Cognitive Hacking and Digital Government:

Digital Identity

Paul Thompson
Institute for Security Technology Studies
Thayer School of Engineering, Dartmouth College
Hanover, New Hampshire 03755, U.S.A.

ABSTRACT

Recently the National Center for Digital Government held a workshop on "The Virtual Citizen: Identity, Autonomy, and Accountability: A Civic Scenario Exploration of the Role of Identity in On-Line. Discussions at the workshop focused on five scenarios for future authentication policies with respect to digital identity. The underlying technologies considered for authentication were: biometrics: cryptography, with a focus on digital signatures; secure processing/computation; and reputation systems. Most discussion at the workshop focused on issues related to authentication of users of digital government, but, as implied by the inclusion of a scenario related to ubiquitous identity theft, there was also discussion of problems related to misinformation, including cognitive hacking. Cognitive hacking refers to a computer or information system attack that relies on changing human users' perceptions and corresponding behaviors in order to succeed. This paper describes cognitive hacking, suggests countermeasures, and discusses the implications of cognitive hacking for identity in digital government. In particular, spoofing of government websites and insider misuse are considered.

Keywords: computer security, cognitive hacking, website spoofing, insider misuse, computer security countermeasures

1. INTRODUCTION

On 28-29 April 2003, the National Center for Digital Government at Harvard's Kennedy School of Government held a workshop on "The Virtual Citizen: Identity, Autonomy, and Accountability: A Civic Scenario Exploration of the Role of Identity in On-Line Governance [1]. The National Center for Digital Government is exploring issues related to the transition from traditional person-to-person provision of government services to the provision of such services over the Internet. As excerpted from the Center's mission statement:

> Government has entered a period of deep transformation heralded by rapid developments in information technologies. The promise of digital government lies in the potential of the Internet to connect government actors and the public in entirely new ways. The outcomes of fundamentally new modes of coordination, control, and communication in government offer great benefits and equally great peril [2].

Discussions at the workshop focused on five scenarios for future authentication policies with respect to digital identity:

- Adoption of a single national identifier
- Sets of attributes
- Business as usual, i.e., continuing growth of the use of ad-hoc identifiers
- Ubiquitous anonymity
- Ubiquitous identify theft.

The underlying technologies considered for authentication were: biometrics: cryptography, with a focus on digital signatures; secure processing/ computation; and reputation systems. Most of the discussion at the workshop focused on issues related to authentication of users of digital government, but, as the scenario related to ubiquitous identity theft implies, there was also discussion of problems related to misinformation, including cognitive hacking.

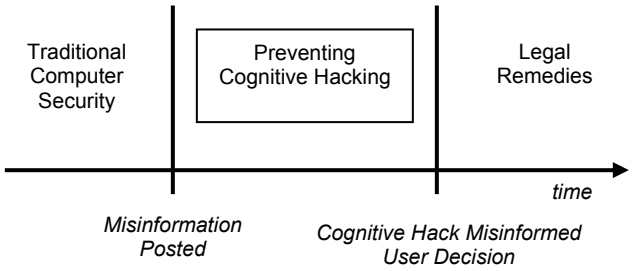
Cognitive hacking refers to a computer or information system attack that relies on changing human users' perceptions and corresponding behaviors in order to succeed. This is in contrast to denial of service (DOS) and other kinds of well-known attacks that operate solely within the computer and network infrastructure. With cognitive attacks neither hardware nor software is necessarily corrupted. There may be no unauthorized access to the computer system or data. Rather the computer system is used to influence people's perceptions and behavior through misinformation. The traditional definition of security is protection of the computer system from three kinds of threats: unauthorized disclosure of information, unauthorized

modification of information, and unauthorized withholding of information (denial of service). Cognitive attacks, which represent serious breaches of security with significant economic implications, are not covered well by this definition. Social engineering via a computer system, i.e., a hacker's psychological tricking of legitimate computer system users to gain information, e.g., passwords, in order to launch an autonomous attack on the system, is a special case of cognitive hacking.

In face to face interaction with other people, there is normally some context in which to evaluate information being conveyed. One associates a level of reliability to information depending on who the speaker is and on what is known of the person. This type of evaluation cannot be transferred to the Web [3]. The Internet's open nature makes it an ideal arena for dissemination of misinformation. The issue is how to deal with false information on the Web and how to decide whether a source is reliable. What happens if a user makes a decision based on information found on the Web that turns out to be misinformation, even if the information appears to come from a government website? In reality, the information might be coming from a spoofed version of a government website. Spoofed websites that are difficult to distinguish from the true website can be readily created [4,5].

