

1. INTRODUCTION

Misinformation, however it arises, can have a significant effect on the economic actions of individuals, organizations, and nations. In this chapter we focus on the deliberate use of misinformation to influence the actions of: a) an individual human user of information system technology, b) autonomous agents, and c) corporations and nations in the context of tactical and strategic information warfare. We also discuss automated countermeasures to these attacks. In section 2 we define cognitive and semantic hacking and discuss the related issues of perception management, deception detection, information warfare, and the role of cognitive, or semantic, hacking in intelligence and security informatics. In section 3 we discuss several recent examples of cognitive hacking on the Internet and then give a more detailed discussion of the problems of insider misuse, digital government, and of attacks on the financial infrastructure. In section 4 we discuss cognitive hacking in terms of several economic models, including information-theoretic models of the value of information, the theory of the firm, and more recent theories developed by Borg. In section 5 we discuss a variety of cognitive hacking countermeasures. In section 6 we discuss semantic hacking, a natural extension to the concept of cognitive hacking. Section 7 describes our plans for future work in this area, while section 8 provides a summary and conclusions.

2. BACKGROUND

Computer and network security present great challenges to our evolving information society and economy. 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. Physical and syntactic attacks operate totally within the fabric of the computing and networking infrastructures. For example, the well-know Unicode attack against older, unpatched versions of Microsoft's Internet Information Server (IIS) can lead to root/administrator access. Once such access is obtained, any number of undesired activities by the attacker is possible. 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. By contrast, a *cognitive* attack requires some change in users' behavior, accomplished 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 the sequencing of a cognitive attack.

Cognitive attacks can be overt or covert. No attempt is made to conceal overt cognitive attacks, e. g., website defacements. Provision of misinformation, the intentional distribution or insertion of false or misleading information intended to influence reader's decisions and / or activities, is covert cognitive hacking. The Internet's open nature makes it an ideal arena for dissemination of misinformation. Cognitive hacking differs from social engineering, which, in the computer domain, involves 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.

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. Our concern 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.

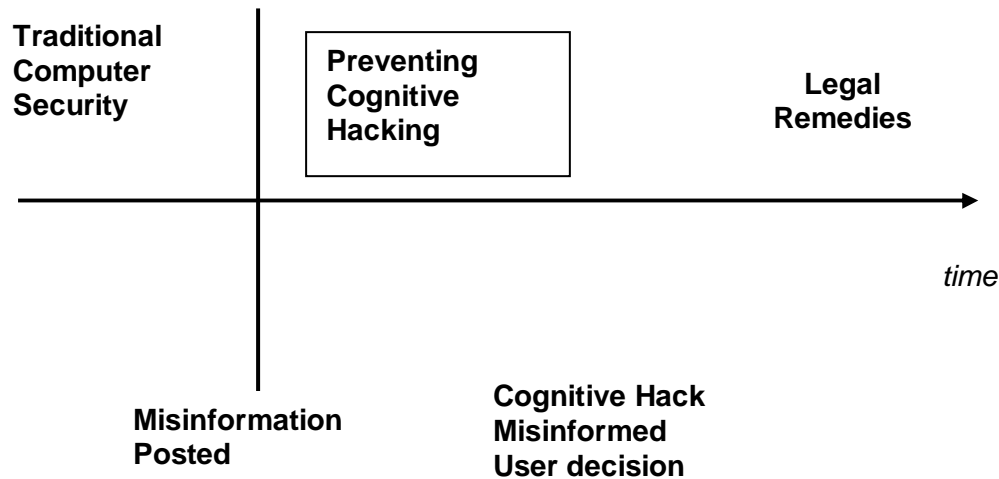


Figure -1. The sequencing of a cognitive attack

By contrast, a *cognitive* attack requires some change in users' behavior, effected by manipulating their perceptions 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. To illustrate what we mean by a cognitive attack, consider the following news report (Mann, 2000):

"Friday morning, just as the trading day began, a shocking company press release from **Emulex** (Nasdaq: EMLX) hit the media waves. The release claimed that Emulex was suffering the corporate version of a nuclear holocaust. It stated that the most recent quarter's earnings would be revised from a \$0.25 per share gain to a \$0.15 loss in order to comply with Generally Accepted Accounting Principles (GAAP), and that net earnings from 1998 and 1999 would also be revised. It also said Emulex's CEO, Paul Folino, had resigned and that the company was under investigation by the Securities and Exchange Commission. Trouble is, none of it was true. The real trouble was that Emulex shares plummeted from their Thursday close of \$113 per share to \$43 -- a rapid 61% haircut that took more than \$2.5 billion off of the company's hide -- before the shares were halted an hour later. The damage had been done: More than 3 million shares had traded hands at the artificially low rates. Emulex vociferously refuted the authenticity of the press release, and by the end of the day the company's shares closed within a few percentage points of where they had opened."

Mark Jacob, 23 years old, fraudulently posted the bogus release on Internet Wire, a Los Angeles press-release distribution firm. The release was picked up by several business news

services and widely redistributed scale without independent verification. The speed, scale and subtlety with which networked information propagates have created a new challenge for society, outside the domain of classical computer security which has traditionally been concerned with ensuring that all use of a computer and network system is authorized.

The use of information to affect the behavior of humans is not new. Language, or more generally communication, is used by one person to influence another. Propaganda has long been used by governments, or by other groups, particularly in time of war, to influence populations (Coombs and Nimmo, 1993; Doob, 1935; Ellul, 1966). Although the message conveyed by propaganda, or other communication intended to influence, may be believed to be true by the propagator, it usually is presented in a distorted manner, so as to have maximum persuasive power, and, often, is deliberately misleading, or untrue. Propaganda is a form of perception management. Other types of perception management include psychological operations in warfare (Information Warfare, 2001), consumer fraud, and advertising (Coombs and Nimmo, 1993; Pratkanis and Aronson, 1992). As described in section 1.5, deception detection has long been a significant area of research in the disciplines of psychology and communications.

2.1 Perception Management

As noted by many authors, e.g. (Coombs and Nimmo, 1993; Denning, 1999; Ellul, 1966; Pratkanis and Aronson, 1992) perception management is pervasive in contemporary society. Its manifestation on the Internet is one aspect of the broader phenomenon. Not all perception management is negative, e.g., education can be considered a form of perception management; nor is all use of perception management on the Internet cognitive hacking (see definition in the next section). Clearly the line between commercial uses of the Internet such as advertising, which would not be considered cognitive hacking, and manipulation of stock prices by the posting of misinformation in news groups, which would be so considered, is a difficult one to distinguish.

Cognitive hacking is defined here 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. The integrity of such a system would for example include correctness or validity of the information the user gets from such a system. In this context, the integrity of a computer system can be defined more broadly than the definition implicit in Landwehr's classic definition of computer security in terms of confidentiality, integrity, and accessibility (Landwehr, 1981). Smith (2001) refers to breaches in computer security as violations of the semantics of the computer system, i.e., the intended operation of the system. Wing (1998) argues a similar view. In this sense the World Wide Web itself can be seen as a computer system used for communication, e-commerce, and so on. As such, activities conducted over the Web that violate the norms of communication or commerce, for example, fraud and propaganda, are considered to be instances of cognitive hacking, even if they do not involve illegitimate access to, or breaking into, a computer. For example, a person might maintain a website that presents misinformation with the intent of influencing viewers of the information to engage in fraudulent commercial transactions with the owner of the website.

2.2 Semantic Attacks and Information Warfare

A definition of semantic attacks closely related to our discussion of cognitive hacking has been described by Schneier (2000), who attributes the earliest conceptualization of computer system attacks as physical, syntactic, and semantic to Martin Libicki (1994), who describes semantic attacks in terms of misinformation being inserted into interactions among intelligent agents on the Internet. Schneier (2000), 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”.

Denning’s (1999) discussion of information warfare overlaps our 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. Applied to the Emulex example, described below, Jakob is the offensive player and Internet Wire and the other newswire services are the defensive players. The outcome is decreased integrity of the newswires’ content. From the perspective of cognitive hacking, while the above analysis would still hold, the main victims of the cognitive hacking would be the investors who were misled. In addition to the decreased integrity of the information, an additional outcome would be the money the investors lost.

2.3 Deception Detection

Deception of detection in interpersonal communication has long been a topic of study in the fields of psychology and communications Buller and Burgoon, 1996; Cornetto, 2001; Cao, Burgoon, and Nunamaker, 2003). The majority of interpersonal communications are found to have involved some level of deception. Psychology and communications researchers have identified many cues that are characteristic of deceptive interpersonal communication. Most of this research has focused on the rich communication medium of face to face communication, but more recently other forms of communication have been studied such as telephone communication and computer-mediated communication (Zhou, Twitchell, Qin, Burgoon, and Nunamaker, 2003). A large study is underway Cao, Burgoon, and Nunamaker, 2003; George, Biros, Burgoon, and Nunamaker, 2003) to train people to detect deception in communication. Some of this training is computer-based. Most recently a study has begun to determine whether psychological cues indicative of deception can be automatically detected in computer-mediated communication, e.g., e-mail, so that an automated deception detection tool might be built (Zhou, Burgoon, and Twitchell, 2003; Zhou, Twitchell, Qin, Burgoon, and Nunamaker, 2003).

