

Weak Models for Insider Threat Detection

Paul Thompson
Thayer School of Engineering and
Department of Computer Science
Dartmouth College
Hanover, New Hampshire 03755
Paul.Thompson@dartmouth.edu

ABSTRACT

This paper describes the design for a content-based approach to detecting insider misuse by an analyst producing reports in an environment supported by a document control system. The approach makes use of Hidden Markov Models to represent stages in the Evidence-Based Intelligence Analysis Process Model (EBIAPM). This approach is seen as a potential application for the Process Query System / Tracking and Fusion Engine (PQS/TRAFEN). Actions taken by the insider are viewed as processes that can be detected in PQS/TRAFEN. Text categorization of the content of analyst's queries, documents accessed, and work product are used to disambiguate multiple EBIAPM processes.

Keywords: insider threat, intrusion detection, multiple hypothesis tracking

1. INTRODUCTION

This paper describes the design for a content-based approach to detecting insider misuse. The focus here is on intelligence analysis in the context of a document control system, but this approach could be applied to other analytic environments, or other information technology environments where sensitive information is stored, e.g., those involving financial, or medical records. The overall framework for this approach is to use the Process Query System / TRAFEN [PQS/TRAFEN]¹ to detect processes characterized in terms of Hughes' Evidence-based Intelligence Analysis Process Model [EBIAPM]². The EBIAPM is represented as a Hidden Markov Model [HMM]³. This approach detects the progression of the insider through multiple EBIAPM HMMs, using a textual analysis of: a) the insider's queries; b) descriptions of the insider's task, if available; c) documents accessed; and d) the insider's work product.

2. BACKGROUND

The focus of this research is on being able to detect authorized and unauthorized misuse by an insider committing high stakes espionage using documents stored in a document control system, i.e., a document management system capable of secure access control which maintains a detailed log of all transactions. The trusted insiders who have historically caused the most damage to national security, such as Robert Hanssen and Aldrich Ames, were caught only after prolonged counterintelligence operations. Both of 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. In the case of Robert Hanssen once he was caught it became clear in retrospect that even the most cursory of information system auditing could have picked up the fact that he frequently queried counterintelligence databases about himself to see whether he was yet under suspicion.

Often in the past 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 and the fact that the insider seemed to be living at a higher standard than would be afforded by his / her salary or other sources of income.

Moving to an electronic document management, retrieval, and analysis environment makes it possible to detect signs of insider misuse much earlier than has previously been possible. Document control systems and document search 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. Academic information retrieval researchers have shown that searcher's relevance feedback judgments can significantly improve retrieval performance. Until recently use of relevance judgments in real world systems was considered impractical, because it was believed that searchers could not be induced to provide judgments⁴. With the development of e-commerce on the Web, various companies have successfully used implicit relevance judgments to build noisy, but useful, models of individual searchers. These implicit relevance judgments include various measurements derived from clickstream mining and session analysis, such as how long a user looks at a given web page, or whether the user prints it out. A sophisticated insider may be aware of countermeasures deployed against him. 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 attempt to manipulate Web search engine rankings on behalf of their clients. Recent research supported by the National Institute of Justice^{5,6} has begun to develop countermeasures against semantic, or cognitive, hacking, which will be adapted here to the problem of detection of deceptive queries by an insider attempting to camouflage misuse.

3. MODELING THE INSIDER THREAT

This approach employs defense in depth to attack the problem of insider misuse. First, the insider models provide a way to: a) monitor normal usage, b) take into account signatures of past misuse, and detect inconsistencies among the models. Software sensors attached to these models will feed observations in TRAFEN, a multiple hypothesis tracking engine³. Second, in recognition of the fact that related intrusion detection software, e.g., intrusion detection systems, have problems with dependence on known signatures of attack or too high false positive rates when detecting anomalous behavior, this research will make use of autonomic design principles and real-time feedback control to reduce the high false positive rate. An intrusion detection system based on these principles, α LAD, has been developed by Alphatech^{7,8}. Using such an autonomic defense may reduce the false alarm rate when detecting clear signs of misuse, but when confronting a skilled insider adversary, it is to be expected that some misuse will only be detected over a significant period of time, when the insider has come under suspicion, either through suspicious patterns of use of documents, through observations of the insider's broader activities, or both. It is therefore important that an approach to counter the insider threat also provide a way to unobtrusively monitor insider behavior in greater depth and also to have the capability to perform a detailed post-compromise analysis of the insider's interaction with the document control system.

