Digital Fingerprints

Tiny behavioral differences can reveal your identity online

Julie J. Rehmeyer

Early during World War II, British intelligence officers eavesdropped on German radio transmissions, but because the messages were in an encrypted version of Morse code, the British couldn't understand the content. The dots and dashes came in distinctive rhythms, and the Allied spies quickly learned to recognize each Morse code operator's particular style, which the listeners called the operator's "fist."

Having identified the individual code senders, the intelligence officers triangulated signals and traced the operators' movements across the continent—thus tracking the movement of their military units.

Morse code transmissions have, for the most part, been supplanted by more-elaborate forms of electronic communication, the latest being the Internet. And differences remain in the way that people tap out their electronic secrets. Internet users have characteristic patterns of how they time their keystrokes, browse Web sites, and write messages for posting on online bulletin boards. Scientists are learning to use these typeprints, clickprints, and writeprints, respectively, as digital forms of fingerprints.

While the aims of this research are to strengthen password security, reduce online fraud, identify online pornographers, and catch terrorists, the technology is raising some troubling possibilities. "It's a bit scary," says Jaideep Srivastava, a Web researcher at the University of Minnesota in Minneapolis. "The privacy implications are huge." This technology might make it impossible for a person to use the Web anonymously.

As people type messages on their computer keyboards and browse Web sites, they leave a trail of electronic fingerprints. Scientists are investigating those keystroke and mouse-use patterns to develop methods to strengthen security and reduce online fraud.
iStockphoto

Typeprints

In 1980, researchers at the Rand Corporation in Santa Monica, Calif., were looking for ways to increase the security of passwords used for logging into computers. They hit on an idea inspired by the World War II fists. Typists, like Morse code operators, might be identifiable by their rhythms.

The scientists kept track of the time between strokes as seven trained typists each entered three passages of about 300 words. Four months later, the volunteers repeated the task. The researchers found that even without any sophisticated analysis, a person could look at the grids of data showing average pauses between pairs of letters and, without fail, match each pair of samples from each of the typists.

Several companies already sell software packages that take advantage of this phenomenon to strengthen password security. Steven Bender, chief operating officer of iMagic Software in Solvang, Calif., says that because people type passwords so frequently, "we start to move it from the conscious mind to the unconscious, just like a dance step or golf swing." As a result, password typing has a nearly identical rhythm every time a person does it.

The typical typeprint-security package asks a user initially to type in his or her password several times. The program then derives statistics, such as the average time between the strokes. The next time the user logs in, the program permits access only if the keystroke timing is sufficiently similar to its initial data.

A major advantage of this kind of identity verification, unlike retinal scanning and other forms of biometrics, is that it doesn't require any sophisticated equipment at the user's end, Bender says.

Researchers are now developing the technique for application beyond password verification. Daniele Gunetti and Claudia Picardi of the University of Torino in Italy are creating a system that examines typing rhythms—sometimes called keystroke dynamics—while a person uses a computer, not just at log-in. "We are particularly interested in applying the system to track illegal activities around the Internet," Picardi says.

The researchers' system scans a person's normal typing to learn all his or her various typing rhythms, not just the ones that occur in a password. It then continually monitors these rhythms.

If a hacker manages to get into someone's computer account, the typeprint system will notice the different pattern and raise an alarm, perhaps by notifying the system administrator. The researchers reported in 2005 that the system produced about one false alarm in every 200 typing sessions.

This approach could also be used for identifying users of a Web site that requires a significant amount of typing. Online e-mail services such as Gmail or Yahoo are candidates for such protection, Picardi says.

Picardi also points to online bulletin boards. The program could identify posters performing illegal activities, such as soliciting sex from children, says Picardi.

Typeprint analysis raises a number of Orwellian possibilities. Conceivably, police could compile a log of many individuals' typing patterns and then identify users of public computers, such as those in libraries, Picardi says.

Even without a database of individuals' typeprints, authorities might glean information about someone on a public computer or online bulletin board just from that person's keystroke rhythms. For example, they might learn a person's native language because the common keystroke combinations that are typed most quickly vary depending upon the person's native language.