2.4 Cognitive Hacking and Intelligence and Security Informatics

Intelligence and security informatics (Chen, Zeng, Schroeder, Miranda, Demchak, and Madhusdan, 2003) will be supported by data mining, visualization, and link analysis technology, but intelligence and security analysts should also be provided with an analysis environment supporting mixed-initiative interaction with both raw and aggregated data sets (Thompson, 2003). Since analysts will need to defend against semantic attacks, this environment should include a toolkit of cognitive hacking countermeasures. For example, if faced with a potentially deceptive news item from FBIS, an automated countermeasure might provide an alert using adaptive fraud detection algorithms (Fawcett and Provost, 2002) or through a retrieval mechanism allow the analyst to quickly assemble and interactively analyze related documents bearing on the potential misinformation. The author is currently developing both of these countermeasures.

Information retrieval, or document retrieval, developed historically to serve the needs of scientists and legal researchers, among others. Despite occasional hoaxes and falsifications of data in these domains, the overwhelming expectation is that documents retrieved are honest representations of attempts to discover scientific truths, or to make a sound legal argument. This

assumption does not hold for intelligence and security informatics. Most information retrieval systems are based either on: a) an exact match Boolean logic by which the system divides the document collection into those documents matching the logic of the request and those that do not, or b) ranked retrieval. With ranked retrieval a score is derived for each document in the collection based on a measure of similarity between the query and the document's representation, as in the vector space model (Salton and McGill, 1983), or based on a probability of relevance (Maron and Kuhns, 1960; Rijsbergen, 1979).

Although not implemented in existing systems, a utility theoretic approach to information retrieval (Cooper and Maron, 1978) shows promise for a theory of intelligence and security informatics. In information retrieval predicting relevance is hard enough. Predicting utility, although harder, would be more useful. When information contained in, say, a FBIS document, may be misinformation, then the notion of utility theoretic retrieval, becomes more important. The provider of the content may have believed the information to be true or false, aside from whether it was true or false in some objective sense. The content may be of great value to the intelligence analyst, whether it is true or false, but, in general, it would be important to know not only whether it was true or false, but also whether the provider believed it to be true or false. Current information retrieval algorithms would not take any of these complexities into account in calculating a probability of relevance.

Predictive modeling using the concepts of cognitive hacking and utility-theoretic information retrieval can be applied in two intelligence and security informatics settings which are mirror images of each other, i.e., the user's model of the system's document content and the systems model of the user as a potential malicious insider. Consider an environment where an intelligence analyst accesses sensitive and classified information from intelligence databases. The accessed information itself may represent cognitive attacks coming from the sources from which it has been gathered, e.g., FBIS documents. As discussed above, each of these documents will have a certain utility for the analyst, based on the analyst's situation, based on whether or not the documents contain misinformation, and, if the documents do contain misinformation, whether, or not, the analyst can determine that the misinformation is present. On the other hand, the analyst might be a malicious insider engaged in espionage. The document system will need to have a cost model for each of its documents and will need to build a model of each user, based on the user's transactions with the document system and other external actions.

Denning's theory of information warfare (1999) and an information theoretic approach to the value of information (Cover and Thomas, 1991) can be used to rank potential risks given the value of each document held by the system. Particular attention should be paid to deception on the part of the trusted insider to evade detection. Modeling the value of information to adversaries will enable prediction of which documents are likely espionage targets and will enable development of hypotheses for opportunistic periods and scenarios for compromise. These models will be able to detect unauthorized activity and to predict the course of a multi-stage attack so as to inform appropriate defensive actions.

Misinformation, or cognitive hacking, plays a much more prominent role in intelligence and security informatics than it has played in traditional scientific informatics. The status of content as information, or misinformation, in turn, influences its utility for users. Cognitive hacking countermeasures are needed to detect and defend against cognitive hacking.

3. EXAMPLES OF COGNITIVE HACKING

This section summarizes several documented examples of cognitive hacking on the Internet and provides a more detailed discussion of the problems of insider misuse and of attacks on the financial infrastructure.

3.1 Internet Examples

3.1.1 NEI Webworld case.

In November 1999 two UCLA graduates students and one of their associates purchased almost all of the shares of the bankrupt company NEI Webworld at a price ranging from 0.05 to 0.17 per share. They opened many Internet message board accounts using a computer at the UCLA BioMedical Library and posted more than 500 messages on hot web sites to pump up the stock of the company, stating false information about the company with the purpose of convincing others to buy stock in the company. They claimed that the company was being taken over and that the target price per share was between 5 and 10 dollars. Using other accounts they also pretended to be an imaginary third party, a wireless telecommunications company, interested in acquiring NEI Webworld. What the three men did not post was the fact that NEI was bankrupt and had liquidated assets in May 1999. The stock price rose from \$0.13 to \$15 in less than one day, and they realized about \$364,000 in profits. The men were accused of selling their shares incrementally, setting target prices along the way as the stock rose. On one day the stock opened at \$8 and soared to \$15 5/16 a share by 9:45 a.m. ET and by 10:14 a.m. ET, when the men no longer had any shares, the stock was worth a mere 25 cents a share.

On Wednesday, December 15, 1999, the U.S. Securities and Exchange Commission (SEC) and the United States Attorney for the Central District of California charged the three men with manipulating the price of NEI Webworld, Inc. In late January 2001, two of them, agreed to give up their illegal trading profits (approximately \$211,000). The Commission also filed a new action naming a fourth individual, as participating in the NEI Webworld and other Internet manipulations. Two of the men were sentenced on January 22, 2001 to 15 months incarceration and 10 months in a community corrections center. In addition to the incarcerations, Judge Fees ordered the men to pay restitution of between \$566,000 and \$724,000. The judge was to hold a hearing on Feb. 26 to set a specific figure. Anyone with access to a computer can use as many screen names as desired to spread rumors in an effort to pump up stock prices by posting false information about a particular company so that they can dump their own shares and give the impression that their own action has been above board.

3.1.2 The Jonathan Lebed case

A 15 years old student using only AOL accounts with several fictitious names was able to change the behavior of many people around the world making them act to his advantage (Lewis, 2001a). In six months he gained between \$12,000 and \$74,000 daily each time he posted his messages and, according to the US Security Exchange Commission, he did that 11 times increasing the daily trading volume from 60,000 shares to more than a million. His messages sounded similar to the following one (Lewis, 2001b):

DATE: 2/03/00 3:43pm Pacific Standard Time

FROM: LebedTG1

FTEC is starting to break out! Next week, this thing will EXPLODE . . .

Currently FTEC is trading for just \$21/2. I am expecting to see FTEC at \$20

VERYSOON . . .

Let me explain why . . .

Revenues for the year should very conservatively be around \$20 million. The average company in the industry trades with a price/sales ratio of 3.45. With 1.57 million shares outstanding, this will value FTEC at . . . \$44. It is very possible that FTEC will see \$44, but since I would like to remain very conservative . . . my short term price target on FTEC is still \$20! The FTEC offices are extremely busy . . . I am hearing that a number of HUGE deals are being worked on. Once we get some news from FTEC and the word gets out about the company . . . it will take-off to MUCH HIGHER LEVELS! I see little risk when purchasing FTEC at these DIRT-CHEAP PRICES. FTEC is making TREMENDOUS PROFITS and is trading UNDER BOOK VALUE!!! This is the #1 INDUSTRY you can POSSIBLY be in RIGHT NOW. There are thousands of schools nationwide who need FTEC to install security systems . . . You can't find a better positioned company than FTEC! These prices are GROUND-FLOOR! My prediction is that this will be the #1 performing stock on the NASDAQ in 2000. I am loading up with all of the shares of FTEC I possibly can before it makes a run to \$20. Be sure to take the time to do your research on FTEC! You will probably never come across an opportunity this HUGE ever again in your entire life.

He sent this kind of message after having bought a block of stocks. The purpose was to influence people and let them behave to pump up the price by recommending the stock. The messages looked credible and people did not even think to investigate the source of the messages before making decisions about their money. Jonathan gained \$800,000 in six months. Initially the SEC forced him to give up everything, but he fought the ruling and was able to keep part of what he gained. The question is whether he did something wrong, in which case the SEC should have kept everything. The fact that the SEC allowed Jonathan to keep a certain amount of money shows that it is not clear whether or not the teenager is guilty from a legal perspective. Certainly, he made people believe that the same message was post by 200 different people.

