Some data on the Internet can be really intimate

by Keith Darce

First, you had to worry about your credit card number being exposed on the Internet. Then, it was your Social Security number and bank account information.

Add orders for sex products to the list.

The names and addresses of tens of thousands of people who requested free samples of Astroglide - a personal lubricant from Vista, Calif.-based Biofilm - made their way onto Google recently when the massive Internet search engine began digging deeper into Biofilm's servers.

Biofilm has spent the past two weeks trying to erase the customer information from Google and trying to figure out how it leaked.

"If we had had better security, if our security was impeccable, it wouldn't have happened," Biofilm spokeswoman Lynne Merrill said recently.

The incident put Biofilm among a growing number of organizations that have lost control of personal data that they were supposed to protect. Burglars have lifted computers full of sensitive data from homes, offices and automobiles. Government employees have lost work laptops. Hackers have broken into corporate computer networks.

In what has been described as the biggest data theft, at least 45.7 million credit and debit card numbers for customers of TJX Co.'s T.J. Maxx, Marshalls, HomeGoods, HomeSense, A.J. Wright and Winners stores were discovered stolen in recent months.

Although the Biofilm breach didn't put at risk information that could lead to identity theft, it was embarrassing for the company and possibly for the customers whose names popped up on Google.

"We are learning more and more that information that is collected electronically has a way of spreading," said Steven Bellovin, an Internet security researcher at Columbia University in New York. "When you supply information to someone else (on the Internet), that information can be indexed and used in ways you might not want."

Biofilm's problems started in early April when Googlebot, a "crawler" program that mines servers for new information, discovered the Astroglide customer files buried inside the same server that houses Biofilm's Web page data, said Biofilm webmaster Matthew Eckmann.

Google shouldn't have found the files, Eckmann said. The company had taken a number of steps to block outsiders from discovering the data, including using password protection and robots.txt files to instruct friendly Internet crawlers to stay away from the customer lists.

But other unprotected files on the server made reference to the sensitive Astroglide files and provided pathways for Googlebot to follow, Eckmann said.

The first hint of trouble came April 12. A caller told Biofilm that a Google search for the caller's name produced a link to an Astroglide list with information the person provided when requesting a sample. Reports of more leaked lists soon arrived from others doing similar searches.

"I was surprised because in the four years that this information had been stored in the archive, it had never been accessed" by Google, Eckmann said.

More than 550 lists containing more than 260,000 entries were exposed through Google, according to Wired News.

Merrill said Biofilm didn't know exactly how many customers were involved in the data breach.

Information from Biofilm's paying customers always has been kept behind a protective firewall. The company believes the only files involved in the breach were those from the free sample requests.

After learning about the leak, Eckmann moved the free sample files behind a firewall, blocking Internet users from accessing the lists directly. But copies of the pages, known as caches, and index references to the files remained on Google's servers.

The company also took down its free-sample order pages indefinitely to prevent new lists from being created.

It took Google a week to remove the caches. The pages would have been removed more quickly if the files had contained more sensitive information, such as Social Security numbers, according to a Google Web page that instructs users on how to remove pages from the search engine.

A Google spokeswoman, who wouldn't comment on the Biofilm matter, referred to that Web page.

Index references to the Biofilm customer pages, which pop up during searches but do not contain personal information, still could be found on Google in late April. Clicking on the links led an innocuous error message page.

The fallout from the breach might linger for some time.

Biofilm could face lawsuits over the leaks from some of the people on the lists, said Chris Hoofnagle, a senior staff attorney and privacy expert with the University of California Berkeley's Samuelson Law, Technology and Public Policy Clinic.

Lawyers might argue that Biofilm violated a California law that bars companies from engaging in unfair or deceptive trade practices, Hoofnagle said. They also could argue that the company breached another state law requiring Internet businesses to abide by online privacy policies.

Web pages for ordering Astroglide samples included this disclaimer: "All information will be used for mailing purposes only and will not be distributed to any outside organizations."

Additionally, a privacy policy Web page for the company promises not to share personal information with outside parties without receiving permission or disclosing plans to distribute the information when it's collected.

Any financial hit resulting from litigation probably won't be too large, Hoofnagle said. "This type of leak raises certain legal risks. But these cases settle for very reasonable amounts."

Biofilm's troubles should serve as a warning to everyone who uses the Internet to transfer personal information, said Pam Dixon, executive director of the World Privacy Forum, a nonprofit public interest research group in Carlsbad, Calif.

"The technology is advancing, but I don't think people's expectations of privacy have kept up with the way technology has changed," she said.

Bellovin, the Columbia researcher, said people can do little to protect personal information once they have shared it online. Even a sound company privacy policy is no guarantee against mistakes and lapses that can lead to security breaches.

He offered this sobering advice: "Buy things over the counter for cash."

Some data on the Internet can be really intimate by Keith Darce