Consistent, Yet Anonymous, Web Access with LPWA

The Lucent Personalized Web Assistant offers a single, effective method for adopting differing personae.

The World-Wide Web has become an immensely popular and powerful medium in recent years. To attract more users, many Web sites offer personalized services, whereby users identify themselves and register their information preferences. On return visits, they conveniently receive a personalized selection of information. These personalized services, however, raise user concerns with respect to convenience and privacy.

Registration for these services lets information providers use a variety of tools to collect extensive profiles of users who visit their Web sites. Moreover, registration typically requires the user to specify a unique username and a secret password. Upon each return visit, the user must provide the same username and password. Sound security would dictate that users choose (and remember!) a different password for each site. An additional problem arises when naive users choose the same username and password for a Web site as they use for their own company's computers, thus potentially providing an intruder an easy way to break into the company's intranet.

Many sites ask for an email address at registration time to verify the user's registration and, often, to provide part of a personalized service, such as a newsletter or a personalized news digest. For example, the travel site expedia.com emails best fares for user-preferred airline routes. Moreover, an increasing number of Web sites send a confirmation code to the user's email address. This code must be provided in order to access the account. Hence, the user often must supply a valid email address to use the service at all. But the email address can also effectively serve as a (nearly) unique identifier for the user, and thus provides an avenue for profile aggregation across Web sites. Furthermore, a database of user email addresses can be easily abused to send out junk email (spam). To counter these concerns, wary users either avoid sites that require them to register, or they register with false information.

This article describes the Lucent Personalized Web Assistant (LPWA), a software system designed to address such user concerns. Users may browse the Web in a personalized, simple, private, and secure fashion using LPWA-generated aliases and other LPWA features. LPWA generates secure, consistent, and pseudonymous aliases (personae) for Web users. Each alias consists of an alias username, an alias password, and an alias email address. The alias email addresses

Eran Gabber, Phillip B. Gibbons, David M. Kristol, Yossi Matias, and Alain Mayer

allow Web sites to send messages to users and enable effective spam filtering. LPWA forwards mail addressed to the alias email address to the actual user. LPWA allows users to filter incoming messages based on the recipient address (the email alias), which is an effective method for detecting and blocking spam.

More specifically, the LPWA system supports the following features:

- *Automatic, secure, consistent, and pseudonymous generation of aliases.* Aliases present a different persona (username, password, email address) to each Web site. Personae for different Web sites, but belonging to the same user, appear to be independent and unrelated. (We will use "persona" and "alias" interchangeably through the remainder of this article.) The generated aliases are consistent, which means the user will present the same alias on return visits to the same Web site. They are pseudonymous in the sense that one cannot correlate between different aliases of the same user, nor between a user and her aliases.
- *Email service.* Web sites can use the email address of the supplied persona to send information to the user.
- *Anti-spamming support.* Users can filter junk email based on the recipient email address, which happens to be the persona email address. Furthermore, the user can infer which Web site is responsible for compromising her email address, even when the message is sent by a third party, or includes false headers.
- *Filtering of privacy-sensitive HTTP header fields.*
- *Indirection.* The TCP connection between the user and the Web site passes through a proxy, which thwarts tracking of the originating computer.
- *Statelessness.* LPWA does not keep any long-term state. In particular, it does not keep translation tables between users and their aliases. In this way, the LPWA site does not attract break-ins, and the LPWA service may be replicated easily. The absence of state information also implies there are no records to which an outside agency can demand access.

Overview of LPWA

The LPWA system consists of three functional components: *Persona Generator, Browsing Proxy,* and *Email Forwarder*. The Persona Generator generates a unique, consistent site-specific persona on demand by a user. It requires two pieces of identity information from the user: a User ID, that is, a valid Internet email address for the user; and a Secret serving as a universal password. Using these two pieces of information, plus the destination Web site address, the Persona Generator computes a persona for this Web site on the user's behalf. The Browsing Proxy increases the user's privacy by indirecting the connection on the TCP level and filtering headers on the HTTP level. The Email Forwarder forwards mail, addressed to a persona email address, to the corresponding user.