Richard Walker, the SEC's director of enforcement, referring to similar cases, stated that on the Internet there is no clearly defined border between reliable and unreliable information, investors must exercise extreme caution when they receive investment pitches online.

3.1.3 Fast-Trade.com website pump and dump

In February and March 1999, Douglas Colt, a Georgetown University law student, manipulated four traded stocks using the web site Fast-trade.com. Together with a group of friends he posted hundreds of false or misleading messages on Internet message boards such as Yahoo! Finance Raging Bull with the purpose of encouraging people to follow Fast-trade.com advice. The site offered a trial subscription and in less then two months more than 9,000 users signed up. The group was able to gain more than \$345,000.

3.1.4 PayPal.com

"We regret to inform you that your username and password have been lost in our database. To help resolve this matter, we request that you supply your login information at the following website."

Many customers of PayPal received this kind of email and subsequently gave personal information about their PayPal account to the site linked by the message (<http://paypalsecure.com> not <http://www.paypal.com>) (Krebs, 2001). The alleged perpetrators apparently used their access to PayPal accounts in order to purchase items on eBay.

3.1.5 Emulex Corporation

Mark S. Jakob, after having sold 3,000 shares of Emulex Corporation in a "short sale" at prices of \$72 and \$92, realized that, since the price rose to \$100, he lost almost \$100,000 (Mann, 2000). This kind of speculation is realized by borrowing shares from a broker and selling them in hope that the price will fall. Once this happens, the shares are purchased back and the stock is returned to the broker with the short seller keeping the difference.

On August 25th 2000, when he realized the loss, he decided to do something against the company. The easiest and most effective action was to send a false press release to Internet Wire Inc. with the goal of influencing the stock price. He claimed that Emulex Corporation was being investigated by the Security and Exchange Commission (SEC) and that the company was forced to restate 1998 and 1999 earnings. The story quickly spread, and half an hour later other news services such as Dow Jones, Bloomberg and CBS Marketwatch picked up the hoax. Due to this false information, in a few hours Emulex Corporation lost over \$2 billion dollars. After sending misinformation about the company, Jakob executed trades so that he earned \$236,000. Jakob was arrested and charged with disseminating a false press release and with security fraud. He is subject to a maximum of 25 years in prison, a maximum fine of \$220 million, two times investor losses, and an order of restitution up to \$110 million to the victims of his action.

3.1.6 Non-financial fraud - web search engine optimization

Con artists have defrauded consumers for many years over the telephone and via other means of communication, including direct personal interaction. Such financially-motivated fraud continues over the Internet, as described above. Some cognitive hacking uses misinformation in a fraudulent way that does not directly attack the end user.

One such use of misinformation is a practice (Lynch 2001) that has been called "search engine optimization", or "index spamming". Because many users of the Internet find pages through use of web search engines, owners of web sites seek to trick web search engines to rank their sites more highly when searched by web search engines. Many techniques, for example inaccurate metadata, printing white text on white background (invisible to a viewer of the page, but not to a search engine) are used. While this practice does not directly extort money from a user, it does prevent the user from seeing the search results that the user's search would have returned based on the content of the web site. Thus the primary attack is on the search engine, but the ultimate target of the attack is the end user. Developers at web search engines are aware of this practice by web site promoters and attempt to defeat it, but it is an on-going skirmish between the two camps.

3.1.7 Non-financial fraud - CartoonNetwork.com

Another common misinformation practice is to register misleading web site names, e.g., a name that might be expected to belong to a known company, or a close variant of it, such as a slight misspelling. In October 2001, the FTC (Washtech.com, 2001) sought to close thousands of web sites that allegedly trap web users after they go to a site with a misleading name. According to the FTC, John Zuccarini registered slight misspelling of hundreds of popular Internet domain names. When a user goes to one of these sites a series of windows advertising various products opens rapidly, despite user attempts to back out of the original site. Zuccarini allegedly made \$800,000 to \$1,000,000 annually in advertising fees for such attacks.

3.1.8 Bogus virus patch report

Although computer viruses are syntactic attacks, they can be spread through cognitive attacks. The W32/Redesi-B virus (Sophos, 2001) is a worm which is spread through Microsoft Outlook. The worm is contained in an e-mail message that comes with a subject chosen randomly from 10 possible subjects, e.g., "FW: Security Update by Microsoft". The text of the e-mail reads "Just received this in my email I have contacted Microsoft and they say it's real" and then provides a forwarded message describing a new e-mail spread virus for which Microsoft has released a security patch which is to be applied by executing the attached file. The attached file is the virus. Thus a virus is spread by tricking the user into taking action thought to prevent the spread of a virus.

3.1.9 Usenet perception management

Since the Internet is an open system where everybody can put his or her opinion and data, it is easy to make this kind of attack. Each user is able to influence the whole system or only a part of it in many different ways, for example by building a personal web site or signing up for a Newsgroup. Blocking the complete freedom to do these activities, or even checking what people post on the web, goes against the current philosophy of the system. For this reason technologies for preventing, detecting and recovering from this kind of attack are difficult to implement (Chez.com, 1997).

3.1.10 Political Web site defacements – Ariel Sharon site

Web site defacements are usually overt cognitive attacks. For example, in January 2001, during an Israeli election campaign, the web site of Likud leader Ariel Sharon was attacked (BBC News Online, 2001b). In this attack, and in the retaliatory attack described in example 11, no attempt was made to deceive viewers into thinking that the real site was being viewed. Rather the real site was replaced by another site with an opposing message. The Sharon site had included a service for viewers that allowed them to determine the location of their voting stations. The replacement site had slogans opposing Sharon and praising Palestinians. It also had a feature directing viewers to Hezbollah "polling stations".

3.1.11 Political Web site Defacements – Hamas site

Following the January attack on the Sharon web site, the web site of the militant group Hamas was attacked in March 2001 (BBC News Online, 2001a). When the Hamas website was hacked, viewers were redirected to a hard-core pornography site.

3.1.12 New York Times site

In February 2001 the New York Times web site was defaced by a hacker identified as “splurge” from a group called “Sm0ked Crew”, which had a few days previously defaced sites belonging to Hewlett-Packard, Compaq, and Intel (Register, 2001a; Register, 2001b). The New York Times defacement included html, a .MID audio file, and graphics. The message stated, among other things, “Well, admin I’m sorry to say by you have just got sm0ked by splurge. Don’t be scared though, everything will be all right, first fire your current security advisor . . .” Rather than being politically motivated, such defacements as these appear to be motivated by self-aggrandizement.

3.1.13 Yahoo site

In September of 2001 Yahoo’s news web site was edited by a hacker (MSNBC, 2001). This cognitive hacking episode, unlike the defacements discussed above, was more subtle. While not as covert as hacking with the intent to engage in fraud or perception management, neither were the changes made to the website as obvious as those of a typical defacement. A 20-year old researcher confessed that he altered a Reuters news article about Dmitry Sklyarov, a hacker facing criminal charges. The altered story stated that Sklyarov was facing the death penalty and attributed a false quote to President Bush with respect to the trial.

3.1.14 Web site defacements since 11 September terrorist incident

Since the 11 September terrorist incident, there have been numerous examples of web site defacements directed against web sites related to Afghanistan (Latimes.com, 2001). While official Taliban sites have been defaced, often sites in any way linked with Afghanistan were defaced indiscriminately, regardless of which sides they represented in the conflict.

3.1.15 Fluffi Bunni declares Jihad

Another type of politically motivated cognitive hacking attack has been perpetrated by “Fluffi Bunni”, who has redirected numerous websites to a page in which Bunni’s opinion on current events is presented. This redirection appears to have been accomplished through a hacking of the Domain Name System Server of NetNames (Hacktivist, 2001).

3.1.16 Web site spoofing - CNN site

On 7 October 2001, the day that the military campaign against Afghanistan began, the top-ranked news story on CNN’s most popular list was a hoax, “Singer Britney Spears Killed in Car Accident”. The chain of events which led to this listing started with a web site spoofing of CNN.com (Newsbytes, 2001). Then, due to a bug in CNN’s software, when people at the spoofed site clicked on the “E-mail This” link, the real CNN system distributed a real CNN e-mail to recipients with a link to the spoofed page. At the same time with each click on “E-mail This” at the bogus site, the real site’s tally of most popular stories was incremented for the bogus story. Allegedly this hoax was started by a researcher who sent the spoofed story to three users of AOL’s Instant Messenger chat software. Within 12 hours more than 150,000 people had viewed the spoofed page.

In 1997 Felton and his colleagues showed that very realistic website spoofings could be readily made. More recently, Yuan, Ye, and Smith (2001) showed that these types of website spoofs could be done just as easily with more contemporary web technologies.

