

# Counter Denial and Deception and Utility-Theoretic Information Retrieval for Intelligence Analysis

*Paul Thompson*

Department of Computer Science  
Dartmouth College  
6211 Sudikoff Laboratory, Hanover, NH 03755  
*Paul.Thompson@dartmouth.edu*

and

National Center for the Study of Counter-Terrorism and CyberCrime  
61 Wall Street  
Northfield, VT 05663

*Eugene Santos Jr., Qunhua Zhao, Gregory Johnson, and Hien Nguyen*  
Computer Science and Engineering Department  
University of Connecticut  
371 Fairfield Road, U-155, Storrs, CT 06269-3155  
*{Eugene,qzhao,gjohnson,hien,amohamed}@engr.uconn.edu*

**Keywords:** Multiple Competing Hypotheses, Information Sharing and Collaboration, HUMINT, OSINT, terrorism

## Abstract

Counter denial and deception plays a prominent role in intelligence analysis, but automated tools intended to support the intelligence analyst generally do not take denial and deception into account. Analysts often must choose a most likely analytic hypothesis among multiple competing hypotheses using information, or knowledge, from many diverse sources, both human and machine. As an analyst's sources become more numerous, detecting deceptive information becomes vital to making a correct, or at least more informed, analytic decision. This paper outlines the role for utility-theoretic, personalized information retrieval as an important foundational technologies for intelligence analysis. Such utility-theoretic retrieval will account for the value of information and will include countermeasures for denial and deception.

## 1. Introduction

This paper describes research conducted on the Semantic Hacking project (Cybenko et al. 2002, 2004) which addressed counter deception in the domain of computer network security, but the countermeasures it designed can be applied more broadly within intelligence and security informatics.

## 2. Background

Libicki first created a taxonomy of computer system attacks as being: a) physical, b) syntactic, or c) semantic (1994). He described semantic attacks in terms of misinformation inserted into interactions among intelligent, or autonomous, agents in the battlespace.

## 3. The Semantic Hacking Project

The Semantic Hacking project proposed several countermeasures to semantic, or cognitive, attacks. The News Verifier prototype was developed as a proof of concept for one of the countermeasures.

Countermeasures are divided into two categories depending on whether they involve detecting deception where there is a single, or multiple sources of information. Single source situations are those in which redundant, independent sources of information about the same topic are not available. Multiple source situations are those where multiple, presumably redundant, sources of information are available about the same subject of interest.

### 3.1 News Verifier Prototype

The News Verifier, a prototype cognitive hacking countermeasure, allows a 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 (Thompson, 1990a,b) 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.

#### **4. Utility-Theoretic IR and Counter Denial and Deception**

Intelligence and security informatics 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 semantic hacking countermeasures.

Information, 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.

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. When information contained in, say, an 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.

#### **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 and F49620-03-1-0014; 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.

#### **References**

- Cooper, William 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.
- Cybenko, G.; Giani, A.; and Thompson, P. August 2002. Cognitive Hacking: A Battle for the Mind. *IEEE Computer* 35(8): 50-56
- Cybenko, G.; Giani, A.; and Thompson, P. 2004. Cognitive Hacking. Zelkowitz, M. (ed.) *Advances in Computers*. vol. 60.
- Libicki, M. The mesh and the Net: Speculations on armed conflict in an age of free silicon National Defense University McNair Paper 28, 1994.  
<http://www.ndu.edu/ndu/inss/macnair/mcnair28/m028cont.html>
- 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.
- Thompson, P. 1990a. A Combination of Expert Opinion Approach to Probabilistic Information Retrieval, Part 1: The Conceptual Model. *Information Processing and Management*, vol. 26, no. 3, 1990, p. 371-382
- Thompson, P. 1990b. A Combination of Expert Opinion Approach to Probabilistic Information Retrieval, Part 2: Mathematical Treatment of CEO Model 3. *Information Processing and Management*, vol. 26, no. 3, 1990, p. 383-394