The Persona Generator consists of the Janus function designed to support pseudonymous client/server schemes. The Janus function is based on a suitable combination of cryptographic functions. (Its specification and implementation are detailed in [1].) Briefly, though, the Janus function takes as inputs a User ID, Secret, and Web site domain name and produces as output an LPWA username and password. (The LPWA alias email address is an encryption of the User ID by a fixed secret key.) The current implementation of LPWA replaces certain escape sequences in the user input by the appropriate component of the alias identity. (See the sidebar for more details.)

LPWA's functional components can potentially reside at various places. The Persona Generator can be implemented directly within the user's browser or on the Browsing Proxy that

system	connection anonymity	data anonymity	personalization
Anonymizer	low	high	n/a
Onion Routing	high	n/a	n/a
Crowds	high	n/a	n/a
P3P	n/a	medium	medium
LPWA	low	medium	high

Table 1. Capabilities of anonymizing systems

might reside on a firewall, an ISP access point, or a neutral site on the Internet. The Email Forwarder needs to reside away from the user's machine, since the goal is that various persona email addresses would be unlinkable to the user. Obviously, there are various trade-offs involved:

- *Trust.* The Persona Generator receives the user's real email address and a secret. The user opens a direct TCP connection to the Browsing Proxy. Depending on the design, the Email Forwarder must either store or forward the received messages reliably. Hence, all components must be trusted to various degrees.
- *Anonymity.* Neither the Browsing Proxy's nor the Email Forwarder's location should make it possible to infer a user's identity.
- *Performance.* If the location of the Browsing Proxy is too far away (in terms of Internet connections), then the performance degradation when browsing becomes noticeable to the user. This is an inherent problem of all HTTP proxies since all traffic to and from the user's browser is routed through the proxy.

The design of the public trial version of LPWA takes into consideration ease of deployment and restrictions on the distribution of software that contains cryptographic modules. We selected an implementation comprising two components. The first component is an HTTP proxy server, located on our premises in Murray Hill, N.J., that implements both the Browsing Proxy and the Persona Generator. This configuration is depicted in Figure 1. The second component is a remailer, located on the same machine as the proxy server, that implements the Email Forwarder.

In [1], we discuss in detail schemes with components residing on users' machines, ISP access points, or firewalls. Compared to our choice, such configurations have advantages in terms of trust and performance. On the other hand, our design choices allowed a fast deployment of a public trial version, showcasing our ideas and attracting thousands of users. (We have also implemented an internal trial version for users within the Lucent corporate firewall; this version plays the role of a firewall proxy.)

Privacy and convenience. The user's true identity is protected by LPWA, since aliases cannot be translated back to usernames. In addition, the user has different aliases for different Web sites, which prevents collusion of Web sites and creation of user profiles or dossiers based on common keys. However, the user should be careful not to provide additional information to Web sites, such as her mailing address or credit card numbers, which would reveal her true identity.