3.1.17 Web site spoofing - WTO site

Use of misleading domain names can also be political and more covert. Since 1999, a site, www.gatt.org, has existed which is a parody of the World Trade Organization site, www.wto.org (NetworkWorldFusion, 2001). Again, as in the case of the spoofing of the Yahoo new site mentioned above, the parody can be seen through fairly easily, but still could mislead some viewers.

3.2 Insider Threat

Trusted insiders who have historically caused the most damage to national security were caught only after prolonged counterintelligence operations. These insiders carried out their illegal activities for many years without raising suspicion. Even when it was evident that an insider was misusing information, and even when attention began to focus on the insider in question as a suspect, it took more years before the insider was caught. Traditionally apprehension of trusted insiders has been possible only after events in the outside world had taken place, e.g., a high rate of double agents being apprehended and executed that led to an analysis eventually focusing on the insider. Once it was clear that there was likely a problem with insider misuse of information, it was eventually possible to determine the identity of the insider by considering who had access to the information and by considering other factors such as results of polygraph tests.

The insider threat, is much more pervasive, however, than a small number of high profile national security cases. It has been estimated that the majority of all computer security breeches are due to insider attacks, rather than to external hacking (Anderson et al., 2000).

As organizations move to more and more automated information processing environments, it becomes potentially possible to detect signs of insider misuse much earlier than has previously been possible. Information systems can be instrumented to record all uses of the system, down to the monitoring of individual keystrokes and mouse movements. Commercial organizations have made use of such clickstream mining, as well as analysis of transactions to build profiles of individual users. Credit card companies build models of individuals' purchase patterns to detect fraudulent usage. Companies such as Amazon.com analyze purchase behavior of individual users to make recommendations for the purchase of additional products, likely to match the individual user's profile.

A technologically adept insider, however, may be aware of countermeasures deployed against him, or her, and operate in such a way as to neutralize the countermeasures. In other words, an insider can engage in cognitive hacking against the network and system administrators. A similar situation arises with Web search engines, where what has been referred to as a cold war exists between Web search engines and search engine optimizers, i.e., marketers who manipulate Web search engine rankings on behalf of their clients.

Models of insiders can be built based on: a) known past examples of insider misuse, b) the insider's work role in the organization, c) the insider's transactions with the information system, and d) the content of the insider's work product. This approach to the analysis of the behavior of the insider is analogous to that suggested for analyzing the behavior of software programs by Munson and Wimer (2001). One aspect of this approach is to look for known signatures of insider misuse, or for anomalies in each of the behavioral models individually. Another aspect is

to look for discrepancies among the models. For example, if an insider is disguising the true intent of his, or her, transactions by making deceptive transactions that disguise the true nature of what the insider is doing, then this cognitive hacking might be uncovered by comparing the transactions to the other models described above, e.g., to the insider's work product.

User models have long been of interest to researchers in artificial intelligence and in information retrieval (Rich, 1983; Daniels, Brooks, and Daniels, 1997). Several on-going research programs have been actively involved in user modeling for information retrieval. The Language Modeling approach to probabilistic information retrieval has begun to consider query (user) models (Lafferty and Chengziang, 2001). The Haystack project at MIT is building models of users based on their interactions with a document retrieval system and the user's collections of documents. The current focus of this project, however, seems to be more on overall system architecture issues, rather than on user modeling as such (Huynh, Karger, and Quan, 2003).

The current type of user modeling that might provide the best basis for cognitive hacking countermeasures is recommender system technology (Varian, 1996, 1997; Hofmann, 2001). One of the themes of the recommender systems workshop held at the 1999 SIGIR conference (Herlocker, 2001) was the concern to make recommender systems applicable to problems of more importance than selling products. Since then, recommender systems technology has developed, but applications are generally still largely commercial. Researchers are concerned with developing techniques that work well with sparse amounts of data (Drineas, Kerendis, and Raghavan, 2002) and with scaling up to searching tens of millions of potential neighbors, as opposed to the tens of thousands of today's commercial systems (Sarwar, Karypis, Konstan, and Riedl, 2001). Related to this type of user modeling, Anderson and Khattak (1998) described preliminary results with the use of an information retrieval system to query an indexed audit trail database, but this work was never completed (Anderson, 2002).

3.3 Financial Infrastructure and Cognitive Hacking

3.4 Digital Government and Cognitive Hacking

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 (National Center for Digital Government, 2003a)

A digital government workshop held in 2003 (National Center for Digital Government, 2003b), 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 consideration of problems related to misinformation, including cognitive hacking.

In the face to face interaction with other people associated with traditional provision of government services, there is normally some context in which to evaluate the reliability of information being conveyed. As we have seen, this type of evaluation cannot be directly transferred to digital government. The Internet's open nature makes it an ideal arena for dissemination of misinformation. 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. Furthermore, the insider threat is a serious concern for digital government.

4. VALUE OF INFORMATION – INFORMATION THEORETIC AND ECONOMIC MODELS

Information theory has been used to analyze the value of information in horse races and in optimal portfolio strategies for the stock market (Cover and Thomas 1991). We have begun to investigate the applicability of this analysis to cognitive hacking. So far we have considered the simplest case, that of a horse race with two horses.

4.1 An Information Theoretic Model of Cognitive Hacking

Sophisticated hackers can use information theoretic models of a system to define a gain function and conduct a sensitivity analysis of its parameters. The idea is to identify and target the most sensitive variables of the system, since even slight alterations of their value may influence people's behavior. For example, specific information on the health of a company might help stock brokers predict fluctuations in the value of its shares. A cognitive hacker manipulates the victim's perception of the likelihood of winning a high payoff in a game. Once the victim has decided to play, the cognitive hacker influences which strategy the victim chooses.

4.1.1 A Horse Race

Here is a simple model illustrating this kind of exploit. A horse race is a system defined by the following elements (Cover and Thomas 1991)

- There are m horses running in a race
- each horse i is assigned a probability p_i of winning the race (so $\{p_i\}_{i=1..m}$ is a probability distribution)
- each horse i is assigned an odds o_i signifying that a gambler that bet b_i dollars on horse i would win $b_i o_i$ dollars in case of victory (and suffer a total loss in case of defeat).

If we consider a sequence of n independent races, it can be shown that the average rate of the wealth gained at each race is given by

$$W(b, p, o) = \sum_{i=1}^m p_i \log b_i o_i$$

where b_i is the percentage of the available wealth invested on horse i at each race. So the betting strategy that maximizes the total wealth gained is obtained by solving the following optimization problem

$$W(p, o) = \max W(b, p, o) = \max \sum_{i=1}^m p_i \log b_i o_i$$

subject to the constraint that the b_i 's add up to 1. It can be shown that this solution turns out to be simply $b_i = p_i$ (proportional betting) and so $W(p, o) = \sum p_i \log p_i o_i$.

Thus, a hacker can predict the strategy of a systematic gambler and make an attack with the goal of deluding the gambler on his/her future gains. For example, a hacker might lure an indecisive gambler to invest money on false prospects. In this case it would be useful to understand how sensitive the function W is to p and o and tamper with the data in order to convince a gambler that it is worth playing (because W appears illusionary larger than it actually is).

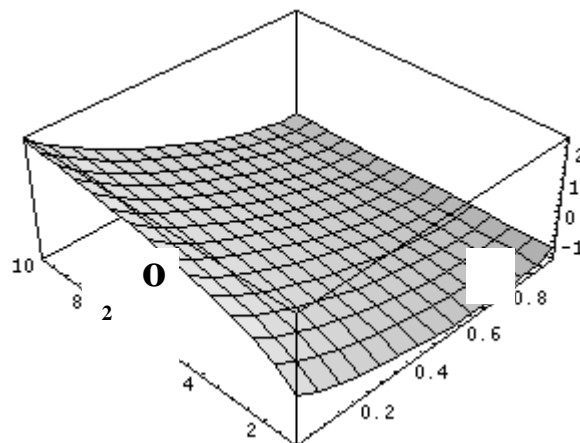
To study the sensitivity of W to its domain variables we consider the partial derivatives of W with respect to p_i and o_i and see where they assume the highest values. This gives us information on how steep the function W is on subsets of its domain.

If we consider the special case of races involving only two horses ($m=2$), then we have

- $\frac{\partial W}{\partial p}(p, o_1, o_2) = \log \left(\frac{p o_1}{1-p o_2} \right)$
- $\frac{\partial W}{\partial o_1}(p, o_1, o_2) = \frac{p}{1-p}$
- $\frac{\partial W}{\partial o_2}(p, o_1, o_2) = \frac{1-p}{o_2}$

Thus, if we fix one of the variables then we can conduct a graphic analysis of those functions with a 3D plot.

