Cognitive Process

John Yen, Robert F. Erbacher, Chen Zhong, and Peng Liu

1 Introduction

The previous chapter showed that our understanding about the cognitive reasoning process of cyber analysts is rather limited. Here, we focus on ways to close this knowledge gap. This chapter starts by summarizing the current understanding about the cognitive processes of cyber analysts based on the results of previous cognitive task analyses. It also discusses the challenges and the importance to capture "fine-grained" cognitive reasoning processes. The chapter then illustrates approaches to overcoming these challenges by presenting a framework for non-intrusive capturing and systematic analysis of the cognitive reasoning process of cyber analysts. The framework includes a conceptual model and practical means for the non-intrusive capturing of a cognitive trace of cyber analysts, and extracting the reasoning process of cyber analysts by analyzing the cognitive trace. The framework can be used to conduct experiments for extracting cognitive reasoning processes from professional network analysts. When cognitive traces are available, their characteristics can be analyzed and compared with the performance of the analysts.

Detecting complex multi-step cyber attacks are challenging for cyber analysts for several reasons. First, the alerts received by cyber analysts include false positives. This requires the analyst to filter out false positive alerts in a timely fashion. The false positive alert may mislead the analysts such that their time is wasted on

J. Yen (✉) • C. Zhong • P. Liu
College of Information Sciences and Technology, The Pennsylvania State University,
University Park, PA 16802, USA
e-mail: jyen@ist.psu.edu; czz111@ist.psu.edu; pliu@ist.psu.edu

R.F. Erbacher
The U.S. Army Research Laboratory, Adelphi, MD 20783, USA
e-mail: robert.f.erbacher.civ@mail.mil

A. Kott et al. (eds.), *Cyber Defense and Situational Awareness*,
Advances in Information Security 62, DOI 10.1007/978-3-319-11391-3_7

false alarms, delaying their attention to the alerts related to actual attacks. Second, an alert related to the attack may be missing (i.e., false negative) due to an unknown vulnerability or a new way of exploiting a known vulnerability. Due to missing alerts, analysts may not be able to identify certain attack steps in an attack chain, and hence delay the time to detect the multi-step attack.

One way to deal with false positive alerts and missing alerts is to leverage previous experience (both successful and failure experience) of cyber analysts in handling similar situations. For example, a failure experience associated with a previous false alarm can prevent an analyst from pursuing a similar false alarm. Similarly, a successful experience associated with a previous missing alert can help the analyst to adapt the experience to deal with a similar missing alert in a new cyber attack. A senior analyst, with years of rich experience in cyber analysis, accumulates many experiences of different types. If the cognitive process of these experiences can be effectively captured and analyzed such that they can be aggregated and effectively reused by other analysts, it will provide several important benefits.

Previous Cognitive Task Analyses (CTAs) about cyber defense have provided valuable insights about the high-level cognitive processes of cyber analysts in the real world. Biros and Eppich (2001) identified four cognitive capabilities. D'Amico and Whitley (2008) generated six analysis roles of cyber analysts: **triage analysis**, **escalation analysis, correlation analysis, threat analysis, incident response**, and **forensic analysis**. We will elaborate on these roles and their relationship with other related cognitive processes. Erbacher et al. (2010a, b) extended the scope of the CTA further to include **vulnerability assessment** and a "big picture" component to highlight the interaction between the tactical-level cyber analysis (e.g., analyzing attacks within an enterprise's regional network) and strategic-level cyber analysis (e.g., detecting attacks involving multiple regions or multiple countries around the globe).

Based on the results of these CTA's, we synthesized and summarized the high-level cognitive processes of cyber analysts and their dependency relationships in Fig. 1. The ovals in the figure represent processes, and the rectangles in the figure represent Data or Information. Because some of the processes are performed by human analysts while some are performed by machine, we distinguish them using solid ovals for cognitive processes of cyber analysts, and white ovals for processes automated by software. For example, "IDS" refers to "intrusion detection system" such as SNORT.