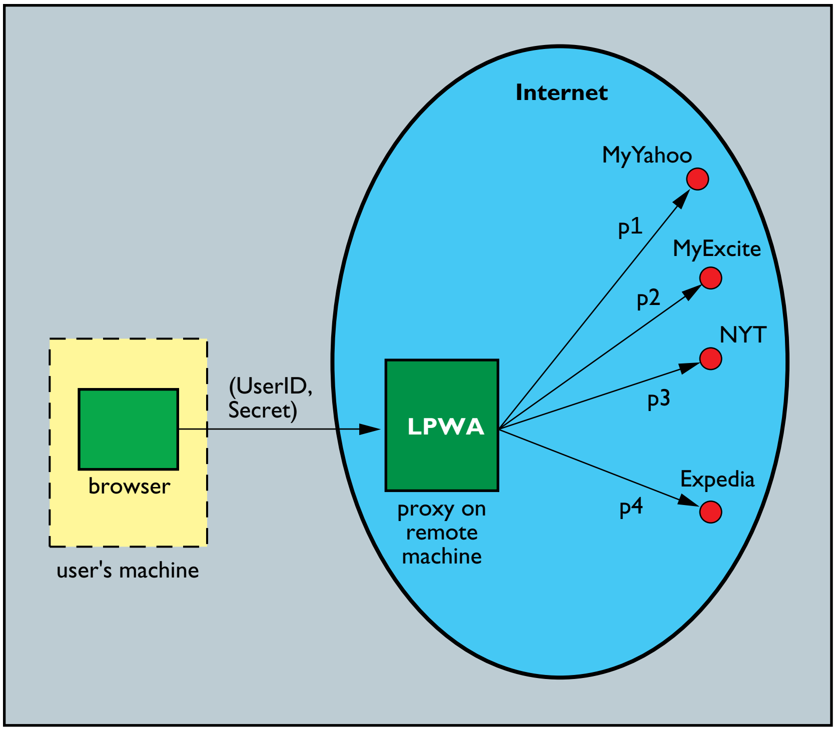


Figure 1. LPWA HTTP proxy configuration

When a Web site asks a user to supply her username, password, or email address, the user simply types the appropriate two character escape sequence (`\u`, `\p` or `\@`, respectively), and LPWA supplies the appropriate alias. The user does not have to remember her alias for each Web site she visited. The user does not have to type a (possibly long) username, password, or email address.

Email forwarding. The LPWA proxy creates an alias email address for a user when the user provides the \@ escape sequence. In [1], we describe an email scheme in which the alias email address generated is the alias username at an appropriate domain, lpwa.com in our case. The email system then stores incoming messages, and a user agent retrieves messages for all aliases that belong to a particular user. This scheme has the advantage that the alias email address generation is trivial and no privacy-compromising information has to be stored on the email system. However, such a scheme is better suited for environments in which the proxy resides on a firewall or an ISP access point.

In our trial configuration of an external proxy, the user typically expects email to be forwarded to his or her real mailbox. In [3], we describe such a scheme and show the resulting alias email address has the same desirable properties as the alias username and password. Actively forwarding without maintaining state implies the alias email address is some sort of encryption of the user's real email address (User ID). The drawback of such a scheme is the proxy and the forwarder must store the secret encryption/decryption key. Possession of this key compromises user privacy, and hence security of this key is paramount. It should be noted that storing the encryption/decryption key does not contradict the statelessness of the proxy since the key is fixed and may be considered as a part of the proxy code.

Anti-spam tool. As part of the Persona Generator, a user obtains a different and seemingly unrelated alias email address for each Web site for which she registered. For example, a user might be known as hwfyh8yocY8XUKm9t5OKvnNW@lpwa.com to my.yahoo.com and as lN8illidPtFk50SthNoXzGuS@lpwa.com to www.expedia.com. This feature enables effective filtering of junk email (spam), as follows: Whenever the LPWA Email Forwarder decrypts an alias email address in order to forward a message to the user's real email address, it includes the alias email address in the CC email header of the forwarded message. We decided to use the CC field because many commercial email readers support filtering of incoming email messages based on this field.

Assume a user registers at www.example.com and LPWA computes bd1YnEW0mot3CX-_UxonbznP@lpwa.com as his or her alias email address. Now the address database at example.com gets sold to spammers. As soon as the user gets the first piece of junk email, he or she can install a local mail filter for the string bd1YnEW0mot3CX-_UxonbznP. This will eliminate all email caused by the selling of the database to spammers, while at the same time email from all other sites is unaffected. Most current anti-spam tools filter according to sender addresses or keywords, both of which are easily changed by spammers (such as address spoofing). Our method is the first to filter according to the recipient address. A spammer who bought the address database from example.com knows the user only as bd1YnEW0mot3CX-_UxonbznP@lpwa.com and therefore cannot change (spoof) this string!

Furthermore, the user can easily keep a small local database, mapping alias email addresses to the Web site for which the address was created. Then, when receiving junk email, the user can determine which Web site is responsible, even when the junk email was sent by a third party. The user can complain to the Web site or take other action as needed.