Case 1. o_1 is constant.



$$W(p, 3, o_2)$$

This is the doubling rate function. The most sensitive parameter to let W increase is o_2 . Increasing this variable W grows at a fast rate for low values of p and grows with a smaller rate for higher values of p .

4.1.2 Applying the Horse Race Example to the Internet

Misinformation was used by Mark Jakob in a cognitive attack (Mann, 2002). Jakob posted a bogus release regarding the company Emulex on Internet Wire, a Los Angeles press-release distribution firm. The release was picked up by several business news services and widely redistributed without independent verification. Jakob sold Emulex short and profited, while other investors lost large sums of money selling the stock as its value fell sharply in response to the misinformation.

In this example the two horses are: horse 1, Emulex stock goes up; and horse 2, Emulex stock goes down. First the cognitive hacker makes the victim want to play the game by making the victim think that he can make a large profit through Emulex stock transactions. This is done by spreading misinformation about Emulex, whether positive or negative, but news that, if true would likely cause the stock's value to either sharply increase, or decrease, respectively. Positive misinformation might be the news that Emulex had just been granted a patent that could lead to a cure for AIDS. Negative misinformation might be that Emulex was being investigated by the

Securities and Exchange Commission (SEC) and that the company was forced to restate 1998 and 1999 earnings. This fraudulent negative information was in fact posted by Jakob.

4.2 Theories of the Firm and Cognitive Hacking

Much attention in economics has been devoted to theories of the market. The economic actor has been modeled as enjoying perfect, costless information. Such analyses, however, are not adequate to explain the operation of firms. Theories of the firm provide a complementary economic analysis taking into account transaction and organizing costs, hierarchies, and other factors left out of idealized market models. It has been argued that information technology will transform the firm, such that “. . . the fundamental building blocks of the new economy will one day be ‘virtual firms’, ever-changing networks of subcontractors and freelancers, managed by a core of people with a good idea” (Economist, 2002). Others argue that more efficient information flows not only lower transaction costs, thereby encouraging more outsourcing, but also lower organization costs, thereby encouraging the growth of larger companies (Agre, 2001). More efficient information flow implies a more standardized, automated processing of information, which is susceptible to cognitive attack. Schneier (2000) 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 software agents (1994). Libicki was describing information warfare, but semantic, or cognitive, attacks can be directed against business systems, as well.

4.3 Borgian Theory and Cognitive Hacking

5. COGNITIVE HACKING COUNTERMEASURES

Cognitive hacking on the internet is an evolving and growing activity, often criminal and prosecutable. Technologies for preventing, detecting and prosecuting cognitive hacking are still in their infancies. Given the variety of approaches to and the very nature of cognitive hacking, *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. We are not considering detection in that sense at this time.

Our discussion of methods for preventing cognitive hacking will be restricted to approaches that could automatically alert users of problems with their information source or sources (information on a web page, newsgroup, chat room and so on). Techniques for preventing unauthorized access to information assets fall under the general category of computer and network security and will not be considered here. Similarly, detecting that users have already modified their behaviors as a result of the misinformation, namely that a cognitive hack has been successful, can be reduced to detecting misinformation and correlating it with user behavior.

The cognitive hacking countermeasures discussed here will be primarily mathematical and linguistic in nature. The use of linguistic techniques in computer security has been pioneered by Raskin and colleagues at Purdue University's Center for Education and Research in Information Assurance and Security (Atallah, McDonough, Raskin, and Nirenburg, 2001). Their work, however, has not addressed cognitive hacking countermeasures.

5.1 Single Source Cognitive Hacking

In this section, we develop a few possible approaches for the single source problem. By single source, we mean situations in which redundant, independent sources of information about the same topic are not available. An authoritative corporate personnel database would be an example.

5.1.1 Authentication of Source

This technique involves due diligence in authenticating the information source and ascertaining its reliability. Various relatively mature certification and PKI 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 (2001) 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.

5.1.2 Information "Trajectory" Modeling

This approach requires building a model of a source based on statistical historical data or some sort of analytic understanding of how the information relates to the real world. For example, weather data coming from a single source (website or environmental sensor) could be calibrated against historical database (from previous years) or predictive model (extrapolating from previous measurements). A large deviation would give reason for hesitation before committing to a behavior or response.

As an interesting aside, consider the story lines of many well-scripted mystery novels or films. We believe that the most satisfying and successful stories involve a sequence of small deviations from what is expected. Each twist in the story is believable but when aggregated, the reader or viewer has reached a conclusion quite far from the truth. In the context of cognitive hacking, this is achieved by making a sequence of small deviations from the truth, not one of which fails a credibility test on its own. The accumulated deviations are however significant and surprise the reader or viewer who was not paying much attention to the small deviations one by one. However, a small number of major "leaps of faith" would be noticed and such stories are typically not very satisfying. Modeling information sources is something that can be done on a case-by-case basis as determined by the availability of historical data and the suitability of analytic modeling.

5.1.3 Ulam Games

Stanislaw Ulam (1991), in his autobiography *Adventure of a Mathematician* posed the following question

“Someone thinks of a number between one and one million (which is just less than 2^{20}). Another person is allowed to ask up to twenty questions, to which the first person is supposed to answer only yes or no. Obviously, the number can be guessed by asking first: "Is the number in the first half-million?" and then again reduce the reservoir of numbers in the next question by one-half, and so on. Finally, the number is obtained in less than $\log_2(1000000)$.”

Now suppose one were allowed to lie once or twice, then how many questions would one need to get the right answer?"

Of course, if an unbounded number of lies are allowed, no finite number of questions can determine the truth. On the other hand, if say k lies are allowed, each binary search question can be repeatedly asked $2k + 1$ times which is easily seen to be extremely inefficient. Several researchers have investigated this problem, using ideas from error-correcting codes and other areas (Mundici and Trombetta, 1997).

This framework involves a sequence of questions and a bounded number of lies, known a priori. For these reasons, we suspect that this kind of model and solution approach may not be useful in dealing with the kinds of cognitive hacking we have documented, although it will clearly be useful in cognitive hacking applications that involve a sequence of interactions between a user and an information service, as in a negotiation or multi-stage handshake protocol.

5.1.4 Linguistic Countermeasures with Single Sources

5.1.4.1 Genre Detection and Authority Analysis

A careful human reader of some types of misinformation, e.g., exaggerated pump-and-dump scheme postings on the Web about a company's expected stock performance, can often detect the misinforming posting from other legitimate postings, even if these legitimate postings are also somewhat hyperbolic. Since Mosteller and Wallace's (1964) seminal work on authorship attribution, statistical linguistics approaches have been used to recognize the style of different writings. In Mosteller and Wallace's work this stylistic analysis was done to determine the true author of anonymous Federalist papers, where the authorship was disputed. Since then Biber and others (Biber, 1986, 1995; Karlgren and Cutting, 1994) have analyzed the register and genre of linguistic corpora using similar stylistic analysis. Kessler, Nunberg, and Schultze (1997) have developed and tested algorithms based on this work to automatically detect the genre of text.

5.1.5 Psychological Deception Cues

The approach to genre analysis taken, e.g., by Biber and Kessler et al., is within the framework of corpus linguistics, i.e., based on a statistical analysis of general word usage in large bodies of text. The work on deception detection in the psychology and communications fields (see section 2.3) is based on a more fine-grained analysis of linguistic features, or cues. Psychological experiments have been conducted to determine which cues are indicative of deception. To date this work has not led to the development of software tools to automatically detect deception in computer-mediated communication, but researchers see the development of such tools as one of the next steps in this line of research (Zhou, Twitchell, Qin, Burgoon, and Nunamaker, 2003).

5.2 Multiple Source Cognitive Hacking

In this section, we discuss possible approaches to preventing cognitive hacking when 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.

Several aspects of information dissemination through digital, network media, such as the Internet and World Wide Web, make cognitive hacking possible and in fact relatively easy to perform. Obviously, there are enormous market pressures on the news media and on newsgroups to quickly disseminate as much information as possible. In the area of financial news, in particular, competing news services strive to be the first to give reliable news about breaking

stories that impact the business environment. Such pressures are at odds with the time consuming process of verifying accuracy. A compromise between the need to quickly disseminate information and the need to investigate its accuracy is not easy to achieve in general.

Automated software tools could in principle help people make decisions about the veracity of information they obtain from multiple networked information systems. A discussion of such tools, which could operate at high speeds compared with human analysis, follows.

5.2.1 Source Reliability via Collaborative Filtering & Reliability Reporting

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 (Yahalom, Klein, and Beth; Dellarocas, 2001). 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. Another technology, closely related to reliability reporting is collaborative filtering (Thornton, 2001). This can be useful in cognitive hacking situations that involve opinions rather than hard objective facts.

