

SCIENTIFIC AMERICAN

[Features](#) - August 18, 2008

Pedophile Proof Chat Rooms?

Can Lancaster University's Isis Project keep children safe online without invading our privacy?

By Brendan Borrell

Czech police [nab a man](#) suspected of raping 12-year-old girls after offering them car rides via an Internet Web site. In Ohio, a 400-pound man, likewise, uses a Web site to [impersonate](#) a 15-year-old boy in order to convince a 12-year-old girl to send photographs and videotapes of herself naked. A [sting operation](#) in the U.K. shuts down a pedophile chat room Web site, and the site's leader is caught with over 75,000 pornographic images.

Social networking over the Web has helped connect millions of Internet users, but all of this online interaction can also have a serious downside: a proliferation of pedophiles who use code words to trade in child pornography or prowl chat rooms and befriend underage victims, peppering their messages with words like "kewl" and other [youthful colloquialisms](#).



In a move that pits technology against criminals (and, some fear, privacy), a group of researchers at Lancaster University in England and law-enforcement officials at the United Kingdom's [Child Exploitation and Online Protection Center](#) (CEOP) is developing software that tracks the Web's evolving child pornography lexicon as well as predators' chat strategies to help law-enforcement agencies catch the most secretive of these criminals before they strike.

"There's a list of about 50 key words that are very indicative of child pornography," says Skinner, who sometimes coordinates with CEOP as part of the Global Virtual Task Force. But, he says, "The terms do change."

That's why [Awais Rashid](#), a Lancaster University computer science professor, has launched the three-year [Isis Project](#), which uses linguistic analysis to keep tabs on these Internet-savvy pedophiles. "There's so much activity it's virtually impossible to police," he says. Currently, investigators at ICE (U.S. Department of Homeland Security) and CEOP are left waiting for a potential victim to report suspicious activity, but by then, it is often too late.

Rashid's strategy is to create automated monitoring tools that operators of chat rooms, social networks, and file-sharing networks would install on their sites. This will provide law-enforcement officials with an automatically updating dictionary of these code words, along with an alert system which will inform them when users are detected masquerading as children.

In preparation to write the software module for file-sharing networks (a prototype by the end of the year), the team sifted through an entire month of search traffic on the number one peer-to-peer file-sharing network, [Gnutella](#), between February 27th and March 27th 2005. Because each peer participates in routing network messages to and from other peers, Rashid's team could set up a specialized client to intercept and log these queries throughout large segments of the network. Then, two specialists at CEOP analyzed 10,000 keyword searches from three separate days to determine whether they contained references to child pornography. About one in every 100 searches was for such material, and about 1.6 percent of search results received contained such material. Because of the size of the Gnutella network, which had a population of 1.81 million users that year, thousands of child-pornography related searches are being

conducted every minute. For comparison, ICE arrests about 2,500 child predators in the U.S. per year.

Even the experts were puzzled by the ciphers in some of these searches. About 53 percent of search terms and 88 percent of the search results contained code words that had not been tallied by CEOP. The agency may have eventually discovered them during the course of their investigations, but Rashid's team realized they could stay ahead of this "cat-and-mouse game" with the help of a computerized strategy.

As a proof-of-concept, human volunteers unversed in the child-pornography lexicon were given 10 popular Gnutella code words, such as "ITA" (Italy) or "PTHC" (Preteen Hardcore) and then asked to guess which were related to child pornography. The volunteers succeeded less than half of the time. But after these same volunteers had a chance to look at the entire search query in which these key words were nested, their success rocketed to 94%.

Rashid now needs to use this principle, termed collocation, in a module that will provide law enforcement with an evolving "dictionary" of the code words. The second phase of Isis, for monitoring chat rooms, is still in its infancy but will require analyzing not just code words but word frequency and sentence patterning.

Other technological efforts have focused on developing image analysis software for the [National Child Victim Identification Program](#) database and in developing surveillance systems. For example, the FBI's now-retired Carnivore, which did not use linguistic analysis, but could "sniff" email traffic and monitor keywords.

Some experts have reservations about the Isis plan, particularly if it allows law-enforcement agencies to amass dossiers on specific individuals outside of a criminal investigation. "If this is something that any government is mandating a social-networking service do," says John Morris, general counsel for the Center for Democracy and Technology, "then that raises enormous challenges." He says he would have no problem with social networks voluntarily cooperating with law-enforcement agencies and disclosing their privacy practices to customers—as they do now—but a government mandate would place an enormous burden on sites that play no role in the distribution of child pornography and threaten the privacy of law-abiding citizens. "A huge problem with any sort of mandate is it is very hard to define what a social-network site is without sweeping in every blog in this country and Ebay and Amazon, all of which allow you to have profiles about yourself."

[Anthony Finkelstein](#), a computer scientist at University College London who has worked on privacy tools to help agencies share data about child welfare cases, simply feels like the Isis Project's blanket monitoring strategy is misguided. "Do I regard this as being a critical issue? On balance, I'm not 100% convinced, but I think it's worth some further investigation." His biggest concern is that Isis, like a neighbor's infuriating home alarm, is bound to produce a lot of false positives that need to be investigated by law-enforcement officials. "Even if you are able to identify those false positives," he says, "it takes effort to do so, and that's effort that's not devoted to something else."

Finkelstein believes investigators should stick to traditional intelligence-gathering efforts rather than a blanket monitoring scheme and that improved funding of social programs for teenagers could prevent them from falling victim to online pedophiles. Indeed, one [recent study of online crimes against children](#) estimates that deception took place in only 5 percent of these cases, and most illicit activity involved teenagers who were aware they were meeting an adult who was looking for sex.

"I think this is an important area to research, undertaken by competent and well-motivated people," he says, however, "At the end of the day there are a difficult set of choices to be made as to whether or not this is the right type of technology to implement."

Further Reading

[International Report: What Impact Is Technology Having on Privacy around the World?](#)

[How I Stole Someone's Identity](#)

[Do Social Networks Bring the End of Privacy?](#)

[How Loss of Privacy May Mean Loss of Security](#)

[Big Brother Sees All in the Technological Fishbowl](#)

[Phoning It In: Software Turns Mobile Phone into Personal Newscam](#)

[I Hear Ya: Bush Signs Expanded Wiretap Power into Law](#)

[Software's Dirty Little Secret](#)