A set of user models for the insider will be developed, including among possible others: a) a model based on known past examples of insider misuse, b) a model based on the user's work role in the organization, c) a model based on the insider's queries of the document control system, d) a model based on an analysis of the insider's interactions with the document control system as recorded by fine-grained logging, e) a model based on auditing of the insider's other actions on the information system, f) a model based on the content of documents accessed via the document control system, and g) a model based on the content of the insider's work product. In analyzing the behavior of the user a perspective analogous to that suggested for an analysis of the behavior of software programs by Munson and Wimer⁹ will be followed. These models will be represented in a common vector space representation as used in current spectral analysis-based, state-of-the-art recommender systems^{10,11}. One aspect of this approach will be to look for known signatures of insider misuse, or for anomalies in each of the behavioral models individually. Another aspect will be to look for discrepancies among all of the models. For example, if an insider is disguising the true intent of his queries by making deceptive queries that still lead him, if indirectly, to the documents that he is seeking, then this might be revealed

by comparing the queries to the texts of which the analyst actually makes use. Similarly, a discrepancy between the documents analyzed and the reports produced, could also reveal potential misuse.

3.1. EBIAPM HMMs

It is assumed that an intelligence analyst will have n tasks on which the analyst is working at a given time. These tasks will have been delegated and thus it will be assumed that these tasks are known to the computer system. Some may be long-term, on-going tasks, while others will be short-term. In addition to these delegated tasks, the malicious insider will be engaged in one, or more, additional non-delegated tasks related to the insider's malicious activity. It is assumed that an analyst will conduct intelligence analysis along the lines of the EBIAPM, the Evidence-Based Intelligence Analysis Process Model

It is supposed that an analyst who is working on multiple, delegated tasks will follow such a model for each of these tasks. His, or her, activity can be observed allowing identification of the state transitions of the analytic process. An HMM can be build for each of these tasks. The state transitions through the EBIAPM HMMs will be inferred by real-time tracking of document control system logs. While these logged events will be low-level, it will be possible to map the events to higher order observations indicating state transitions. Since all non-noise observations are assumed to be generated by either delegated or malicious EBIAPM HMMs, or, in the case of malicious activities, possibly partial EBIAPM HMMs; it will be necessary to associate observations with a single EBIAPM HMM, or a combination of EBIAPM HMMs. This will be done through text categorization. The categorization step will characterize each transition node based on its content, allowing a prediction as to which HMM generated the observation. Applying the Viterbi algorithm¹² to the observations enables determination of the most likely sequence of stages that support those observations and, based on the text categorization, the most likely user delegated task associated with the EBIAPM HMM.

Here, for purposes of illustration, a simplified version of the EBIAPM HMM is presented. In the appendix a more detailed HMM is shown. The states of the HMM are the five stages of the process, i.e., Define, Search, Marshal, Justify, and Present. So $S = (D, S, M, J, P)$.

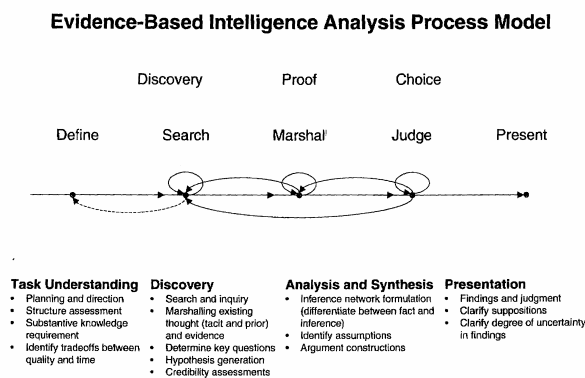


Figure 1 EBIAPM.