Both of these approaches involve user feedback about information that they receive from a particular information service, building up a community notion of reliability and usefulness of a resource. The automation in this case is in the processing of the user feedback, not the evaluation of the actual information itself.

5.2.1.1 An Example of a Multiple Source Collaborative Filtering Model for Multiple News Sources

Consider the following scenario. An end user is examining a posting to the business section of Google News (Google News beta, 2003). The document purports to provide valuable news about a publicly traded company that the user would like to act on quickly by purchasing, or selling stock. Although this news item might be reliable, it might also be misinformation being fed to unwary users by a cognitive hacker as part of a pump-and-dump scheme, i.e., a cognitive hacker’s hyping of a company by the spread of false, or misleading information about the company and the hacker’s subsequent selling of the stock as the price of its shares rise, due to the misinformation. The end user would like to act quickly to optimize his or her gains, but could pay a heavy price, if this quick action is taken based on misinformation.

News Verifier, a prototype cognitive hacking countermeasure, allows an end user to effectively retrieve and analyze documents from the Web that are similar to the original news item. When the end user receives a news item that he, or she, suspects, may represent a cognitive attack, i.e., contain deliberate misinformation, the user can run the News Verifier. First, a query is automatically generated from the text of the news item. This query is then sent automatically to an API for Google News. Then, a set of documents is retrieved by the Google News clustering algorithm. The Google News ranking of the clustered documents is generic, not necessarily optimized as a countermeasure for cognitive attacks. News Verifier uses a combination process in which several different search engines are used to provide alternative rankings of the documents initially retrieved by Google News. The ranked lists from each of these search engines, along with the original ranking from Google News, are combined using the Combination of Expert Opinion algorithm (Mateescu, Sosonkina, and Thompson, 2002) to provide a more optimal ranking. Relevance feedback judgments from the end user are used to train the

constituent search engines. It is expected that this combination and training process will yield a better ranking than the initial Google News ranking. This is an important feature in a countermeasure for cognitive hacking, because a victim of cognitive hacking will want to detect misinformation as soon as possible in real time.

5.2.2 Byzantine Generals Models

Chandy and Misra (1988) define the Byzantine General's Problem as follows:

A message-communicating system has two kinds of processes, *reliable* and *unreliable*. There is a process, called *general*, that may or may not be reliable. Each process x has a local variable $byz[x]$. It is required to design an algorithm, to be followed by all reliable processes, such that every reliable process x eventually sets its local variable $byz[x]$, to a common value. Furthermore, if *general* is reliable, this common value is $d0[g]$, the initial value of one of *generals* variables. The solution is complicated by the fact that unreliable processes send arbitrary messages. Since reliable processes cannot be distinguished from the unreliable ones, the straightforward algorithm--*general* transmits its initial value to all processes and every reliable process u assigns this value to $byz[u]$ --does not work, because *general* itself may be unreliable, and hence may transmit different values to different processes.

This problem models a group of generals plotting a coup. Some generals are reliable and intend to go through with the conspiracy while others are feigning support and in fact will support the incumbent ruler when the action starts. The problem is to determine which generals are reliable and which are not.

Just as with the Ulam game model for a single information source, this model assumes a sequence of interactions according to a protocol, something that is not presently applicable to the cognitive hacking examples we have considered, although this model is clearly relevant to the more sophisticated information sources that might arise in the future.

5.2.3 Detection of Collusion by Information Sources

Collusion between multiple information sources can take several forms. In pump and dump schemes, a group may hatch a scheme and agree to post misleading stories on several websites and newsgroups. In this case, several people are posting information that will have common facts or opinions, typically in contradiction to the consensus.

Automated tools for preventing this form of cognitive hack would require natural language processing to extract the meaning of the various available information sources and then compare their statistical distributions in some way. For example, in stock market discussion groups, a tool would try to estimate the "position" of a poster, from "strong buy" to "strong sell" and a variety of gradations in between. Some sort of averaging or weighting could be applied to the various positions to determine a "mean" or expected value, flagging large deviations from that expected value as suspicious.

Similarly, the tool could look for tightly clustered groups of messages, which would suggest some form of collusion. Such a group might be posted by the one person or by a group in collusion, having agreed to the form of a cognitive hack beforehand.

Interestingly, there are many statistical tests for detecting outliers but much less is known about detecting collusion which may not be manifest in outliers but in unlikely clusters that may not be outliers at all. For example, if too many eyewitnesses agree to very specific details of a suspect's appearance (height, weight, and so on), this might suggest collusion to an investigator.

For some interesting technology dealing with corporate insider threats due to collusion, see (SRD, 2003).

Automated software tools that can do natural language analysis of multiple documents, extract some quantitative representation of a "position" based on that document and then perform some sort of statistical analysis of the representations are in principle possible, but we are not aware of any efforts working at developing such a capability at this time.

5.2.4 Linguistic Countermeasures with Multiple Sources

5.2.4.1 Authorship Attribution

Authorship attribution using stylometry is a field of study within statistics and computational linguistics with a long history. Mosteller and Wallace (1964) resolved a longstanding debate on the authorship of certain of the Federalist Papers. More recently, principal components analysis 'approach has been pioneered by Burrows (1987) in the field of literary and linguistic computing, while Rao and Rhatgi (2000) have shown that Burrows' techniques can be employed even more successfully with text taken from the Internet. A recent account of research on authorship attribution is given by Harold Love (2002); while works on forensic linguistics include Rieber and Stewart (1990), McMenamin and Choi (2002), Shuy (1998), and Grant (2004).

Stylometry techniques can be used to determine the likelihood that two documents of uncertain authorship are written by the same author, or that a document of unknown authorship is written by an author from whom sample writings are available. Similarly, given a set of documents with several authors, it is possible to partition the documents into subsets of documents all written by the same author. There are two parameters in such techniques: a) the data requirements per pseudonym, and b) the discriminating power of the technique. Using only semantic features, Rao and Rhatgi demonstrated that anonymity and pseudonymity cannot preserve privacy. Rao and Rhatgi did some exploratory research to confirm that inclusion of syntactic features, e.g., misspellings or other idiosyncratic features much more prevalent in web, as opposed to published, documents, could provide stronger results.

6. SEMANTIC HACKING: EXPANDING THE CONCEPT OF COGNITIVE HACKING

Pump and Dump vs. DDOS

Law Enforcement – Identity Theft

National Security

Bringing the Internet down vs. supply chain attacks

Insider Threat – deception detection

Utility-theoretic retrieval and new science of ISI

7. FUTURE WORK

In this chapter a variety of cognitive hacking countermeasures have been described, but implementation has begun on only a few of them. Our future work lies in implementation of the remaining countermeasures and in the development of countermeasures that can be used not only against cognitive attacks, but against semantic attacks more broadly, such as the attacks with

misinformation against autonomous agents, as described in Libicki's original definition of semantic hacking.

8. SUMMARY AND CONCLUSIONS

This chapter has defined a new concept in computer network security, cognitive hacking. Cognitive hacking is related to other concepts, such as semantic hacking, information warfare, and persuasive technologies, but is unique in its focus on attacks via a computer network against the mind of a user. Psychology and Communications researchers have investigated the closely related area of deception and detection in interpersonal communication, but have not yet begun to develop automated countermeasures. We have argued that cognitive hacking is one of the main features which distinguishes intelligence and security informatics from traditional scientific, medical, or legal informatics. If, as claimed by psychologists studying interpersonal deception, most interpersonal communication involves some level of deception, then perhaps communication via the Internet exhibits a level of deception somewhere between that of face to face interpersonal communication, on the one hand, and scientific communication on the other. As the examples from this chapter show, the level of deception on the Internet and in other computer networked settings is significant, and the economic losses due to cognitive hacking are substantial. The development of countermeasures against cognitive hacking is an important priority.

9. ACKNOWLEDGMENTS

Support for this research was provided by a Department of Defense Critical Infrastructure Protection Fellowship grant with the Air Force Office of Scientific Research, F49620-01-1-0272; Defense Advanced Research Projects Agency projects F30602-00-2-0585 and F30602-98-2-0107; and the Office of Justice Programs, National Institute of Justice, Department of Justice award 2000-DT-CX-K001 (S-1). The views in this document are those of the authors and do not necessarily represent the official position of the sponsoring agencies or of the US Government.

10. REFERENCES

1. Abel, S. (1998). "Trademark issues in cyberspace: The brave new frontier"
<http://library.lp.findlaw.com/scripts/getfile.pl?file=/firms/fenwick/fw000023.html>
2. Agre, P. (2001). "The market logic of information". *Knowledge, Technology, and Policy* vol. 13, no. 1, p. 67-77.
3. Anderson, R. (2002). Personal Communication
4. Anderson, R. H., Bozek, T., Longstaff, T., Meitzler, W., Skroch, M. and Wyk, K. Van. (2000). Research on Mitigating the Insider Threat to Information Systems - #2: Proceedings of a Workshop Held August 2000. RAND Technical Report CF163, Santa Monica, CA: RAND.