People who write and sell software that directly records the content of what's being typed have been prosecuted for violating wiretap laws. Because keystroke-dynamics programs don't record contents, they aren't expected to be subject to such laws, and no legal difficulties have arisen so far. But in some circumstances, keystroke-timing data might be used to reconstruct a password or even the content of a message.

Gunetti and Picardi's program, for example, records the average time elapsed between keystrokes for each pair of letters but doesn't keep track of the order of the keystroke pairs. In a short typing session, however, that might be enough for someone to guess how to put together the keystrokes into the full message.

Typeprint analysis could also be troublesome in hackers' hands. In 2001, researchers pointed out that typeprints could be used by hackers to listen in when people are working on a computer from a remote location. Secure communication protocols send each keystroke across the Internet encoded in a separate data packet. A hacker can't read the encoded packets directly, but by analyzing the rhythm of the packets, he or she might narrow the possibilities for what has been typed. This vulnerability would be difficult to remove but, so far, it has also proved difficult to exploit.

Challenges remain even for using keystroke analysis to strengthen passwords or to identify the user of a Web site. Keystroke-dynamics software may be fooled if people type differently when they're using an unfamiliar keyboard or when they're tired or drunk or distracted. On the other hand, those variations may be valuable to detect fatigue in situations where alertness is essential.

Clickprints

The keyboard isn't the only method of computer input. With the rise of the Internet and its click-through format, input devices such as the computer mouse are playing an increasingly important role.

SIGNING BY MOUSE-STROKE. To strengthen passwords, researchers developed a system that requires users to move a mouse to mimic their pen-on-paper signatures or to create a doodle.
McOwan

Picardi and Gunetti are testing ways to detect intruders on a computer system by their mouse movements. The researchers suspect that people have identifiable patterns in the shapes and speeds of their usual mouse motions.

Mouse movements can be used to produce signatures, says Peter McOwan of Queen Mary, University of London. He recorded his test subjects as they drew signatures using the mouse—either an imitation of their normal, pen-and-paper signatures or a drawing of their choosing. He used these digital signatures as additions to password entry to strengthen authentication of computer users' identities.

To challenge the strength of his program, he gave test participants the password of a person whose keystroke pattern and tracing signature had been previously recorded. The combined digital signature and keystroke-dynamic analyses rejected more than 95 percent of participants who were acting as intruders, while accepting the legitimate users more than 99 percent of the time, McOwan reported in 2003.

Other researchers are working to identify patterns in the ways in which people click and scroll through Web sites. Balaji Padmanabhan of the Wharton School in Philadelphia and Yinghui Yang of the University of California, Davis are looking for ways to employ what they call clickstream data—what a user clicks on and when—to verify Web site visitors' claimed identities and to prevent fraud online.

Suppose that a person ordinarily visits an online bookseller only on Sunday afternoons, spends around 15 minutes looking through the site, reads reviews of gardening books, and always buys one book with a registered credit card. If on a Monday morning, someone claims to be that person and after 8 minutes tries to buy five books on science fiction, the seller might well suspect fraudulent activity. The seller could then ask for additional verification of the visitor's identity, for example by sending a message to that person's e-mail address on file.

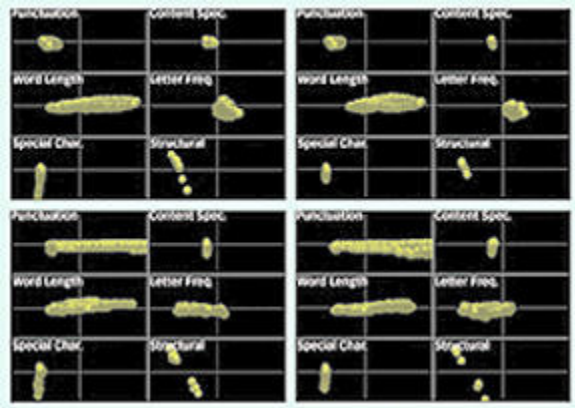
The key to verifying someone's identity lies in accumulating data about that person's behavior from multiple browsing sessions. The researchers' experimental program kept track only of the session's length, time of day, and day of the week and the number of pages viewed. In their study, Padmanabhan and Yang found that a clickstream-data program within a Web site getting small amounts of traffic would need at least 30 browsing sessions to discern the habits of a user. And even then, the program would be only about 80 percent accurate.