The observations which indicate the underlying states are those available to the system, such as the clickstream of the analyst and the analyst's transactions with the document control system. The transition matrix between states is a matrix in which a_{ij} is the probability of being in state S_j at time $t+1$ given that at time t the process was in state S_i . So, for example, there must be an index of the matrix that establishes the probability of being in state Judge immediately after state Define.

The observation symbol probability distribution in state j is the probability of a single observation given the fact that the process is in state j , for example, the probability of observing connections to web search engines given the fact that the process is in state Search. The initial state distribution is then the probability to have a particular state as initial state. If it is assumed that the process starts with the definition of the task, it should be $\pi = (1,0,0,0,0)$.

Once the five parameters of the model have been defined, a Viterbi algorithm is applied to determine the most likely state sequence that best explains the observations. If this sequence does not conform to a sequence corresponding to a delegated task, it suggests the possibility that the analyst is engaging in misuse, which should be further investigated.

3.2. Supporting content-based models

Modeling of the insider involves multiple models:

- Past Examples of Insider Misuse
- Insider's Work Profile, e.g., some system representation of the insider's role, clearances
- Insider's Queries
- Insider's Interactions with the Document Control System
- Insider's other Interactions with the larger Information System
- Documents accessed by the Insider
- Content of the Insider's Work Product.

Depending on the nature of an insider's malicious activity, there may, or may not, be a full underlying HMM EBIAPM associated with the malicious non-delegated task. If the analyst is carrying out the full analytic process for the malicious task, it can be assumed that there is an associated HMM EBIAPM and the problem is one of determining whether the observations are generated by the HMM for the malicious task.

3.3. EBIAPM HMMS and the document control system

The four HMMs presented in the appendix are sub-models of the overall EBIAPM HMM. Because the model is normative, that is a model of recommended practice of intelligence analysis, it will not necessarily be followed by all analysts. Nevertheless it is comprehensive and general enough that it should be able to model analysts not explicitly following the EBIAPM.

In the idealized model to be considered here, it will be assumed that analysts are using a single, integrated intelligence analysis product that supports the full range of activities represented in the EBIAPM. It will be assumed that an analyst has n delegated intelligence analysis tasks at any given time that are known to the system and that have a corresponding textual description. It will be assumed that the system has a record of which task is ostensibly being pursued at any time. Many of the states described in the HMMs, e.g., "Clarify with Customer" from the Task Understanding HMM are states that may involve some activities that go beyond even the larger computer information system. An analyst may leave his or her workstation and travel to the customer's work site for a conversation. On the other hand, these states may be observable, if, as in this example, the analyst instead sends the customer an e-mail message.

As discussed in section 3.1, the state transitions through the EBIAPM HMMs will be inferred by real-time tracking of document control system logs, i.e., using the fourth model, the insider's interactions with the document control system. Observations will be associated with a single EBIAPM HMM, or a combination of EBIAPM HMMs. This will be done through automatic text categorization based on four of the other six models, that is, the models: 1) insider's work profile, 2) insider's queries, 3) documents accessed by the insider, and 4) content of the insider's work product. Unlike the case with the fourth model which is based on analysis of state transitions, the categorization step will characterize each transition node based on its content, allowing a prediction as to which HMM generated the observation. The remainder of this section describes how each of these models will be used.

3.3.1. Past examples of insider misuse

From an analysis of past examples of insider misuse partial EBIAPM HMMs will be produced. These partial models will be used in a similar way to the full EBIAPM HMMs associated with modeling the insider's interactions with the document control system.

3.3.2. Insider work profile

The insider work profile will include, among other things, a representation of all delegated tasks for the insider. For sake of simplicity in modeling it will be assumed that there is a narrative, textual description of the task available to the system. Each task will have its associated EBIAPM HMM. In addition the textual narrative will provide an initial content-based characterization of the task.

3.3.3. Insider's queries

The text of all of the insider's queries will be maintained. In an advanced information retrieval environment it might be assumed that at the time that an analyst works on a particular task, the system will have an explicit representation of the fact that the analyst is working on that particular task. In lieu of such an environment, the system will be able to form a representation of the task based on the narrative textual description mentioned in 3.2 and measure the similarity of each query to all task descriptions in order to infer on which task the analyst is engaged.

