005	White	Male	No	No	fosterCare	90	32107	20060115	20060131	Shelter 185	1	TempShelter
7	1969	Black	Female	No	No			32108	20060302	20060320	Center		Meal
2	1958	White	Female	No	Yes	rental	60	32109	20060317	20060331	Psychiatric		In-house
3	1960	White	Female	No	No			32108	20060401	20060425	Center		Meal
8	1973	Black	Female	No	No			32108	20060302	20060320	Shelter 132		TempShelter
9	1959	Black	Female	No	No			32108	20060401	20060405	Shelter 132		TempShelter
3	1960	White	Female	No	No	TempShelter	26	32108	20060426	20060525	Shelter 132		TempShelter

(b)

Figure 69. De-duplicated results. Information in the computer's memory after mixing (a). In the next step, the PrivaMix Demonstration System matches complete mixes to identify which clients are the same clients (rows 2 and 8; and, rows 3, 9, and 12). Planning Office receives a copy of de-duplicated results (b) with all UIDs replaced with numbering from 1 to the number of distinct Clients, repeating numbers to identify which records relate to the same Clients. Dates appear as year, month, day (yyyymmdd). GroupID identifies Clients belonging to the same household; these are also renumbered. Client 2 visited Shelter 185 and a psychiatric facility. Client 3 visited two Shelters and received meals.

9.9 Comparison to prior recommendations

Figure 70 below is a summary of the PrivaMix Demonstration System in terms of prior recommendations.

1	Outside scope.	Coordination of Systems across neighboring CoC's.
2	Outside scope.	Not share Shelter PIN beyond Shelter.
3	Implemented. (Section 9.8)	De-duplicated results should not include PINs, UIDs, or Household IDs.
4	Outside scope.	Shelters only include Clients who have left the Shelter.
5	Outside scope.	Train personnel on accepted practices for handling Client data.
6	Implemented. (Section 8.2)	UIDs should be inconsistently assigned across Shelters.
7	Outside scope.	Shelters should privacy notices for Client inspection.
8	Outside scope.	Fields date of birth and ZIP should be less specific.
9	Outside scope.	Planning Office should delete any fields in the Universal Data Elements not needed.
10	Outside scope.	Planning Office should sign Data Use Agreement with Shelters regarding linking.
11	Implemented. (Section 8.4)	Skilled person should certify System's risk of re-identification.
12	Implemented. (Section 8.4)	Skilled person should certify utility of de-duplicated results.
13	Implemented. (Section 8.4)	System using non-verifiable source information should instill trust.
14	Implemented. (Section 9.2)	System using encryption or hashing should use strong cryptographic methods.
15	Implemented. (Section 9.3)	System using encryption or hashing should control access to the function.
16	Outside scope.	System using scan cards/RFID should avoid issuing multiple cards to the same Client.
17	Implemented. (Section 9.8)	UIDs should be removed from de-duplicated results.
18	Outside scope.	Fields date of birth and ZIP must be less specific.
19	Implemented. (Section 8.4)	System must satisfy VAWA's requirements limiting re-idenification.
20	Implemented. (Section 9)	A PrivaMix System must avoid Shelters producing the same UID for Clients.
21	Addressed .(Section 9.1)	Computers transmitting UDE over a network must adhere to accepted security standards.
22	Not done. (Section 9.7)	If desirable, have a party other than the Planning Office orchestrate mixing.
23	Not done. (Section 9.8)	A PrivaMix System should anonymize or aggregate, rather than provide Client-level data.
24	Not done. (Section 9)	An economical PrivaMix System can result from using existing web browsers.
25	Implemented. (Section 9.2)	A PrivaMix Function must satisfy six noted requirements.
26	Implemented. (Section 9.3)	In a PrivaMix System. A Shelter value must be sufficiently large.
27	Implemented. (Section 9)	In a PrivaMix System, a Shelter should not even know its own private value.
28	Implemented. (Section 9.7)	In a PrivaMix System, unauthorized parties should be unable to use a Shelter's PrivaMix function.
29	Not done. (Section 9.6)	In a PrivaMix System, Shelters should validate the number of UIDs requested to mix.
30	Not done. (Section 9.8)	In order to provide collusion with an HMIS, provide only aggregate or anonymized results.
31	Outside scope.	Shelters only include Clients who have left the Shelter.
32	Implemented. (Section 9.8)	UIDs should be removed from de-duplicated results.
33	Implemented. (Section 9)	Claims must be assessed for any particular PrivaMix implementation.

Figure 70. Assessment of the PrivaMix Demonstration System in terms of prior recommendations made.