Computer and network security present great challenges to our evolving information society and economy [6]. The variety and complexity of cyber security attacks that have been developed parallel the variety and complexity of the information technologies that have been deployed, with no end in sight for either. Two classes of information systems attacks are distinguished: *autonomous* attacks and *cognitive* attacks. Autonomous attacks operate totally within the fabric of the computing and networking infrastructures. For example, files containing private information such as credit card numbers can be downloaded and used by an attacker. Such an attack does not require any intervention by users of the attacked system, hence can be called an "autonomous" attack. By contrast, a *cognitive* attack requires some change in users' behavior, affected by manipulating their perception of reality. The attack's desired outcome cannot be achieved unless human users change their behaviors in some way. Users' modified actions are a critical link in a cognitive attack's sequencing.

Consider the graph below. Most analyses of computer security focus on the time before misinformation is posted, i.e., on preventing unauthorized use of the system. A cognitive hack takes place when a user's behavior is influenced by misinformation. At that point the focus is on detecting that a cognitive hack has occurred and on possible legal action. The concern here is with developing tools to prevent cognitive hacking, that is, tools that can recognize and respond to misinformation before a user acts based on the misinformation.



Ultimately each individual is responsible for his or her use of technology and for decisions taken based on information gathered from the web. The primary concern here is with misinformation that cannot be easily detected. Who is responsible for a large loss incurred resulting from misinformation posted on the Web? Is this simply a matter of "buyer beware," or "government information consumer beware", or can users be protected by technology or policy?

2. DEFINITION OF COGNITIVE HACKING

Cognitive hacking is defined [7] is defined as gaining access to, or breaking into, a computer information system for the purpose of modifying certain behaviors of a human user in a way that violates the integrity of the overall user-information system.

A definition of semantic attacks closely related to this discussion of cognitive hacking has been described by Schneier [9], who attributes the earliest conceptualization of computer system attacks as physical, syntactic, and semantic to Martin Libicki, who describes semantic attacks in terms of misinformation being inserted into interactions among intelligent agents on the Internet [10]. Schneier, by contrast, characterizes semantic attacks as "... attacks that target the way we, as humans, assign meaning to content." He goes on to note, "Semantic attacks directly target the human/computer interface, the most insecure interface on the Internet" [9].

Denning's discussion of information warfare [11] overlaps this concept of cognitive hacking. Denning describes information warfare as a struggle over an information resource by an offensive and a defensive player. The resource has an exchange and an operational value. The value of the resource to each player can differ depending on factors related to each player's circumstances. The outcomes of offensive information warfare are: increased availability of the resource to the offense, decreased availability to the defense, and decreased integrity of the resource.

3. COGNITIVE HACKING COUNTERMEASURES

Given the wide variety of cognitive hacking approaches, *preventing* cognitive hacking reduces either to preventing unauthorized access to information assets (such as in web defacements) in the first place or detecting posted misinformation before user behavior is affected (that is, before behavior is changed but possibly after the misinformation has been disseminated). The latter may not involve unauthorized access to information, as for instance in "pump and dump" schemes that use newsgroups and chat rooms. By definition, *detecting* a successful cognitive hack would involve detecting that the user behavior has already been changed. Detection in that sense is not being considering at this time. Rather, this discussion of methods for preventing cognitive hacking is restricted to approaches that could automatically alert users of problems with their information sources.

In general, there are two classes of cognitive attacks: those where there is a single source of potentially misleading information, and those where there are multiple sources. Furthermore, countermeasures can be mathematical, or linguistic, in nature. Single source situations are those in which redundant, independent sources of information about the same topic are not available. An authoritative corporate personnel database would be an example. Countermeasures for single source cognitive attacks involve due diligence in authenticating the information source and ascertaining its reliability. Various relatively mature certification and Public Key Infrastructures technologies can be used to detect spoofing of an information server. Additionally, reliability metrics can be established for an information server or service by scoring its accuracy over repeated trials and different users. In this spirit, Lynch [12] describes a framework in which trust can be established on an individual user basis based on both the identity of a source of information, through PKI techniques for example, and in the behavior of the source, such as could be determined through rating systems. Such an approach will take time and social or corporate consensus to evolve. Other countermeasures for single source situations include information trajectory modeling and Ulam games [7,8].