3.3.4. Insider's other interactions with the larger information system

In this paper the focus is mainly on building models of the insider based on analysis of transactions with the document control system with the support of text categorization of content provided by other models, e.g., the insider's queries. The role of the insider's actions with the larger information system is analogous to the insider's interactions with the document control system, but is beyond the scope of the discussion in this paper.

3.3.5. Documents accessed by the insider

The text of the documents accessed by the insider will also provide a textual representation of the insider's task. The entire set of documents accessed can be analyzed to provide evidence of task, but each individual document will be considered as well. Documents which are outliers may be evidence of non-delegated, or malicious tasks.

3.3.6. Content of the insider's work product

The insider's work product will be analyzed in a similar way to the queries and documents accessed by the insider to provide evidence of the task on which the insider is working. Again, in an advanced information retrieval environment the system will have an explicit representation of the fact that a given work product is associated with a given task.

3.4. Observations

The observations corresponding to the hidden states of these HMMs are action / response pairs. That is for each action taken by the analyst, there is a response by the system. For example, an analyst may issue a text query and get back a ranked set of documents. The analyst's query and the ranked set of documents together are the observations. Analysis of the content of the query and documents will provide evidence for which hidden state the HMM is now in and to which EBIAPM HMM the state belongs. Anomalies in: a) state transition or b) inconsistencies among the different models of the analyst formed through the five other types of models, that is, past examples of insider misuse, insider's work profile, insider's queries, documents accessed by the insider, and content of the insider's work product will be used to detect insider misuse.

3.5. Deception by the malicious insider

A malicious insider is likely to attempt to conceal malicious actions. Thus even if an advanced information retrieval environment is in place, the knowledge that the system has, e.g., that a given query is associated with a given task, may be manipulated by the insider. The insider also may be able to make each of the individual models appear to reflect normal behavior. For example, an insider may issue deceptive queries which return documents from which the insider can infer information needed for malicious activity without triggering suspicion when considering query or document models separately. To counter insider deception, the approach described here also checks for inconsistencies among the various models. Thus, if an insider's queries and the content of work product seem to belong to the EBIAPM HMM of a particular delegated task, but the content of documents accessed seems inconsistent, this may be because some of the documents accessed were in support of a malicious task, even though this was not apparent through analysis of the queries, documents, or work product separately.

4. DISCUSSION

HMMs are considered strong, not weak models³. It may turn out that it will not be possible to obtain the training data needed to accurately estimate the probabilities needed for the EBIAPM HMMs. In this case an alternative approach would be developed using weak models. Moreover, even if the probabilistic parameters required for the EBIAPM HMMs can be obtained, the approach taken here is weak in another sense. Detecting which EBIAPM, corresponding to a delegated task, is being tracked depends on the content-based textual analysis of the other models, i.e., the models based on the user's task, queries, documents accessed, and work product.

It is also apparent that building a system as described here would be a complex undertaking. Software sensors would need to be built to record not only all interactions with the document control system, but the user's interactions with the entire information system in which the document control system is embedded would need to be monitored as well. As discussed in a Mitre workshop on Cyber Indications and Warnings¹³, the user's verbal and non-digital interactions would be monitored, as well.

5. CONCLUSIONS

The approach described here focuses on modeling the insider through the insider's interactions with the document control system and textual models of the insider's task, queries, documents accessed, and work product. The EBIAPM in its full generality includes other factors in modeling the insider threat that have been discussed, for example, in the recent series of Cyber Indications and Warnings workshops at Mitre¹³. Such other factors include an analysis of transactions with the larger information system, e.g., including e-mail and calendar applications, as well as modeling of the insider, according to personality and life style circumstances. Some work has been done on such models in the past and a more complete approach could be developed through their incorporation in the model described here.

ACKNOWLEDGEMENTS

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.