Statelessness. In the LPWA trial, the HTTP proxy, which comprises the Browsing Proxy and the Persona Generator, is stateless. The proxy gets the user's identity information via an HTTP header (Proxy-Authorization) that accompanies each HTTP request. The user's browser is induced to start sending this header as part of the LPWA login process (see sidebar). The Persona Generator computes the user's aliases on the fly from the information in the header, plus the domain name of the destination Web site, thus obviating the need to store any identity information in the proxy. Of course, the user must log in to LPWA with the same identity information each time to get consistent LPWA aliases.

Related Work

Our work concentrates on *data anonymity*, which protects the identity of the user by careful modification of the data she exchanges with the world. In most cases, data anonymity means the system does not reveal identifying information about the user. However, for personalized Web sites, data anonymity means the ability to present distinct persona for each Web site, so the user may establish personalized accounts without revealing her identity. Another type of anonymity is connection anonymity, which protects the identity of the user by disguising the communication path between the user and the rest of the world. Table 1 compares the capabilities of several systems for providing connection anonymity, data anonymity (removal of identifying information), and personalization (presentation of distinct persona).

LPWA provides filtering for data anonymity and full personalization. It also provides limited connection anonymity by using an HTTP proxy. However, tracing all communication to and from the proxy may reveal the user's identity. Also, LPWA does not filter Java and JavaScript, which may leak information from the browser back to the server.

The Anonymizer[1] is a service that provides limited connection anonymity, high-data anonymity, but no personalization. It is an intermediate entity that filters HTTP headers and removes Java and JavaScript for Web browsing. It rewrites all HTTP pages so that clicking on one of the links causes a request to be sent to the Anonymizer server, which in turn issues the original request. However, there are no features provided for anonymous registration at Web sites, and hence, no simple and secure means for users to preserve data anonymity at Web sites that offer personalized services.

Onion Routing [9] and Crowds [8] are two recent systems that provide a high degree of connection anonymity for Web browsing (see Reiter and Rubin, and Goldschag et al. in this section). Similar to mixmaster remailers, Onion Routing transforms a message into several layers of encryptions ("onions"). Each layer determines the next forwarding node ("onion router"). To enable two-way communication, onion routers maintain connection state. Crowds randomly assigns a native route for each crowd's member ("jondo") among other jondo's before the connection is routed outside the crowd. We note that LPWA can be potentially combined with these tools to give a high degree of both data and connection anonymity.

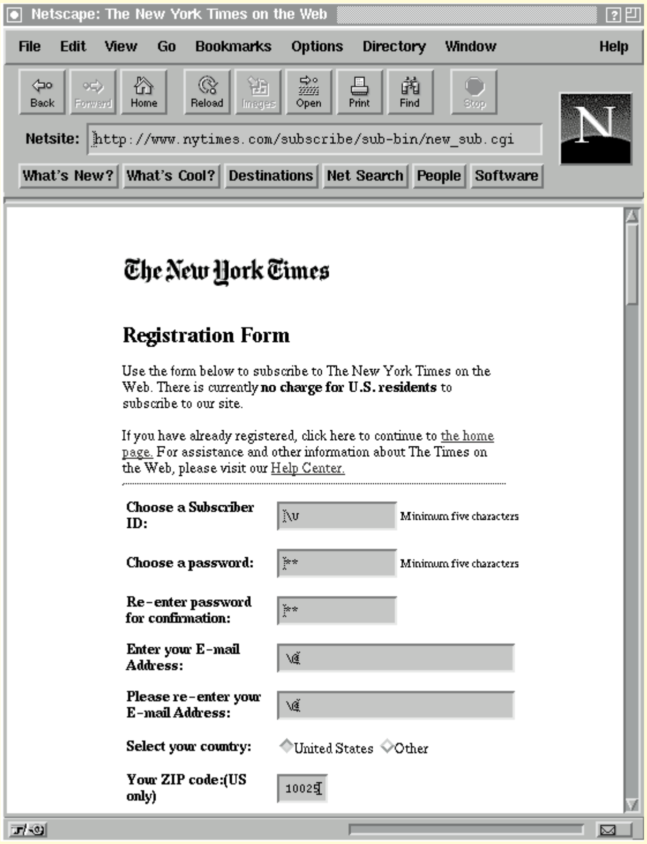
LPWA can also be integrated with the Platform for Privacy Preferences (P3P) standard proposal to make a P3P persona (see [7]) pseudonymous: P3P enables Web sites to express privacy practices and clients to express their preferences about those practices (see Reagle and Cranor in this section). A P3P interaction will result in an agreement between the service and the client regarding the practices associ-

