

## Detecting Deception in the context of Web 2.0

Paul Thompson  
CS Dept. Dartmouth College  
Hanover, NH  
[Paul.Thompson@dartmouth.edu](mailto:Paul.Thompson@dartmouth.edu)

Annarita Giani  
EECS UC Berkeley  
Berkeley, CA  
[agiani@eecs.berkeley.edu](mailto:agiani@eecs.berkeley.edu)

### 1.0 Introduction

Cybenko et al. [1] introduced the concept of cognitive hacking and described several countermeasures for defending against cognitive hacking. Cognitive hacking was defined as a disinformation attack on the mind of the end user of a networked computer system, e.g., a computer connected to the Internet. Cognitive hacking is a type of semantic attack as defined by Libicki, who described computer network attacks as being physical, syntactic, and semantic [6]. Cybenko et al. narrowed Libicki's focus to semantic attacks targeting the mind of a human user, called cognitive attacks. More recently Giani has extended the notion of cognitive hacking to describe cognitive channel attacks [3]. Thompson in subsequent research has emphasized semantic attacks and deception [7,8]. In this paper we use the terminology of semantic attacks and cognitive channel attacks to broaden the scope of Cybenko et al.'s earlier work on detecting pump-and-dump schemes to consider cognitive channel and semantic attacks, and their detection, in the context of Web 2.0 environments.

### 2.0 Cognitive Channel Attacks

A cognitive channel is the communication channel between a person and the information technology used. An attack on a cognitive channel exploits the vulnerabilities between the user, her perception of the information system, and the actual underlying technology. The vulnerabilities are in the gap between the user's mental model of the information system and its actual implementation. The sophistication of modern information systems and their growing presence in human activities has made these channels attractive targets. Traditional computer security protection and attack detection approaches focus mainly on technical vulnerabilities. Cognitive channels are increasingly the weak links in an information system because traditional technical vulnerabilities are being fixed. This has created a significant gap between computer security technology and the threat space. Modern cognitive channel attacks, such as Cognitive Hacking and Phishing, are in fact complex processes that can be *detected* and *tracked*. An effective approach to defending against cognitive channel attacks therefore involves accurate process modeling and the development of new attack models based on processes. Giani [3] and Gregorio-de Souza et al. [4] have identified, implemented and evaluated several approaches based on the Process Query System paradigm for detecting complex multi-stage Phishing attacks. Web2.0 security research should seek to obtain a better understanding of users' mental models, since most of the new attacks succeed because they are able to exploit the fact that what the user thinks is taking place in the human computer interface is not what it is really happening. Research should lead to new systems which reduce this gap between the meaning of the technology and what user thinks. For example, the user might be helped with explicit information about the security and the real status of the system.

### 3.0 Semantic Attacks

Cognitive hacking refers to a computer or information system attack that relies on changing human users' perceptions and corresponding behaviors in order to be successful. 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 disinformation. 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, or denial of service [5]. Semantic attacks, which represent serious breaches of security with significant economic implications, are not well-covered by this definition. In face-to-face interaction with other people, there is normally some context in which to evaluate information being conveyed. Communication in Web 2.0 environments, as compared to earlier Web environments, has more of the flavor of face to face interaction, but there are substantial differences. In face-to-face interactions researchers have shown that there is much deception [2]. There are many cues available to people in face-to-face communication that allow at least some people to reliably detect deception. Deception is also presumably widespread in computer-mediated communication, as typified in Web 2.0 environments. Some of this deception may be innocuous, e.g., analogous to white lies in face-to-face communication, but other deception, e.g., pump-and-dump schemes, is harmful. The semantic hacking project proposed a variety of countermeasures for detecting semantic attacks on the Web. These countermeasures, informed by recent social science research on computer-mediated deception, e.g., Zhou et al. [9] can be adapted for Web 2.0 environments.

#### 4.0 References

- [1] Cybenko, G., Giani, A. and Thompson, P. 2002. "Cognitive Hacking: A Battle for the Mind" *IEEE Computer*, 35(8), 2002, 50-56.
- [2] George, J., Biros, D. P., Burgoon, J. K. and Nunamaker, J. F. Jr. 2003. "Training Professionals to Detect Deception". *NSF / NIJ Symposium on Intelligence and Security Informatics, Lecture Notes in Computer Science*, Berlin: Springer-Verlag, June 1-3, 2003, Tucson, Arizona, 2003, p. 366-370.
- [3] Giani, A. "Detection of Attacks on Cognitive Channels", Ph.D Thesis, Thayer School of Engineering, Dartmouth College, Nov. 2006
- [4] Gregorio-De Souza, I.; Berk V.H., Giani, A.; Bakos, G.; Bates, M.; and Cybenko, G.V. "Detection of Complex Cyber Attacks", In *Proceedings of the SPIE/ Vol. 6201*, Orlando, Florida, April 2006
- [5] Landwehr, C. E. 1981. "Formal models of computer security". *Computing Surveys*, vol. 13, no. 3.
- [6] Libicki, M. 1994. "The mesh and the Net: Speculations on armed conflict in an age of free silicon". National Defense University McNair Paper 28,
- [7] Thompson, P. (to appear 2007). "Cognitive Hacking Five Years Later: Detecting Deception" In Harrington, B. (Ed.), *Deception: What lies beneath* Palo Alto: Stanford University Press.
- [8] Thompson, P. "Deception as a Semantic Attack" In *Adversarial Reasoning: Computational Approaches to Reading the Opponent's Mind* In Kott, A. and W. M. McEneaney, W.M. (Eds.) Boca Raton, Florida: Chapman and Hall, 2006.
- [9] Zhou, L.; Twitchell, D.P.; Qin, T.; Burgoon, J.K.; and Nunamaker, J.F. 2003. "An exploratory study into deception in text-based computer-mediated communications" *Proceedings of the 36<sup>th</sup> Hawaii International Conference on Systems Science*.