Inside Risks | Evgeniy Gabrilovich and Alex Gontmakher

The Homograph Attack

Old-timers remember crossing zeros (Ø) in program listings to avoid confusing them with the letter *O*. This has long been obsoleted by advances in editing tools and font differentiation. However, the underlying problem of character resemblance remains, and has now emerged as a security problem.

On April 7, 2000, an anonymous site published a bogus story intimating that the company PairGain Technologies was about to be acquired for approximately twice its market value. The site employed the look and feel of the Bloomberg news service, and thus appeared authentic to unsuspecting users. A message containing a link to the story was simultaneously posted to the Yahoo message board dedicated to PairGain. The link referred to the phony site by its numerical IP address rather than by name, and thus obscured its true identity. Many readers were convinced by the Bloomberg look and feel, and accepted the story at face value despite its suspicious address. As a result, PairGain stock first jumped 31%, and then fell drastically, incurring severe losses to investors. This hoax was discovered promptly. However, forthcoming Internet technologies may make such attacks more elusive and devastating.

A new initiative, promoted by a number of Internet standards bodies including IETF and IANA, allows one to register domain names in national alphabets. This way, for example, Russian news site gazeта.ru (*gazeta* means *newspaper* in Russian) might register a more appealing *газета.ру*. The initiative caters to the genuine needs of non-English-speaking Internet users, who currently find it difficult to access Web sites otherwise. Several alternative implementations are currently being considered, and we can expect the standardization process to be completed soon.

The benefits of this initiative are indisputable. Yet the very idea of such an infrastructure is compromised by the peculiarities of world alphabets. Revisiting our newspaper example, one can observe that Russian letters *a, e, p, y* are indistinguishable in writing from their English counterparts. Some of the letters (such as *a*) are close etymologically, while others look similar by sheer coincidence. (As it happens, other Cyrillic languages may cause similar collisions.)

With the proposed infrastructure in place, numerous English domain names may be *homographed*—maliciously misspelled by substitution of non-Latin letters. For example, the Bloomberg attack could have been crafted much more skillfully, by registering a domain name bloomberg.com, where the letters *o* and *e* have been faked with Russian substitutes. Without adequate safety mechanisms, this scheme can easily mislead even the most cautious reader.

Sounds frightening? Here is something more scary.

One day John Hacker similarly imitates the name of your bank's Web site. He then uses the newly registered domain to install an eavesdropping proxy, which transparently routes all the incoming traffic to the real site. To make the bank's customers go through his site, John H. hacks several prominent portals that link to the bank, substituting the bogus address for the original one. And now John H. has access to unending streams of passwords to bank accounts. Note that this plot can be in service for years, while customers unfortunate enough to have bookmarked the new link might use it forever.

Several approaches can be employed to guard against this kind of attack. A simple fix would indiscriminately prohibit domain names that mix letters from different alphabets, but this will block names like CNNenEspañol.com. More practically, the browser can highlight international letters present in domain names with a distinct color, although many users may find this technique overly intrusive. A more user-friendly browser may highlight suspicious names, such as ones that mix letters within a single word. For additional security, the browser can use a map of identical letters to search for collisions between the requested domain and similarly written registered ones.

Caveat: To demonstrate the feasibility of the described attack, we registered a homographed domain name www.microsoft.com with Russian letters *c* and *o*. While this name may be tricky to type in, you can conveniently access it from www.cs.technion.ac.il/~gabr/papers/-homograph.html. (Predictably, MICR0S0FT.com, MICR0SOFT.com, and ICROS0FT.com—with *0*s replacing *o*s—are already registered, as is BL00MBERG.com. John H. has not been wasting his time.)

So, next time you see *microsoft.com*, where does it want to go today? **C**

Evgeniy Gabrilovich (gabr@acm.org) and **Alex Gontmakher** (gsasha@cs.technion.ac.il) are Ph.D. students in Computer Science at the Technion–Israel Institute of Technology.

PAUL WATSON