[1] www.anonymizer.com.

Usage of LPWA

The user configures his or her browser's HTTP Proxy setting to use the LPWA HTTP proxy. (The current trial LPWA proxy is located at lpwa.com.) At the beginning of a browsing session, the LPWA login page asks the user to supply a User ID (real email address) and a Secret (universal password). From that point on, LPWA is transparent while the user is browsing the Web.

Whenever a Web site asks the user to supply any username, password, or email address, the user may invoke the persona generator by supplying a corresponding LPWA escape sequence, as depicted in the figure of the *New York Times* Web site. As it passes along the request to the destination Web site, LPWA recognizes these sequences, computes an alias username, password, or email address specific to that Web site, and inserts them into the user's request. In particular, \u is replaced by the alias username, \p is replaced by the alias password, and \@ is replaced by the alias email address. On repeat visits, LPWA will generate those same personae, so when the user returns to a Web site, he or she is recognized as a repeat visitor. When a Web site sends a message to an alias email address, the message arrives at LPWA, which then forwards the message to the corresponding user.



Netsite: http://www.nytimes.com/subscribe/sub-bin/new_sub.cgi

The New York Times

Registration Form

Use the form below to subscribe to The New York Times on the Web. There is currently **no charge for U.S. residents** to subscribe to our site.

If you have already registered, click here to continue to the home page. For assistance and other information about The Times on the Web, please visit our Help Center.

Choose a Subscriber ID:	\u	Minimum five characters
Choose a password:	**	Minimum five characters
Re-enter password for confirmation:	**	
Enter your E-mail Address:	\@	
Please re-enter your E-mail Address:	\@	
Select your country:	◇United States ◇Other	
Your ZIP code:(US only)	10025	

NEW YORK TIMES REGISTRATION PAGE (C) 1997 ated with a client's implicit (such as click stream) or explicit (such as client answered) data. The latter is taken from data stored in a repository on the client's machine, so the client need not repeatedly enter frequently solicited information. A persona is the combination of a set of client preferences and P3P data.

Currently, P3P does not have any mechanisms to assist clients to create pseudonymous personae. For example, a client can choose whether to reveal his or her real email address stored in the repository. If the email address is not revealed, the Web site cannot communicate with the client, and, if the email address is indeed revealed, the Web site has a very good indication of the client's identity. Note that P3P selectively reveals parts of a single persona to each Web site; thus we have classified its personalization capabilities as medium in Table 1. Using LPWA provides a new and useful middle ground—data in the repository that corresponds to usernames, passwords, email addresses, and possibly other fields, can be replaced by macros which, by calling LPWA, expand to different values for different Web sites and create pseudonymous personae for the client.

Conclusions

The LPWA trial has run at lpwa.com since June, 1997, and at press time has attracted over 40,000 unique users. About 40% of those users have logged in more than once. For the last few months, an average of 700 to 800 distinct returning users log into LPWA every day. (We note that LPWA logs the one-way hash value of the User ID and Secret, in order to count users without compromising their anonymity).

This number of users and the network traffic are very encouraging, especially in light of our trial configuration where users outside the New York metropolitan area incur a non-negligible performance degradation by using LPWA, and where the potential user population is mostly restricted to ISP customers, because corporate employees are typically required to use their respective firewall proxy.