Web sites getting more traffic would require analysis of more habits, the researchers say.

If someone didn't want to be identified by clickprint, he or she could easily alter behavior to elude detection, Padmanabhan and Yang say. On the other hand, it would be difficult for crooks to be successful impersonators. "They'd really have to change their behavior in a way that's exactly like the person they're mimicking," Padmanabhan says.

Writeprints

On July 11, 1804, Alexander Hamilton had no idea that he was laying the groundwork for research into online bulletin boards. On that night, as Hamilton prepared for a morning duel with Aaron Burr, he made a list of which of the 85 essays in the *Federalist Papers* he'd written and which ones had been penned by James Madison or John Jay. The duel proved fatal to Hamilton, and Madison subsequently disputed Hamilton's claim of authorship on 12 of the articles.



THE RIGHT WRITEPRINT? A new technique for identifying Internet abusers analyzes a message and plots characteristics of several traits, such as punctuation. The similar shapes show that the top two sets of graphs come from messages by one author, and the bottom two from messages by another.

A. Abbasi and Chen

With the scandal, a puzzle was set for scientists, who have since tried various statistical techniques to characterize the writing styles of the three men. Altogether, researchers have considered more than 1,000 features of writing style. Nearly all the analyses have vindicated Madison.

Hsinchun Chen, a researcher in information systems at the University of Arizona in Tucson, realized that such analysis could be applied to a quite different problem. "It could be used to track anyone who is trying to hide their identity on the Web," Chen says. "They'll leave a trace."

People commonly post anonymously to message boards or employ different user names. Chen seeks to enable law-enforcement officers to detect whether various threatening or illegal posts come from a single user.

Chen and his colleagues have studied messages from the White Knights, a chapter of the Ku Klux Klan; the Al-Aqsa Martyrs, an anti–United States Palestinian group; and English and Chinese bulletin boards where pirated software and music are commonly sold.

The researchers considered the same writing habits that analysts of the *Federalist Papers* had relied on. These include word choice, punctuation, frequency of the passive voice, ratio of upper case letters to lowercase ones, paragraph length, and indentation. Chen's team also analyzed content, looking for hate speech and words such as "for sale."

The online messages presented the researchers with some different challenges from those encountered by analysts of the *Federalist Papers* and other published matter. Web messages tend to be shorter and more casual, with more misspellings and punctuation errors. Furthermore, the researchers had to distinguish individuals among the hundreds of people who post to a bulletin board rather than just among Hamilton, Madison, and Jay.

On the other hand, bulletin board postings offer multiple fonts and colors, greetings, links, varying styles of quotations, and other analyzable features that rarely appear in essays and books.

Chen and his colleagues identified 270 features of English usage and then used a computer program to pick those that most successfully distinguished among writers on bulletin boards. They then employed those 134 features to analyze bulletin board messages. They also chose features valuable for analyzing messages written in Chinese and Arabic.

The team generated a graphic representation, called a writeprint, which showed how consistent each writer was in traits such as punctuation and word length. To do this, the program broke each message into chunks of 50 or 60 words, analyzed the chunks individually, and then plotted the most revealing aspects of the writer's habits. Subsequent messages from that author would be expected to show a similar pattern.

The researchers reported in the April 2006 *Communications of the ACM* (Association for Computing Machinery) that after running an analysis on 30 to 40 messages from any known author, the program could identify subsequent messages by that author with 93 percent accuracy in Chinese, 95 percent in Arabic, and 99 percent in English.

