News: New Perils in Hosted Data?

Greg Goth

One of the newest beta features in the Google Desktop (http://desktop.google.com/about.html) is exposing both theoretical and practical issues about storing data at remote, hosted sites. Search Across Computers, new in Google Desktop 3, lets registered users search the documents on any of their computers via indexed copies stored temporarily on Google servers.

The new feature represents only one of many online service innovations that current network capabilities encourage—and current laws allow. Legal scholars have joined technologists to resolve the potential data-exposure risks without stifling the development of new and useful online services.

Stored copies: Let the user beware

"We're facing a trend where people are being constantly encouraged to store more and more of their data with third parties," says Kevin Bankston, a staff attorney with the Electronic Frontier Foundation, a technology policy watchdog organization. Bankston says the EFF's objection isn't to the trend itself. "To the extent the law keeps up or these things are designed such that the government can't get at your data without you knowing about it, we have no problem with it. But under current law and the design parameters for Search Across Computers, he says, the idea of Google holding copies of documents is "a very frightening thought." Accordingly, the EFF issued a cautionary advisory (http://www.eff.org/news/archives/2006_02.php#004400) urging consumers not to use Search Across Computers.
The Gartner Group technology market research firm, in an initial analysis (pdf, http://www.gartner.com/resources/137800/137896/manage_googles_desktop_search_137896.pdf), also recommended that companies using Google Desktop for Enterprise disable the feature. Analyst Whit Andrews criticized Google's decision to hold a temporary index on a remote server. He also found the company's security assurances about encryption and limited employee access inadequate to the risks posed by transporting the data outside the enterprise.

The old days: ASPs in the 1990s

In many ways, questions surrounding the security of distributed data architectures were first addressed at the tail end of the Internet bubble in the late 1990s, when the application service provider model first gained traction. The original ASP model focused on targeted industry segments, giving birth to a cottage industry of consultants and attorneys with expertise in negotiating ASP service-level agreements (SLAs).

One networking industry veteran remembers the promise of the early ASP days well, when he ran an ASP for a vertical industry. "We were the leading provider of digital asset management for Hollywood motion picture and TV studios," says Marc Orchant, PR and marketing communications manager at VanDyke Software, which provides security technology for data in transmission. "It was all done from a server farm in New Mexico. Everything was done with open source tools, and we had a very solid SLA in place with each of our customers. We built multiple layers of security into our offering because all of this was very proprietary information. The whole point was to have a central depository of all this intellectual property and only accord people access privileges based on who they were and whatever business logic was in place."

Yet ASPs failed to take the network by storm, Orchant says. "They were just a little ahead of their time. The pipes weren't there yet and the tools weren't there yet."

The law lags behind technology

The biggest difference between the ASP landscape and today's distributed capabilities might be the inverse relationship between the technological scale and legal safeguards of
each. Today, data storage and transport costs are falling as fast as Moore's law and the law of supply and demand allow. The bandwidth is available, and tools such as scripting-language frameworks are allowing—some might even say compelling—Google and its competitors to offer ever more remote services to ever more users.

But this leveling of the ASP concept isn't safeguarded by carefully negotiated SLAs. Instead, boilerplate-type end-user agreements tend to protect the service provider. Bankston says that most user contracts or privacy policies are drafted with such "weasly language" providers need only a "good-faith belief that it's important to release your data." Thus, in addition the "black hat" perils that hackers pose to data hosted by third parties, government investigators or perhaps even civil litigants such as divorcing spouses or business rivals might have legal access to it as well.

The EFF's initial advisory highlighted the tiered privacy rights regarding hosted data: if a service is doing nothing more than storing or processing data, the 1986 Electronic Communication Privacy Act protects it. However, if the service is doing anything else, such as scanning data to target delivery of advertising, the protection is very limited. "Google says it is not yet scanning the files it copies from your hard drive in order to serve targeted advertising," the EFF reported, "but it hasn't ruled out the possibility, and Google's current privacy policy appears to allow it." Google didn't respond to a request for comment.

Bankston says that providers generally give themselves a lot of leeway in their contracts and privacy policies. They can get away with it because most people don't read the contract. "Our real hope here is not that companies will stop offering these services," he says. "We want to see innovation; but the leaders in this space, including Google, should be taking a much more public stance seeking to have these laws updated, to increase the level of protection for this remotely stored data and to allow more innovative services. So, even if it's used for more than just storage or processing, it's still protected."