5. Anderson, R. and Khattak, A. (1998). "The Use of Information Retrieval Techniques for Intrusion Detection" *First International Workshop on Recent Advances in Intrusion Detection (RAID)*
6. Atallah, M. J., McDonough, C. J., Raskin, V., and Nirenburg, S. (2001). "Natural Language Processing for Information Assurance and Security: An Overview and Implementations" *Proceedings of the 2000 Workshop on New Security Paradigms*.
7. BBC News Online. (2001). " Hamas hit by porn attack"
http://news.bbc.co.uk/low/english/world/middle_east/newsid_1207000/1207551.stm
8. BBC News Online. (2001). "Sharon's website hacked"
http://news.bbc.co.uk/low/english/world/middle_east/newsid_1146000/1146436.stm
9. Biber, D. (1995). "Dimensions of Register Variation: A Cross-Linguistic Comparison" *Cambridge University Press*. Cambridge, England
10. Biber, D. (1986). "Spoken and written textual dimensions in English: Resolving the contradictory findings" *Language* vol. 62, no. 2, p. 384-413.
11. Buchanan, Ingersoll, P.C. (2001). "Avoiding web site liability—Online and on the hook?"
<http://library.lp.findlaw.com/scripts/getfile.pl?file=/articles/bipc/bipc000056.html>.
12. Buller, D.B. and Burgoon, J.K. (1996). "Interpersonal deception theory" *Communication Theory* vol. 6 no. 3, p. 203-242
13. Burgoon, J. K., Blair, J.P., Qin, T and Nunamaker, J.F. (2003). "Detecting Deception through Linguistic Analysis" *NSF / NIJ Symposium on Intelligence and Security Informatics, Lecture Notes in Computer Science*, Berlin: Springer-Verlag, June 1-3, 2003, Tucson, Arizona, 2003, p. 91-101.
14. Burrows, J.F. 1987. "Word Patterns and Story Shapes: The Statistical Analysis of Narrative Style" *Literary and Linguistic Computing*, vol. 2, p. 61-70.
15. Cao, J, Crews, J. M., Lin, M., Burgoon, J. K. and Nunamaker, J. F. (2003). "Designing Agent99 Trainer: A Learner-Centered, Web-Based Training System for Deception Detection" *NSF / NIJ Symposium on Intelligence and Security Informatics, Lecture Notes in Computer Science*, Berlin: Springer-Verlag, June 1-3, 2003, Tucson, Arizona, 2003, p. 358-365.
16. Chandy, K. M. and Misra, J. (1988). *Parallel Program Design: A Foundation*. Addison Wesley.
17. Chen, H., Zeng, D.D., Schroeder, J., Miranda, R., Demchak, C. and Madhusudan, T. (eds.). (2003). *Intelligence and Security Informatics: First NSF/NIJ Symposium ISI 2003 Tucson, AZ, USA, June 2003 Proceedings*, Berlin: Springer-Verlag.
18. Chez.com. (1997). "Disinformation on the Internet."
http://www.chez.com/loran/art_danger/art_danger_on_internet.htm

19. Cignoli, R. L.O., D'Ottaviano, I. M.L. and Mundici, D. (1999). *Algebraic Foundations of Many-Valued Reasoning* Boston: Kluwer Academic
20. Combs, J. E. and Nimmo, D. (1993). *The new propaganda: The dictatorship of palaver in contemporary politics*. New York: Longman.
21. Cooper, W. S. and Maron, M.E. "Foundations of Probabilistic and Utility-Theoretic Indexing". *Journal of the Association for Computing Machinery* vol. 25, no. 1, 1978, p. 67-80.
22. Cornetto, K. M. (2001). "Identity and Illusion on the Internet: Interpersonal deception and detection in interactive Internet environments" Ph.D.Thesis. University of Texas at Austin.
23. Cover, T. A. and Thomas, J. A. (1991). *Elements of Information Theory*. New York: Wiley
24. DELELTE LATER Cybenko, G., Giani, A. and Thompson, P. "Cognitive Hacking and the Value of Information" *Workshop on Economics and Information Security*, May 16-17, 2002, Berkeley, California.
25. Cybenko, G., Giani, A. and Thompson, P. "Cognitive Hacking: A Battle for the Mind" *IEEE Computer*, 35(8), 2002, 50-56.
26. Cybenko, G., Giani, A., Heckman, C. and Thompson, P. "Cognitive Hacking: Technological and Legal Issues", *Law Tech 2002* November 7-9, 2002.
27. Daniels, P., Brooks, H.M. and Belkin, N.J. (1997). "Using problem structures for driving human-computer dialogues" In Sparck Jones, Karen and Willett, Peter (eds.) *Readings in Information Retrieval San Francisco: Morgan Kaufmann*, p. 135-142, reprinted from RIAO-85 Actes: Recherche d'Informations Assistee par Ordinateur, Grenoble, France: IMAG, p. 645-660.
28. Dellarocas, C. (2001). "Building trust on-line: The design of reliable reputation reporting mechanisms for online trading communities" *Center for eBusiness@MIT* paper 101.
29. Denning, D. (1999). *Information warfare and security*. Reading, Mass.: Addison-Wesley.
30. Denning, D. (1999). "The limits of formal security models". *National Computer Systems Security Award Acceptance Speech*.
31. Doob, L. (1935). *Propaganda, Its psychology and technique* New York: Holt.
32. Drineas, P., Kerendis, I. and Raghavan, P. Competitive recommendation systems STOC'02, May 19-21 2002.

33. Ebay Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058 (N.D. Cal., 2000)
34. *Economist*. (2002). "Re-engineering in real time" 31 January.
http://www.economist.com/surveys/PrinterFriendly.cfm?Story_ID=949093.
35. Ellul, J. (1966). *Propaganda* translated from the French by Konrad Kellen and Jean Lerner New York: Knopf.
36. Farahat, A., Nunberg, G. and Chen, F. (2002). "AuGEAS (Authoritativeness Grading, Estimation, and Sorting)" *Proceedings of the International Conference on Knowledge Management CIKM'02* 4-9 November, McLean, Virginia.
37. Fawcett, T. and Provost, F. in W. Kloesgen and J. Zytchow (eds.). (2002) *Handbook of Data Mining and Knowledge Discovery*, Oxford University Press.
38. Felton, E. W., Balfanz, D., Dean, D., and Wallach, D. (1997). "Web spoofing: An Internet con game". Technical Report 54-96 (revised) Department of Computer Science, Princeton University.
39. 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.
40. Gertz v. Robert Welch, Inc., 428 U.S. 323, 94 S.Ct. 2997, 41 L.Ed.2d 789 (1974).
41. Google News beta. (2003). <http://news.google.com/>.
42. Grant, Tim. 2004. Ph. D. thesis, Forensic Section, School of Psychology University of Leicester (upcoming publication).
43. Hactivist, The. (2001). "Fluffi Bunni hacker declares Jihad"
<http://thehactivist.com/article.php?sid=40>
44. Heckman, C. and J. Wobbrock, J. (2000) "Put Your Best Face Forward: Anthropomorphic Agents, E-Commerce Consumers, and the Law". *Fourth International Conference on Autonomous Agents*, June 3-7, Barcelona, Spain.
45. Herlocker, J. (ed.). (2001). "Recommender Systems: Papers and Notes from the 2001 Workshop" *In conjunction with the ACM SIGIR Conference on Research and Development in Information Retrieval*. New Orleans.
46. Hofmann, T. (2001). "What People (Don't) Want". *European Conference on Machine Learning (ECML)*.
47. Hunt, A. (2001). "Web defacement analysis". ISTS.
48. Huynh, D., Karger, D. and Quan, D. (2003). "Haystack: A Platform for Creating, Organizing and Visualizing Information using RDF". *Intelligent User Interfaces (IUI)*