We also receive email from users indicating they would seek out ISPs that offer an LPWA service. This mail reveals a hunger in a segment of the user population for both data and connection anonymity, and we think that hunger will grow with users' growing awareness of how easily personal information about them can be abused. This hunger, in turn, could be satisfied commercially in several different places. Anonymity can be provided by generating personae in browsers. It can also be provided by a Web proxy and an email forwarder similar to our trial system, as an added-value service offered by ISPs or third-party, for-fee vendors. Finally, it can be provided by corporate firewalls that generate personae to mask identities.

Compared to other systems, LPWA occupies the middle ground on the anonymity spectrum. While the Anonymizer thoroughly protects an individual's privacy by rewriting content and suppressing Java and JavaScript, it provides no assistance to someone who wants to make use of personalized services. Both Crowds and onion routing also do a better job of providing connection anonymity. LPWA provides simple connection anonymity and, uniquely among these other systems, a simple and effective way to generate and use pseudonyms, as well as receive and filter email.

References

1. Bleichenbacher, D., Gabber, E., Gibbons, P.B., Matias, Y., Mayer, A. On secure and pseudonymous client-relationships with multiple servers. In *Proceedings of the Third USENIX Electronic Commerce Workshop.* (Aug. 1998), 99–108.
2. Cranor, L.F., LaMacchia, B.A. Spam! *Commun. ACM 41*, 8 (Aug. 1998), 74–83.
3. Gabber, E., Gibbons, P.B., Matias, Y., Mayer, Y. How to make personalized web browsing simple, secure, and anonymous. In *Proceedings of Financial Cryptography'97.* Springer-Verlag LNCS 1318, 17–31.
4. Gabber, E., Jacobson, M., Matias, Y., Mayer, A. Curbing junk email via secure classification. *Proceedings of Financial Cryptography.* (Feb. 1998, Anguilla, British West Indies), Springer Verlag LNCS 1465, 198–213.
5. Kristol, D.M., Gabber, E., Gibbons, P.B., Matias, Y., Mayer, A. Design and implementation of the Lucent Personalized Web Assistant (LPWA). Submitted for publication.
6. Martin, D. Internet anonymizing techniques. *;login: Mag.* (May 1998) 34–39.
7. P3P Architecture Working Group. General Overview of the P3P Architecture. www.w3.org/TR/WD-P3P-arch.
8. Reiter, M., Rubin, A. Crowds: Anonymous Web transactions. *ACM Trans. Information and System Security.* To appear. www.research.att.com/projects/crowds/.
9. Syverson, P., Goldschlag, D., Reed, M. Anonymous connections and onion routing. In *Proceedings of the IEEE Symposium on Security and Privacy* (1997) .

See [6] for a recent overview of Internet anonymizing techniques; [2] provides an overview of the junk email (spam) problem, statistics, and tools to combat it. Additional references to the theoretical aspects of alias generation are examined in [1]. The design and implementation of the LPWA system are detailed in [5]. An early description of the LPWA system appeared in [3]. An extension of our anti-spam tool to more general email communication is described in [4].

Eran Gabber (eran@research.bell-labs.com) is a member of the technical staff in the Information Sciences Research Center, Bell Laboratories, Lucent Technologies, in Murray Hill, N.J.

Phillip B. Gibbons (gibbons@research.bell-labs.com) is a member of the technical staff in the Information Sciences Research Center, the Bell Laboratories, Lucent Technologies, in Murray Hill, N.J.

David M. Kristol (dmk@research.bell-labs.com) is a member of the technical staff in the Information Sciences Research Center, the Bell Laboratories, Lucent Technologies, in Murray Hill, N.J.

Yossi Matias (matias@math.tau.ac.il) is a senior lecturer in the computer science department of Tel-Aviv University, Israel.

Alain Mayer (alain@research.bell-labs.com) is a member of the technical staff in the Information Sciences Research Center, Bell Laboratories, Lucent Technologies, in Murray Hill, N.J.