Researchers develop prototype to detect fake websites

July 26 2011, By La Monica Everett-Haynes

It seems logical that a more Internet-driven world would translate into a heightened awareness of fake websites. But it isn't so. The vast majority of people still are unable to determine the authenticity of websites, resulting in tremendous monetary loses. That is what is driving the work of UA Artificial Intelligence Lab members who, along with a UA alumnus, have earned a top honor from MIS Quarterly for their research.

(PhysOrg.com) -- Do you go online to pay bills, shop, transfer funds, sign up for classes, send email or instant messages or search for medical information? If so, then this pertains to you.

Members of a University of Arizona Eller College of Management team and a UA alumnus developed a prototype system to detect fake websites. When tested against other existing commercial systems, the team found that its system resulted in effective and more accurate detections of

spoof sites – better than a human can.

The team's subsequent article, “Detecting Fake Websites: The Contribution of Statistical Learning Theory" was published last year in an issue of *MIS Quarterly*, or MISQ. A preeminent peer-reviewed journal in the field of management information systems, MISQ has since been named the article its top paper for 2010.

"Even to get into MISQ is very difficult, and this is probably the first technical paper to receive the Best Paper award," said Hsinchun Chen, the UA Artificial Intelligence Lab director, one of the paper's five authors.

MISQ will formally honor the researchers in Shanghai, China later this year during the International Conference on Information Systems.

"The topic of detecting fake websites and also our computational approach are both considered major contributions. This topic has great relevance to the industry, the society and the citizens in general," said Chen, also the McClelland Professor of Management Information Systems.

"This award is not something just for me, or my lab, but also for our department," he said, adding that the team's eventual goal is technology transfer.

UA alumnus Ahmed Abbasi, now a University of Virginia assistant professor of information technology, is lead author on the paper. Chen served as his dissertation adviser. Other co-authors are UA Eller College's department of management information systems faculty members Zhu Zhang and Jay F. Nunamaker Jr.; and David Zimbra, a doctoral student in the Artificial Intelligence Lab.

For the research, the team used the prototype and several other detection systems to evaluate the authenticity of 900 websites.

It is easy to pick up on a site's authenticity by checking whether the URL contains "http" when it should read "https," when it was last updated, if a security key is missing or if images appear strangely pixelated.

The team found that its system – founded on statistical learning technology, which evaluates a large accumulation of data – was more apt to detect imitation sites and those that were entirely concocted, said Abbasi, who earned his doctoral degree in management information systems from the UA in 2008.

The major difference between the authors' prototype and the other systems? Their system relied on a tremendously rich set of fraud cues.

The team developed five categories with thousands of cues, finding that the best results were attained when utilizing thousands of highly visible and also deeply embedded cues, such as placement, URL length, the number of links, characters types on the site and how thorough the site's "frequently asked questions" section is detailed, among other features.

The project's origins were born out of the Artificial Intelligence Lab, where Abbasi developed the mathematical formula the team eventually used while working as a project lead and research associate. He continued the work after having taken a faculty position at the University of Wisconsin-Milwaukee.

"It creates a greater awareness for a problem that has been around for a while yet still remains an issue as we increasingly move to the Internet for everything – online banking, online health initiatives and medical information," Abbasi said.

Given the pervasive nature of online phishing scams, being able to readily and frequently detect a site's validity is crucial, Abbasi said, also noting research that indicates people are less than 60 percent accurate in detecting fake sites, and other security issues.

"The problem we're looking at is quite big. Fake websites constitute much of the Internet fraud's multi-billion dollar industry, and that is monetary loss…we can't even quantify the social ramifications," Abbasi said. "That's the whole motivation. It is so profitable for fraudsters, and it is slipping through the cracks."

Today, Chen and more than one dozen of his collaborators are continuing to investigate fake sites. Meanwhile, Abbasi is undertaking an investigation of peoples' abilities to detect fake sites through a grant funded by the National Science Foundation.

Today, Chen and more than one dozen of his collaborators are continuing to investigate fake sites. Meanwhile, Abbasi is undertaking an investigation of users and peoples' abilities to detect fake sites.

Abbasi said developing better detection systems requires improved statistical learning technology that utilize larger quantities of cues. It also is important to dismiss long-held perceptions about how fake sites might and should appear.

"The idea of protecting from the front level has been around for a while," Abbasi said, adding that companies have begun to employ software that better detects fake sites. "But we are not where we need to be, and there is a lot of potential in future development."

Provided by University of Arizona