In other situations multiple, presumably redundant, sources of information are available about the same subject of interest. This is clearly the case with financial, political, and other types of current event news coverage. Automated software tools could in principle help people make decisions about the veracity of information they obtain from multiple networked information systems.

The problem of detecting misinformation on the Internet is much like that of detecting other forms of misinformation, for example in newsprint or verbal discussion. Reliability, redundancy, pedigree, and authenticity of the information being considered are key indicators of the overall "trustworthiness" of the information. The technologies of collaborative filtering and reputation reporting mechanisms have been receiving more attention recently, especially in the area of on-line retail sales [13]. This is commonly used by the many on-line price comparison services to inform potential customers about vendor reliability. The reliability rating is computed from customer reports. Other countermeasures for situations with multiple sources include Byzantine General models, detection of collusion by information sources, and the linguistic techniques of authorship attribution and genre analysis [7,8].

4. COGNITIVE HACKING AND DIGITAL GOVERNMENT

Cognitive hacking and digital identity issues with respect to digital government include those where the consumer of virtual government is the victim of the attack, e.g., when a government website is spoofed, and those where the victim of the attack is the government itself, and ultimately the public, i.e., in cases of insider misuse. In this later case the government insider, although an authenticated user, is misusing the government resources to which the insider has access.

The cognitive hacking countermeasures described above in their more general use, can be applied directly to cases where the consumer of government information is the victim of the cognitive attack. The insider misuse problem is a difficult open research problem. A trusted insider will try to conceal his/her unauthorized interactions with a computer system. In this case the victim of the cognitive attack is the system, or security, administrator of the government information system. Some of these deceptive interactions might be detected by some of the cognitive hacking countermeasures described above.

5. CONCLUSIONS

This paper has described some of the threats that cognitive hacking represents to digital government and has proposed countermeasures that can be developed to ameliorate the threat. In particular, from the consumer's perspective, problems of authenticating government websites have been considered, and, from the government's perspective, the insider misuse problem has been considered.

6. REFERENCES

[1] http://www.ksg.harvard.edu/digitalcenter/conference/.

[2] http://www.ksg.harvard.edu/digitalcenter/.

[3] L. Zhou, D.P. Twitchell, T. Qin, J.K. Burgoon, and J.F. Nunamaker, "An exploratory study into deception in text-based computer-mediated communications" *Proceedings of the 36th Hawaii International Conference on Systems Science*. 2003.

[4] E.W. Felton, D. Balfanz, D. Dean, and D. Wallach, "Web spoofing: An Internet con game". Technical Report 54-96 (revised) Department of Computer Science, Princeton University, 1997.

[5] Y. Yuan, E.Z. Ye, and S. Smith, "Web spoofing 2001" Department of Computer Science/Institute for Security Technology Studies *Technical Report TR2001-409*, 2001.

[6] G. Cybenko, A. Giani, and P. Thompson, "Cognitive Hacking and the Value of Information" *Workshop on Economics and Information Security*, May 16-17, 2002, Berkeley, California.

[7] G. Cybenko, A. Giani, & P. Thompson, Cognitive Hacking: A Battle for the Mind *IEEE Computer*, *35*(8), 2002, 50-56.

[8] G. Cybenko, A. Giani, and P. Thompson, "Cognitive Hacking" In M. Zelkowitz (ed.) *Advances in Computers* (to appear).

[9] B. Schneier, "Semantic attacks: The third wave of network attacks" *Crypto-gram Newsletter* October 15, 2000. *http://www.counterpane.com/crypto-gram-0010.htm.*

[10] M. Libicki, *The mesh and the Net: Speculations on armed conflict in an age of free silicon* National Defense University McNair Paper 28, 1994. *http://www.ndu.edu/inss/macnair/mcnair28/m028cont.html*.

[11] D. Denning, *Information warfare and security* Reading, Mass.: Addison-Wesley, 1999.

[12] C. Lynch, "When Documents Deceive: Trust and Provenance as New Factors for Information Retrieval in a Tangled Web", *Journal of the American Society for Information Science & Technology*, *52*(1), 2001, 12-17.

[13] C. Dellarocas, "Building trust on-line: The design of reliable reputation reporting mechanisms for online trading communities" *Center for eBusiness@MIT* paper 101.