Consistent yet Anonymous Web Access with LPWA

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

Information Sciences Research Center
Bell Laboratories, Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974
{eran, gibbons, dmk, matias, alain}@research.bell-labs.com

October 16, 1998

In recent years the World Wide Web (WWW) has become an immensely popular and powerful medium. 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. However, personalized services 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, registering 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 user name and password for a web-site (e.g., for `my.yahoo.com`) 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 e-mail address at registration time as well, which they use both to verify the user's registration, and, often, to provide part of the personalized service, such as a newsletter or a personalized news digest. For example, the travel site `expedia.com` e-mails best fares for user-preferred airline routes. Moreover, an increasing number of web-sites send a confirmation code to the user's e-mail address. This code must be provided in order to access the account. Hence, the user often must supply a valid e-mail address to use the service at all. But the e-mail 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 e-mail addresses can be easily abused to send out junk e-mail (spam). To counter these concerns, wary users either avoid sites that require them to register, or they register with false information.

This paper describes the *Lucent Personalized Web Assistant (LPWA)*, a novel software system designed to address these user concerns. Users may browse the web in a personalized, simple, private, and secure fashion using LPWA-generated aliases and other LPWA features. LPWA gen-

*Also with Computer Science Dept., Tel-Aviv University, Tel-Aviv 69978 Israel. E-mail: matias@math.tau.ac.il.

erates secure, consistent and pseudonymous aliases (personae) for web users. Each alias consists of an alias username, an alias password and an alias e-mail address. The alias e-mail addresses allow web-sites to send messages to users and enable effective filtering of junk e-mail (spam). LPWA forwards mail addressed to the alias e-mail address to the actual user. LPWA allows users to filter incoming messages based on the recipient address (the alias e-mail), 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, e-mail 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 in the rest of the paper.) The generated aliases are consistent, which means that 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 its aliases.

- *E-mail Service:* Web-sites can use the e-mail address of the supplied persona to send information to the user.

- *Anti-Spamming Support:* Users can filter junk e-mail based on the *recipient e-mail address*, which happens to be the persona e-mail address. Furthermore, the user can infer which web-site is responsible for compromising the e-mail 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, LPWA site does not attract break-ins, and LPWA service may be replicated easily. The absence of state information also implies that 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 E-mail 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*, which is a valid Internet e-mail address for the user; and a *Secret*, which serves 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 *E-mail Forwarder* forwards mail, addressed to a persona e-mail address, to the corresponding user.

The Persona Generator consists of the *Janus function* which was designed to support pseudonymous client-server schemes. The Janus function is based on a suitable combination of cryptographic functions. Its specification and implementation are given in detail in a separate paper [BGGMM98], which also discusses the basic notion of pseudonymous client-server schemes. 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 e-mail 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. The Browsing Proxy might reside on a firewall, an Internet Service Provider (ISP) access point, or a neutral site on the Internet. The E-mail Forwarder needs to reside "away from" the user's machine, since the goal is that the various persona e-mail addresses would be unlinkable to the user. Obviously, there are various trade-offs involved:

- *Trust:* The Persona Generator receives the user's real e-mail address and a secret. The user opens a direct TCP connection to the Browsing Proxy. Depending on the design, the E-mail Forwarder must reliably either store or forward the received messages. Hence, all components must be trusted to various degrees.
- *Anonymity:* Neither the Browsing Proxy's nor the E-mail 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, New Jersey, 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 E-mail Forwarder.

In [BGGMM98], we discuss schemes with components residing on user's machines, ISP access points, or firewalls. Compared to our choice, such configurations have advantages in terms of trust and performance, as discussed in detail in that paper. 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.)