Chen says that he isn't free to discuss details about how his system has been used for law enforcement. He offers only, "We've been successful at bringing up clues that will alert authorities about suspicious people."

He acknowledges that his team's creation could be employed in ways that raise privacy concerns. Governments "could use it to probe political forums or to create a profile of people," he says. "That's the part we want to avoid."

Peter Eckersley, staff technologist with the Internet-privacy group Electronic Frontier Foundation in San Francisco, worries that writeprints will have a chilling effect on whistle blowing and public speech in general. "From this point on," he says, "the writer who would remain anonymous in the face of serious scrutiny will have to take unusual recourse to the thesaurus and a syntactic scrambler."

Eckersley has additional worries about the writeprint program's future potential for abuse. "If a malicious linguist decided that she didn't like a particular Muslim community leader, what would stop her from making anonymous, terrorism-inciting posts [to the Web], deliberately crafted to match his writing style?" asks Eckersley. "Could she get his home raided just by doing that?"

It may be many years before the full impact of digital fingerprints become clear. But the effect that telegraphers' fists had on World War II suggests that subtle patterns of people's Internet communication will yield powerful information.

If you have a comment on this article that you would like considered for publication in *Science News*, send it to editors@sciencenews.org. Please include your name and location.

References:

Abbasi, A., and H. Chen. 2005. Applying authorship analysis to extremist-group web forum messages. *IEEE Intelligent Systems* 20(September):67–75. Abstract available at http://dx.doi.org/10.1109/MIS.2005.81.

Everitt, R.A.J., and P. McOwan. 2003. Java-based Internet biometric authentication system. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(September):1166-1172. Abstract available at http://dx.doi.org/10.1109/TPAMI.2003.1227991

Gaines, R.S., *et al*. 1980. Authentication by keystroke timing: Some preliminary results. Rand. Report R-256-NSF. Rand Corporation. Available at http://www.rand.org/pubs/reports/R2526/.

Gunetti, D., and C. Picardi. 2005. Keystroke analysis of free text. *ACM Transactions on Information and System Security* 8(August):312–347. Available at http://portal.acm.org/citation.cfm?id=1085129.

Li, J., R. Zheng, and H. Chen. 2006. From fingerprint to writeprint. *Communications of the ACM* 49(April):76-82. Abstract available at http://doi.acm.org/10.1145/1121949.1121951.

Padmanabhan, B. and Y. Yang. Unpublished manuscript. Clickprints on the Web: Are there signatures in Web browsing data? Available at http://knowledge.wharton.upenn.edu/papers/1323.pdf?CFID=720523&CFTOKEN=57530247.

Song, D., *et al*. 2001. Timing analysis of keystrokes and timing attacks on SSH. Proceedings of the 10th USENIX Security Symposium. Available online at http://www.usenix.org/publications/library/proceedings/sec01/song.html.

Further Readings:

Center For Democracy and Technology Guide to Online Privacy. Available at http://www.cdt.org/privacy/guide/introduction/.

Sources:

Steven Bender
iMagic Software
565 Rancho Alisal Drive
Solvang, CA 93463

Hsinchun Chen
Department of MIS
College of BPA

University of Arizona
Tucson, AZ 85721

Peter Eckersley
Electronic Frontier Foundation
454 Shotwell Street
San Francisco, CA 94110-1914

Peter McOwan
Department of Computer Science
Queen Mary, University of London
Mile End Road
London E1 4NS
United Kingdom

Balaji Padmanabhan
Wharton School
University of Pennsylvania
561 Jon M Huntsman Hall
3730 Walnut Street
Philadelphia, PA 19104

Claudia Picardi
c/o Dipartimento di Informatica
corso Svizzera 185
10149 Torino
Italy

Jaideep Srivastava
University of Minnesota
EE/CS 5-209
200 Union Street SE
Minneapolis, MN 55455

Yinghui Yang
Graduate School of Management
University of California, Davis
Room 145, AOB IV
One Shields Ave.
Davis, CA 95616

From *Science News*, Vol. 171, No. 2, Jan. 13, 2007, p. 26.