**Short- and long-term solutions**

Bankston says there are ways to protect data online, both short-term—via improved legislation and modified technology—and long-term. For example, a software tool that let users communicate peer-to-peer with their computers is an alternative to hosting data on third-party servers. Bankston also suggests third-party encryption such that only the
data owner has the key. "So if they were approached for your data," he explains, "all they could hand over would be encrypted data. Currently, the Search Across Computers files are encrypted, but Google has the key. And if they are compelled, they can and will decrypt that for the government and possibly for civil litigants."

In Bankston's best-case long-term scenario, easily configured home servers would obviate the need for third-party storage. For the moment, however, this scenario is theoretical and unlikely to become available anytime soon. For one thing, Internet service providers have architected asymmetric data rates in current networks, with much faster rates downstream toward the end user than back into the network. Moreover, ISPs have crafted user agreements forbidding server deployments for consumer installations.

Level access and authentication

Other network observers, however, say the trends toward smaller access devices together with remote storage and processing will ensure the evolution of privacy strategies that move away from those who own data and toward their partners on the network.

Attorney Jeff Aresty is the founder of Internetbar.org (http://www.internetbar.org), a virtual association of lawyers and public policy aficionados interested in harmonizing Internet-related laws globally. Aresty says the best way to tackle the great leveling of Internet access smaller and cheaper devices enable is to assure identity and authentication across the network. Proponents of secure storage and transport have long advocated technologies such as Pretty Good Privacy and public key encryption, he says, but nothing has grabbed the mass psyche yet. Once users worldwide understand how critical the need for a universal authentication architecture is, Aresty says, "you're entering into a different realm with regard to agreements. I think the answer is going to depend upon how the technologists gin up the type of technology that will allow us to create an identity everybody can access. And in that regard, if you think of development of cell phones and smart phones—if such a technology exists that can be put into a smart phone, for example—it will change interrelationships between the developed and developing world." It will take harmonization of laws, but Aresty sees a trend that enables anyone with network access to participate in the global economy.
Golden Copy meets Memex

Greg Linden, founder and CEO of findory.com, a company that applies personalization technology to locating information sources, also follows Google's announcements carefully via his blog. Following a presentation that Google made recently to Wall Street Financial that inadvertently contained notes detailing Google's possible future architecture plans, Linden wrote (http://glinden.blogspot.com/ 2006/03/in-world-with-infinite-storage.html) that the company appeared to envision a time when it can store everything for everybody all the time.

"Google's vision is impressive and broad," Linden says. "They said they should be able to 'house all user files, including emails, Web history, pictures, bookmarks, etc. and make it accessible from anywhere (any device, any platform, etc.),' leading to a world where 'the online copy of your data will become your Golden Copy and your local-machine copy serves more like a cache.'"

Linden says the possibility that Google can provide almost instant access to information anywhere and anytime comes startlingly close to the vision of mid-20th century scientist Vannevar Bush, whose microfilm-based Memex (http://www.theatlantic.com/doc/194507/bush) concept influenced the development of hypertext.

"The productivity impact this would have could be remarkable," Linden says. "An extended memory, instant access to anything you had seen before—it would be a powerful tool."

Conclusion

Before such a tool can be realized, however, technology policy experts and the global standards community must address privacy concerns such as those expressed by the EFF, and developers must fill technological holes such as those expressed by Aresty. Even though the fat-client desktop paradigm of computing isn't going away soon, says VanDyke's Orchant, anyone wishing to compete in the future can't avoid the questions the online storage providers are raising.
"People in business have to look for ways to extend the utility of all the data they've got out to the Internet to take advantage of, and respond to, the virtualization of work, because that's the next reality business has to accommodate itself to."

**Related Links**

- DS Online's Security Community, cms:/dsonline/topics/security/index.xml
- "From AWK to Google: Peter Weinberger Talks Search", http://doi.ieeeecomputersociety.org/10.1109/MSP.2005.123

**Cite this article:** Greg Goth, "New Perils in Hosted Data?", *IEEE Distributed Systems Online*, vol. 7, no. 5, 2006, art. no. 0605-05004.