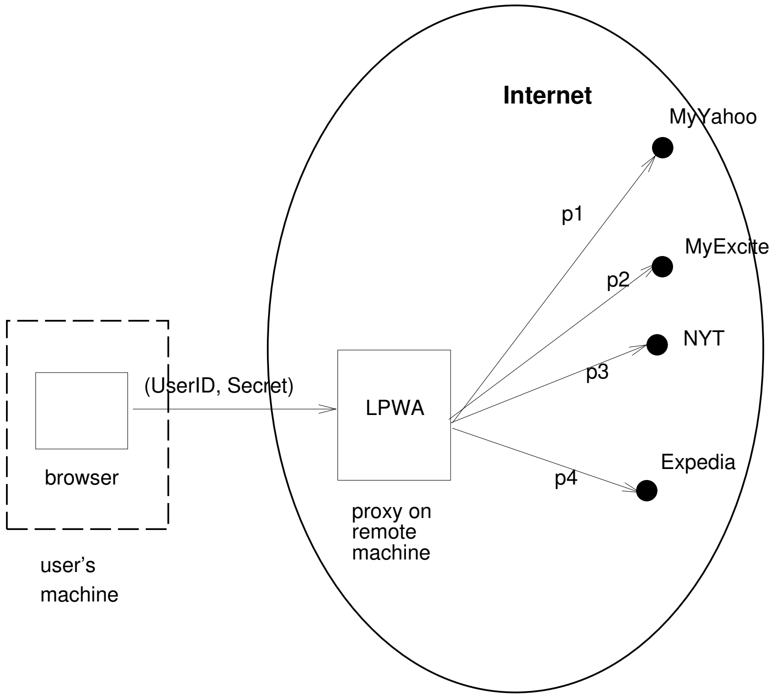

Figure 1: LPWA HTTP proxy configuration

Privacy and Convenience

The user's true identity is protected by LPWA, since aliases cannot be translated back to user names. 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.

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

E-mail Forwarding

The LPWA proxy creates an alias e-mail address for a user in response to the user providing the `\@` escape sequence. In [BGGMM98], we describe an e-mail scheme in which the alias e-mail address generated is the alias username at an appropriate domain; `lpwa.com` in our case. The e-mail 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 e-mail address generation is trivial and that no privacy-compromising information has to be stored on the e-mail 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 as an external proxy, the user typically expects e-mail to be forwarded to her real mailbox. In [GGMM97], we describe such a scheme and show that the resulting alias e-mail address has the same desirable properties as the alias username and password. Actively forwarding without maintaining state implies that the alias e-mail address is some sort of *encryption* of the user's real e-mail address (User ID). The drawback of such a scheme is that 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. Note 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 e-mail 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 `lN8illidPtFk5OSthNoXzGuS@lpwa.com` to `www.expedia.com`. This feature enables effective filtering of junk e-mail (spam), as follows.

Whenever the LPWA E-mail Forwarder decrypts an alias e-mail address in order to forward a message to the user's real e-mail address, it includes the alias e-mail address in the `CC` e-mail header of the forwarded message. We decided to use the `CC` field, since many commercial e-mail readers support filtering of incoming e-mail messages based on this field.

Assume that a user registers at `www.example.com` and LPWA computes `bd1YnEWOmot3CX-_UxonbznP@lpwa.com` as her alias e-mail address. Now the address database at `example.com` gets sold to spammers. As soon as the user gets the first piece of junk e-mail, she can install a local mail filter for the string `bd1YnEWOmot3CX-_UxonbznP`. This will eliminate all e-mail caused by the selling of the database to spammers, while at the same time e-mail 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 (e.g., 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 `bd1YnEWOmot3CX-_UxonbznP@lpwa.com` and hence cannot change (spoof) this string!

Furthermore, the user can easily keep a small local database, mapping alias e-mail addresses to the web-site for which the address was created. Then, when receiving junk e-mail, the user can determine which web-site is responsible, even when the junk e-mail 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 the sidebar.) The Persona Generator computes the user's aliases "on the fly" from the information in the header, plus the domain name