REFERENCES

1. G. Jiang, Weak process models for robust process detection, *Proceedings of the Defense and Security Symposium*. April, Orlando, Florida, 2004.
2. L. Liddy, Scenario based Question Answering, *Intelligence Analysis Futures Day*, Mitre Speaker Series, Bedford, MA, 27 June, 2003.
3. L. Rabiner, A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE* 77:257-286, 1989.
4. G. Matescu, M. Sosonkina, P. Thompson, A New Model for Probabilistic Information Retrieval on the Web, *Second SIAM International Conference on Data Mining (SDM 2002) Workshop on Web Analytics*, Arlington, Virginia, 2002.
5. G. Cybenko, A. Giani, and P. Thompson, Cognitive Hacking: A Battle for the Mind, *IEEE Computer* vol. 35, no. 8, August 2002, p. 50-56.
6. G. Cybenko, A. Giani, and P. Thompson, Cognitive Hacking In: M. Zelkowitz (ed.), *Advances in Computers*, 2004 (to appear).
7. D. Armstrong, S. Carter, G. Frazier, T. Frazier, Autonomic Defense: Thwarting Automated Attacks via Real-Time Feedback Control, *Complexity*, vol. 9, issue 2, p.41-48, 2004.
8. O. Kreidl, and T. Frazier, Feedback Control Applied to Survivability: a Host-Based Autonomic Defense System, *IEEE Transactions on Reliability*, vol. 52, no. 3, September 2003.
9. J. Munson, and S. Wimer, Watcher: the Missing Piece of the Security Puzzle, *17th Annual Computer Security Applications Conference (ACSAC 2001)*.
10. P. Drineas, I. Kerendis, and P. Raghavan, Competitive recommendation systems, *STOC'02*, May 19-21, 2002.
11. Y. Azar, A. Fiat, A. Karlin, F. McSherry, J. Saia, Spectral Analysis of Data, *ACM Symposium on Theory of Computing*, 2000.
12. G. Forney, The Viterbi Algorithm, *Proceedings of the IEEE*, 61:268-278, 1973.
13. M. Maybury, J. Sebring, and P. Chase, Cyber Indications & Warnings, workshop presentation, Mitre, Bedford, Massachusetts, 2003.