49. Information Warfare Site. (2001). <http://www.iwar.org.uk/psyops/index.htm>
50. “Interpersonal Deception: Theory and Critique” Special Issue *Communication Theory* vol 6. no. 3.
51. Johansson, P. (2002). “User Modeling in Dialog Systems”. St. Anna Report SAR 02-2.
52. Karlgren, J. and Cutting, D. (1994). “Recognizing text genres with simple metrics using discriminant analysis”
53. Kessler, B., Nunberg, G. and Schütze, H. (1997). "Automatic Detection of Genre" *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*
54. Krebs, B. (2001). " E-Mail Scam Sought To Defraud PayPal Customers" *Newsbytes* 19 December, <http://www.newsbytes.com/news/01/173120.html>
55. Lafferty, J. and Chengxiang, Z. (2001) Document language models, query models, and risk minimization for information retrieval. *2001 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*.
56. Lafferty, J. and Chengxiang, Z. (2001). “Probabilistic relevance models based on document and query generation” *Proceedings of the Workshop on Language Modeling and Information Retrieval, Carnegie Mellon University*.(Kluwer volume to appear PT reviewing).
57. Landwehr, C. E. (1984). “A security model for military message systems” *ACM Transactions on Computer Systems*. vol. 9, no. 3.
58. Landwehr, C. E. (1981). “Formal models of computer security”. *Computing Surveys*, vol. 13, no. 3.
59. Latimes.com. (2001). “‘Hacktivists,’ caught in web of hate, deface Afghan sites” <http://www.latimes.com/technology/la-000077258sep27.story?coll=la%2Dheadlines%2Dtechnology>
60. Lewis, M. “Jonathan Lebed: Stock Manipulator”, S.E.C. Nemesis – and 15 *New York Times Magazine* 25 February 2001
61. Lewis, M. (2001). *Next: The Future Just Happened* New York: W. W. Norton p. 35-36.
62. Libicki, M. (1994). “The mesh and the Net: Speculations on armed conflict in an age of free silicon”. National Defense University McNair Paper 28 <http://www.ndu.edu/ndu/inss/macnair/mcnair28/m028cont.html>

62. Love, Harold. 2002. *Attributing Authorship: An Introduction* Cambridge, UK: Cambridge University Press.

Lynch, C. (2001). "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*, vol. 52, no. 1, p. 12-17.

McMenamin, Gerald R. and Choi, Dongdoo (eds.). 2002. *Forensic Linguistics: Advances in Forensic Stylistics* Boca Raton, Florida: CRC.

63. Mann, B. (2000). "Emulex fraud hurts all". *The Motley Fool*.
<http://www.fool.com/news/foolplate/2000/foolplate000828.htm>
64. Maron, M.E. and Kuhns, J.L. "On relevance, probabilistic indexing and information retrieval". *Journal of the ACM* vol. 7 no. 3, 1960, p. 216-244.
65. Mateescu, G.; Sosonkina, M.; and Thompson, P. "A New Model for Probabilistic Information Retrieval on the Web" *Second SIAM International Conference on Data Mining (SDM 2002) Workshop on Web Analytic*
66. Matthew Bender and Company. (2001). Title 15.Commerce and Trade. Chapter 22. Trademarks General Provisions. *United States Code Service*. http://web.lexis-nexis.com/congcomp/document?_m=46a301efb7693acc36c35058bee8e97d&_docnum=1&wchp=dGLStS-ISIAA&_md5=5929f8114e1a7b40bbe0a7a7ca9d7dea
67. Mensik, M. and Fresen, G. (1996). "Vulnerabilities of the Internet: An introduction to the basic legal issues that impact your organization"
<http://library.lp.findlaw.com/scripts/getfile.pl?file=/firms/bm/bm000007.html>
68. Mosteller, F. and Wallace, D. L. 1964. *Inference and Disputed Authorship: The Federalist* Reading, MA: Addison-Wesley.
69. MSNBC. (2001). "Hacker alters news stories on Yahoo"
<http://stacks.msnbc.com/news/631231.asp>.
70. Mundici, D. and Trombetta, A. (1997). "Optimal Comparison Strategies in Ulam's Searching Game with Two Errors", *Theoretical Computer Science*, vol. 182, nos 1-2, 15 August.
71. Munson, J. C. and Wimer, S. "Watcher: the Missing Piece of the Security Puzzle", 17th Annual Computer Security Applications Conference (ACSAC'01). December 10 - 14, 2001 New Orleans, Louisiana
72. National Center for Digital Government. (2003). Integrating Information and Government John F. Kennedy School of Government Harvard University.
<http://www.ksg.harvard.edu/digitalcenter/>
73. National Center for Digital Government: Integrating Information and Government "Identity: The Digital Government Civic Scenario Workshop" Cambridge, MA 28-29

April 2003 John F. Kennedy School of Government Harvard University.
<http://www.ksg.harvard.edu/digitalcenter/conference/>

74. NetworkWorldFusion. (2001). "Clever fake of WTO web site harvests e-mail addresses" <http://www.nwfusion.com/news/2001/1031wto.htm>
75. New York v. Vinolas, 667 N.Y.S.2d 198 (N.Y. Crim. Ct. 1997).
76. Newsbytes. (2001). "Pop singer's death a hoax a top story at CNN" <http://www.newsbytes.com/cgi-bin/udt/im.display.printable?client.id=newsbytes&story.id=170973>
77. Pratkanis, A. R. and Aronson, E. (1992). *Age of propaganda: The everyday use and abuse of persuasion* New York: Freeman.
78. Rao, J. R. and Rohatgi, P. (2000). "Can pseudonymity really guarantee privacy?" *Proceedings of the 9th USENIX Security Symposium* Denver, Colorado August 14-17.
79. R.A.V. v. City of St. Paul, 505 U.S. 377, 112 S.Ct. 2538, 120 L.Ed.2d 305 (1992)
80. Register, The. (2001). "Intel hacker talks to The Reg" <http://www.theregister.co.uk/content/archive/17000.html>
81. Register, The. (2001). "New York Times web site sm0ked" <http://www.theregister.co.uk/content/6/16964.html>
82. Rich, E. (1983). "Users are individuals: individualizing user models" *International Journal of Man-Machine Studies* vol. 18 no. 3, p. 199-214.
83. Rieber, Robert W. and Stewart, William A. (eds.) 1990. *The Language Scientist as Expert in the Legal Setting*, *Annals of the New York Academy of Sciences* vol. 606, New York: The New York Academy of Sciences.
84. Rijsbergen, C.J van. *Information Retrieval* 2d. edition, London: Buttersworth, 1979.
85. Salton, G. and McGill, M. (1983). *Introduction to Modern Information Retrieval* New York: McGraw-Hill.
86. Sarwar, B., Karypis, G., Konstan, J. and Reidl, J. "Item-based collaborative filtering recommendation algorithms." *WWW10* May 1-5, 2001 Hong Kong.
87. Schneier, B. (2000). "Semantic attacks: The third wave of network attacks" *Cryptogram Newsletter* October 15, 2000. <http://www.counterpane.com/crypto-gram-0010.html>.
88. Shuy, Roger W. 1998. *The Language of Confession, Interrogation, and Deception* Thousand Oaks, California: SAGE Publications.
89. Smith, A.K. "Trading in False Tips Exact a Price", *U.S. News & World Report*, February 5, 2001, p.40

90. Smith, S. (2001). Personal communication.
91. Sophos. (2001). "W32/Redesi-B"
<http://www.sophos.com/virusinfo/analyses/w32redesib.html>
92. Thompson, P. "Semantic Hacking and Intelligence and Security Informatics" *NSF / NIJ Symposium on Intelligence and Security Informatics, Lecture Notes in Computer Science*, Berlin: Springer-Verlag, June 1-3, 2003, Tucson, Arizona, 2003.
93. Thornton, J. (2001). "Collaborative Filtering Research Papers".
<http://jamesthornton.com/cf/>.
94. Ulam, S.M. (1991). *Adventures of a Mathematician* Berkeley, CA: University of California Press.
95. Varian, H. R. (1996). "Resources on collaborative filtering"
<http://www.sims.berkeley.edu/resources/collab/>
96. Varian, H. R. and Resnik, P. eds. *CACM* Special issue on recommender systems, *CACM* vol. 40, no. 3, 1997.
97. Washtech.com. (2001). "FTC shuts down thousands of deceptive web sites"
<http://www.washtech.com/news/regulation/12829-1.html>
98. Wing, J. M. (1998). "A Symbiotic Relationship Between Formal Methods and Security"
Proceedings from Workshops on Computer Security, Fault Tolerance, and Software Assurance.
99. Yahalom, R., Klein, B., and Beth, Th. (1993). "Trust relationships in secure systems – A distributed authentication perspective. In *Proceedings of the IEEE Symposium on Research in Security and Privacy*, Oakland.
100. Yuan, Yougu; Ye, E. Z.; and Smith, S. (2001). "Web spoofing 2001" Department of Computer Science/Institute for Security Technology Studies Technical Report TR2001-409
101. Zhou, L., Burgoon, J. K. and Twitchell, D. P. (2003). "A Longitudinal Analysis of Language Behavior of Deception in E-mail". *NSF / NIJ Symposium on Intelligence and Security Informatics, Lecture Notes in Computer Science*, Berlin: Springer-Verlag, June 1-3, 2003, Tucson, Arizona, 2003, p. 102-110.
102. 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 36th Hawaii International Conference on Systems Science
103. SRD (2003). see http://www.the3dcorp.com/prod_nora.html