Content providers phishing for demographic data via logins

Oleh Indra Febria Widy
Kamis, 10 Juni 2010 09:07

There has been a steady flow of academic studies into the insecurity of the username/password authentication system (a number of which we've covered at Ars) that suggest it's doomed to failure: humans have a limited memory capacity for unique strings of random characters, which is precisely what most experts recommend as a secure password. A pair of academic researchers from Cambridge have analyzed the use of passwords by many prominent online sites, and found that many sites require passwords as a sort of security theater, requiring them in contexts that are superfluous and failing to do their part to secure the information on their end. The end result, they argue, is a tragedy of the commons, with the commons being the finite memory of the average user.

The paper in which the duo make this argument was presented at the Workshop on the Economics of Information Security, and the paper itself is an interesting mix of economic arguments and security analysis. We'll go through the latter first before tying it back together with the former.

The researchers review the extensive literature on passwords, describing how security experts have identified a series of best practices, such as long passwords that contain non-alphanumeric characters and are changed on a regular basis. They then discuss how users have consistently failed to implement them, and instead settle on shorter, memorable passwords, and typically reuse a small set of them. Alternate systems have their own issues—mnemonics are subject to dictionary attacks using quotations and song lyrics, while leveraging humanity's ability to rapidly recognize images runs up against the reality of the finite ways we have of presenting images.

Given that we haven't adopted anything that's clearly better than the username/password system, the authors look into whether the websites that typically require logins actually do a decent job of promoting best practices. The answer was, in general, no.

To perform the analysis, the authors picked three categories of sites—content providers, e-merchants, and personal identity hosts (think Facebook). Using Alexa traffic rankings, they randomly chose 25 sites from the top 100 in each category, and then 25 more from deeper into the list. They then created accounts and tested all the aspects of password management, from creation to resets.

The results were pretty bad. Of the 150 sites, nearly 80 percent of the sites provided users with no hint as to what makes for a secure password, while less than 10 percent implemented a dictionary check that would prevent the use of "password." Only seven sites enforced the use of a digit in the password, and a grand total of two required a non-alphanumeric character.

Many of the sites transmitted the password back in plaintext, and one even sent it back as POST form data in a way that ensured it remained in the browser's history; only three sites actually used the browser to hash the password before transmitting it. Logins and password resets had their own issues. 126 sites set no restrictions on login attempts, enabling the authors to try a brute-force attack on the system. When using the reset forms, 16 sites simply sent the users' passwords back in plaintext in the body of an e-mail.

The economics of passwords

The authors then broke things down by category, and found an obvious pattern: the worst offenders, security wise, were typically content providers. As the authors recognize, the stakes here are pretty low; most content providers are handing out ad-supported content for free, so the security of a login to those sites might not be a high-stakes affair. But, because users tend to recycle so many passwords, that may not be the case. Getting a user's password to a newspaper's site, for example, may leave that same user vulnerable on a site that stores more significant information.

The authors also point out that many content providers could do without logins entirely, and argue they're required for business, rather than security purposes. In their study, they found that content providers were far more likely to request demographic information during account creation, and more likely to require an account validation via e-mail. From this perspective, the username/password is just a ritual that eases the acquisition of valuable demographic data.

They suggest another economic aspect of this ritual: establishing an account builds trust and establishes a relationship, often committing a user to further interactions. This is true for even noncontent sites, as the authors illustrate by considering OpenID, a single-password signon system based on open standards. Only four sites in their survey enable login via OpenID, and two of the big ones—Google and Yahoo—discourage its use. They were only able to figure out how to use OpenID with Google by finding instructions on a third-party site; Yahoo's instructions refer to it as a "geeky" solution, and the site uses a complex unicode URL to access the feature.

This tendency to require logins for business (rather than security) reasons is what's creating a tragedy of the commons, in the authors' view. "Consumers' finite mental storage capacity for passwords is a common good from the viewpoint of website operators," they argue. The cost of poor security ends up being a negative externality—users end up choosing and reusing weak passwords because they need to generate so many, which puts other companies and their own data at risk.

Is there any good news here? By following the traffic numbers, the authors find that many of the sites with the best password practices are growing at a pace higher than their peers. There are plenty of possible explanations for this—rapidly growing sites might have more resources to apply to security, or may simply be staffed by more technically adept people, etc. But it at least suggests that users will ultimately be exposed to more instances of good password practices.