APPENDIX

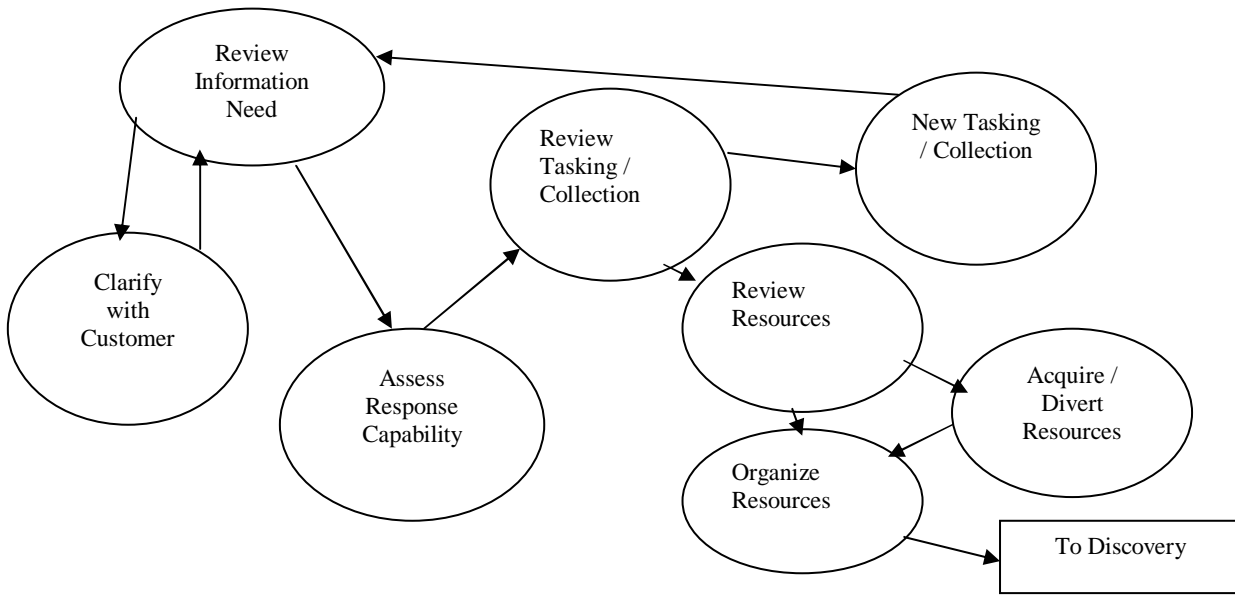


Figure 2. HMM for Task Understanding

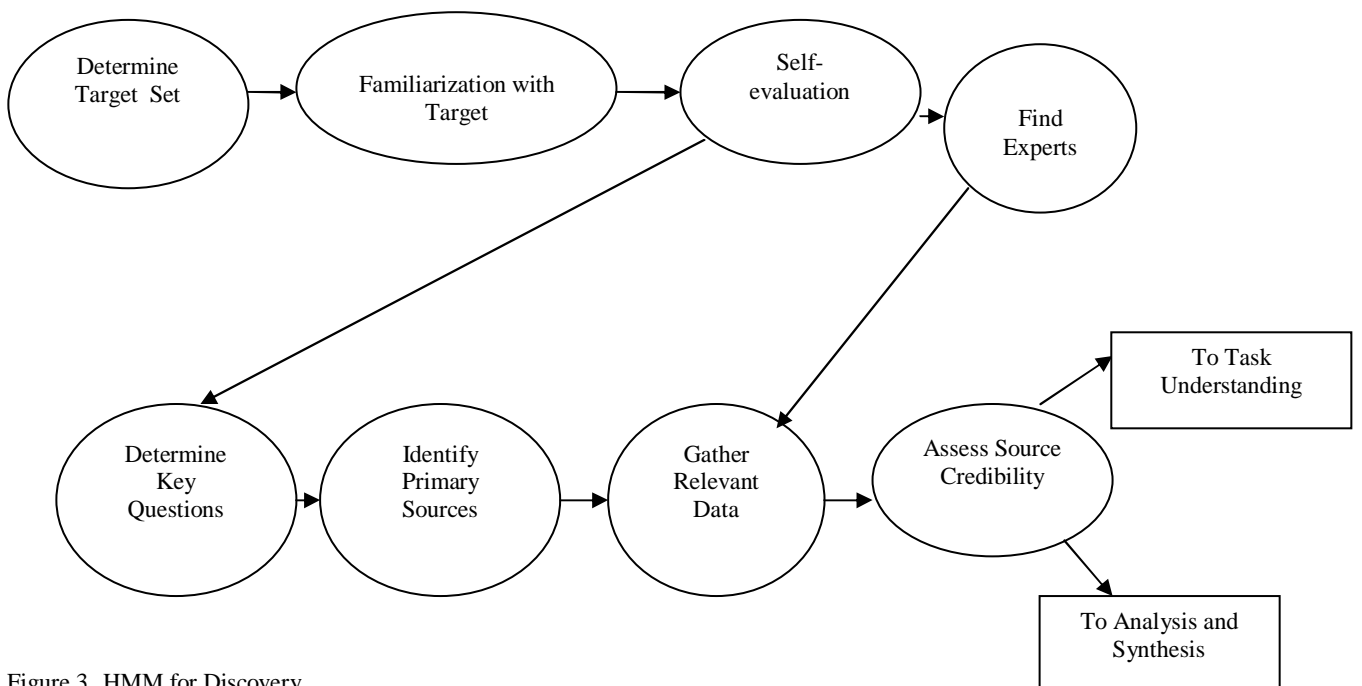


Figure 3. HMM for Discovery

