Researchers develop grammar-aware password cracker

January 24 2013

When writing or speaking, good grammar helps people make themselves be understood. But when used to concoct a long computer password, grammar—good or bad—provides crucial hints that can help someone crack that password, researchers at Carnegie Mellon University have demonstrated.

A team led by Ashwini Rao, a software engineering Ph.D. student in the Institute for Software Research, developed a password-cracking algorithm that took into account grammar and tested it against 1,434 passwords containing 16 or more characters. The grammar-aware cracker surpassed other state-of-the-art password crackers when passwords had grammatical structures, with 10 percent of the dataset cracked exclusively by the team's algorithm.

"We should not blindly rely on the number of words or characters in a password as a measure of its security," Rao concluded. She will present the findings on Feb. 20 at the Association for Computing Machinery's Conference on Data and Application Security and Privacy (CODASPY 2013) in San Antonio, Texas.

Basing a password on a phrase or short sentence makes it easier for a user to remember, but the grammatical structure dramatically narrows the possible combinations and sequences of words, she noted.

Likewise, grammar, whether good or bad, necessitates using different parts of speech—nouns, verbs, adjectives, pronouns—that also can

undermine security. That's because pronouns are far fewer in number than verbs, verbs fewer than adjectives and adjectives fewer than nouns. So a password composed of "pronoun-verb-adjective-noun," such as "Shehave3cats" is inherently easier to decode than "Andyhave3cats," which follows "noun-verb-adjective-noun." A password that incorporated more nouns would be even more secure.

"I've seen password policies that say, 'Use five words,'" Rao said. "Well, if four of those words are pronouns, they don't add much security."

For instance, the team found that the five-word passphrase "Th3r3 can only b3 #1!" was easier to guess than the three-word passphrase "Hammered asinine requirements." Neither the number of words nor the number of characters determined password strength when grammar was involved. The researchers calculated that "My passw0rd is $uper str0ng!" is 100 times stronger as a passphrase than "Superman is $uper str0ng!," which in turn is 10,000 times stronger than "Th3r3 can only b3 #1!"

The research was an outgrowth of a class project for a masters-level course at CMU, Rao said. She and Gananand Kini, a fellow CMU graduate student, and Birendra Jha, a Ph.D. student at MIT, built their password cracker by building a dictionary for each part of speech and identifying a set of grammatical sequences, such as "determiner-adjective-noun" and "noun-verb-adjective-adverb," that might be used to generate passphrases.

Rao said the grammar-aware password cracker was intended only as a proof of concept and no attempt has been made to optimize its performance. But it is only a matter of time before someone does, she predicted.

Provided by Carnegie Mellon University