Table 1: Capabilities of Anonymizing Systems

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

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 that 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 that 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 her 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* (see [Anon]) is a service which provides limited connection anonymity, high data anonymity, but no personalization. It is an intermediate entity which 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 [SGR97] and *Crowds* [RR97] are two recent systems that provide a high degree of connection anonymity for web browsing. 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 *Personal Privacy Preferences* (P3P) standard proposal to make a P3P *persona* (see [P3P]) pseudonymous: P3P enables web-sites to express privacy practices and clients to express their preferences about those practices. A P3P interaction will result in an agreement between the service and the client regarding the practices associated with a client's implicit (i.e., click stream) or explicit (i.e., client answered) data. The latter is taken from data stored in a *repository* on the client's machine, so that 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 her real e-mail address, stored in the the repository. If the e-mail address is not revealed, the web-site cannot communicate with the client, and, if the e-mail 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: The data in the repository that corresponds to usernames, passwords, e-mail addresses, and possibly other fields, can be replaced by macros which, by calling LPWA, expand to different values for different web-sites and thus create pseudonymous personae for the client.

See [M98] for a recent overview of Internet anonymizing techniques. The paper [CL98] provides an overview of the junk e-mail (spam) problem, statistics, and tools to combat it. The paper [BGGMM98] contains additional references to the theoretical aspects of alias generation. The paper [KGGMM98] describes in detail the design and implementation of the LPWA system. An early description of the LPWA system appeared in [GGMM97]. An extension of our anti-spam tool to more general e-mail communication is described in [GJMM98].

Conclusions

The LPWA trial has run at `lpwa.com` since June, 1997, and has thus far attracted over 40,000 unique users (by September, 1998). 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).

The above 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, since corporate employees typically are required to use their respective firewall proxy.

Furthermore, we receive e-mail from users indicating that 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 e-mail 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 more 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, along with a way to receive and filter e-mail.

References

[Anon] The Anonymizer. `http://www.anonymizer.com`.

[BGGMM98] D. Bleichenbacher, E. Gabber, P.B. Gibbons, Y. Matias, A. Mayer, On secure and pseudonymous client-relationships with multiple servers. In *Proc. 3rd USENIX Electronic Commerce Workshop*, August 1998, pp. 99–108.

[CL98] L. F. Cranor, B. A. LaMacchia, Spam! *Communications of the ACM*, Vol. 41, No. 8, August 1998, pp. 74–83.

[GGMM97] E. Gabber, P.B. Gibbons, Y. Matias, A. Mayer, How to make personalized web browsing simple, secure, and anonymous. In *Proc. of Financial Cryptography'97*, Springer-Verlag LNCS 1318, pp. 17–31.

[GJMM98] E. Gabber, M. Jacobson, Y. Matias, A. Mayer, Curbing junk e-mail via secure classification. In *Proc. 2nd International Conf. on Financial Cryptography*, Anguilla, British West Indies, February, 1998.

[KGGMM98] D. M. Kristol, E. Gabber, P.B. Gibbons, Y. Matias, A. Mayer, Design and implementation of the Lucent Personalized Web Assistant (LPWA). Manuscript submitted for publication, 1998.

[M98] D. Martin, Internet anonymizing techniques. *;login: Magazine*(The USENIX Association Magazine), May 1998, pp. 34–39.

[P3P] P3P Architecture Working Group, General Overview of the P3P Architecture. `http://www.w3.org/TR/WD-P3P-arch`.

[RR97] M. Reiter, A. Rubin, Crowds: Anonymous web transactions. *ACM Transactions on Information and System Security*, to appear. Manuscript at `http://www.research.att.com/projects/crowds/`.

[SGR97] P. Syverson, D. Goldschlag, M. Reed, Anonymous connections and onion routing. In *Proc. IEEE Symposium on Security and Privacy*, 1997.

SIDEBAR: Usage of LPWA

The user configures her browser's HTTP Proxy setting to use the LPWA HTTP proxy. (The current trial LPWA proxy is located at `lpwa.com`.) Subsequently, at the beginning of a browsing session, the user is presented with the LPWA login page. This page asks the user to supply her User ID (real e-mail address) and 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 of her username, password, or e-mail address, the user may invoke the persona generator by supplying a corresponding LPWA escape sequence, as depicted in Figure 2 for 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 e-mail 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 e-mail address. On repeat visits, LPWA will generate those same personae, so when the user returns to a web-site, she is recognized as a repeat visitor. When a web-site sends a message to an alias e-mail address, the message arrives at LPWA, which then forwards the message to the corresponding user.