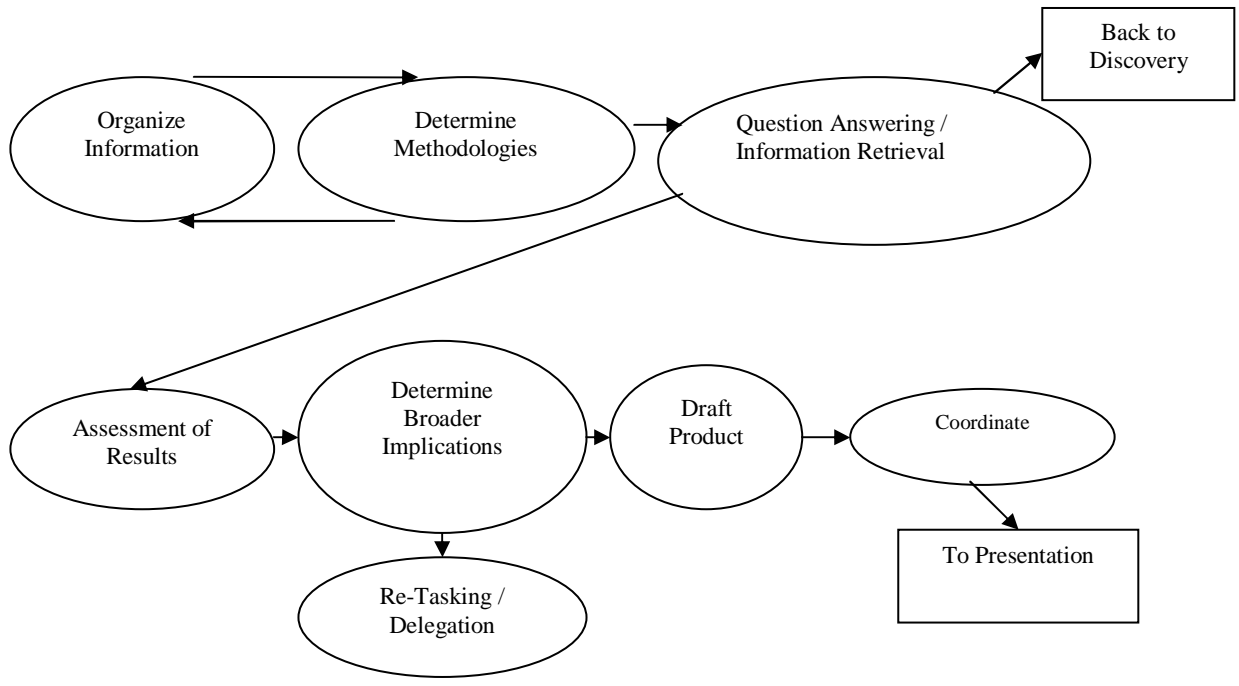


Figure 4. Analysis and Synthesis

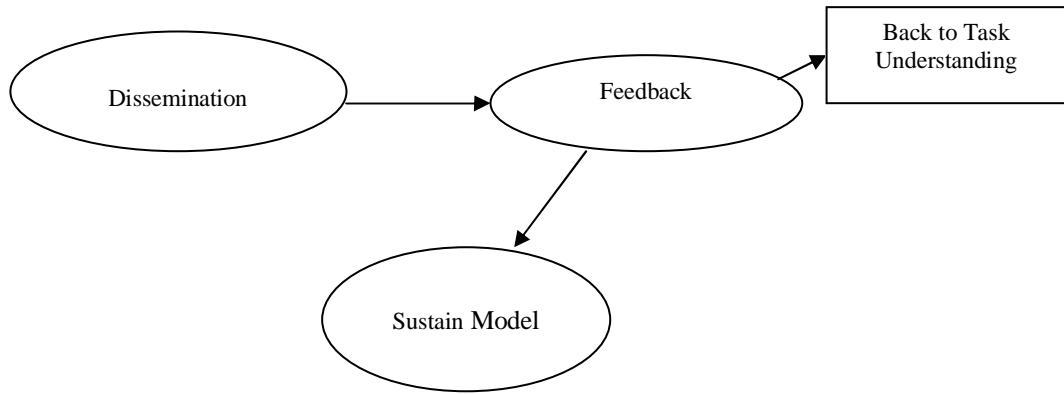


Figure 5. Presentation