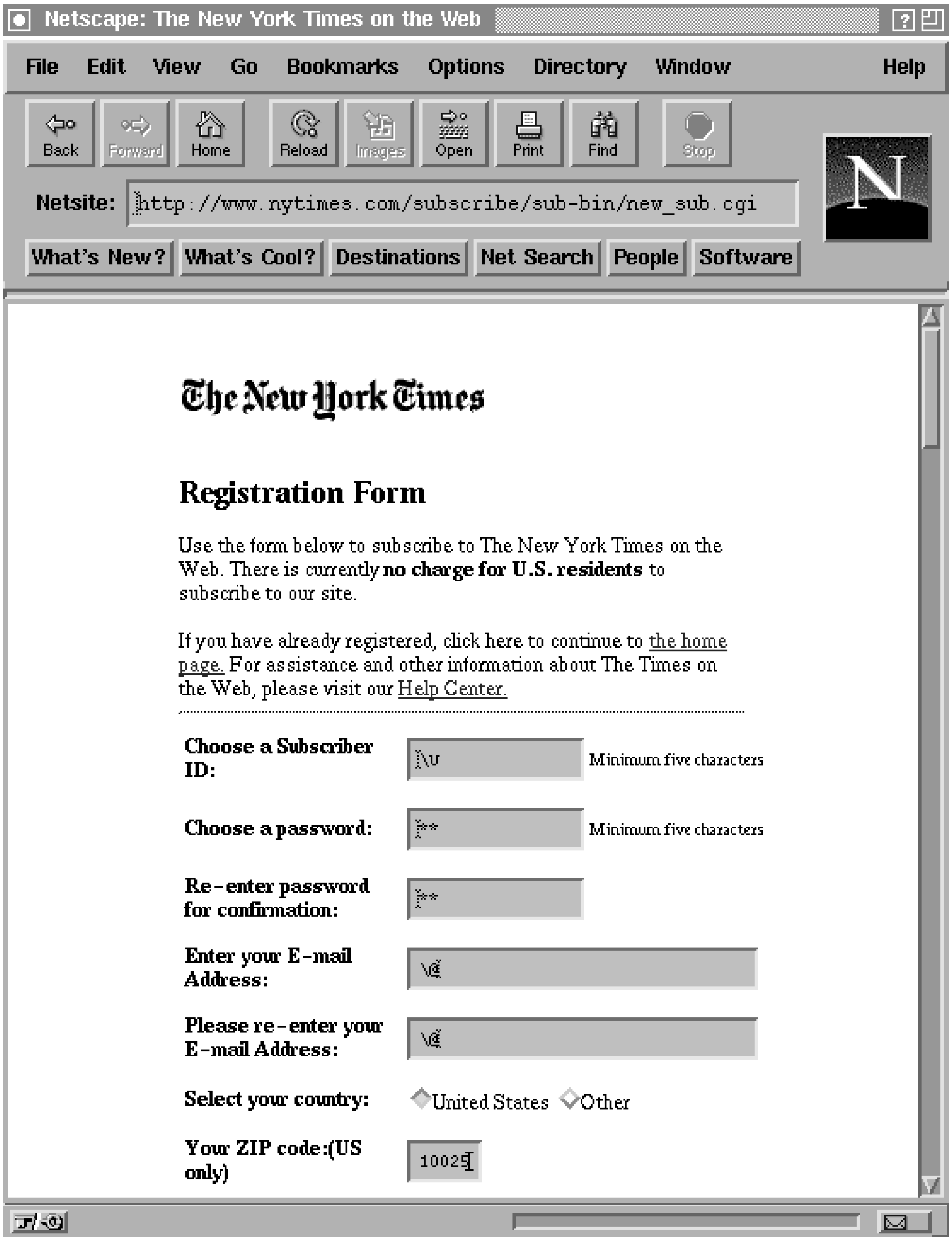

Figure 2: New York Times registration page ©1997