A cyber analysis process transforms a huge amount of raw data in the network (e.g., network packets) and in each computer in the network (e.g., record of system calls such as authentication of a user's password) into decisions about "incident" (which represents a cyber attack that needs to be responded), which lead to response actions (e.g., shutting down a compromised machine) and further actions to mitigate the impact of the incident. This is the tactical level cyber analysis. Cyber analysts also need to correlate related incidents (which may be detected in different regions, different countries, or even possibly far in time) that are parts of a larger attack scheme. This is referred to as the strategic level cyber analysis (D'Amico and Whitley 2008).

The tactical-level cyber analysis also includes vulnerability scanning (typically performed by machine, but can be initiated and scheduled by a human analyst), which perform vulnerability assessment based on known vulnerabilities. Vulnerability of

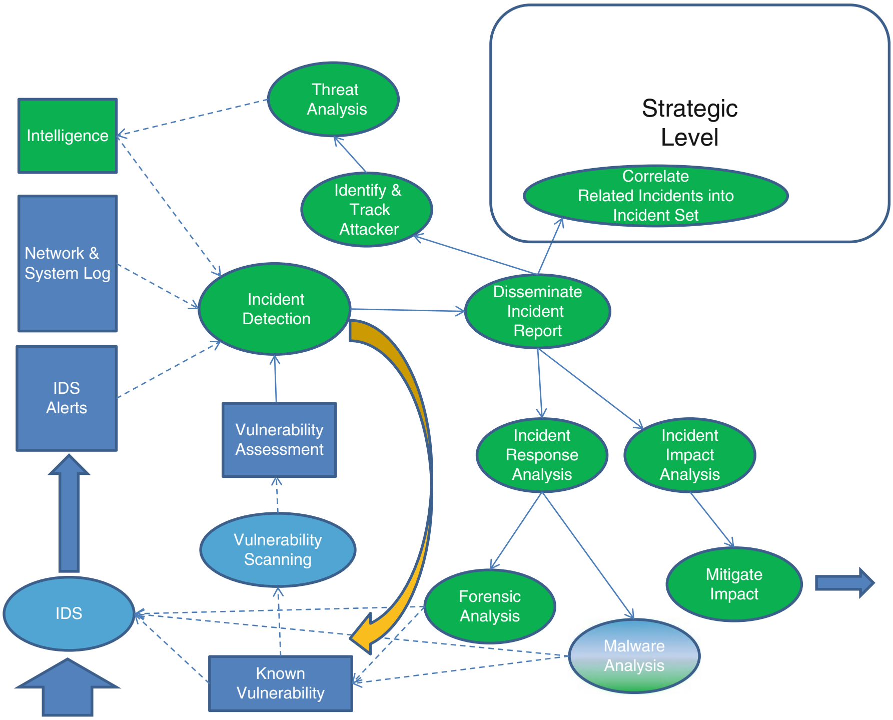


Fig. 1 The tactical-level cognitive processes of cyber analysts

machines often plays a key role for an analyst to confirm an incident. After an incident is detected, a formal report is generated by the analyst and disseminated for four types of further analyses at the tactical level: (1) incident response (for minimizing damage and expedited repairs), (2) impact analysis and mitigation plan (e.g., impacts to the current mission of war fighters), (3) identify and track attackers through threat analysis (i.e., intelligence gathering, analysis, and fusion for identifying the sponsor and the intent of the attack), and (4) forensic and malware analysis to obtain further details about the incident. The fourth step is especially important for a zero-day attack (e.g., an attack exploiting an unknown vulnerability), because they are crucial to identify the "signature" of the attack so that they can be incorporated into IDS for detecting future attacks of the same type.

D'Amico and Whitley (2008) identified six analysis roles that accounted for all the cognitive work performed by cyber analysts: (1) triage analysis, (2) escalation analysis, (3) correlation analysis, (4) threat analysis, (5) incident response, and (6) forensic analysis. While the role of the latter three has been explicitly represented in Fig. 1, the first three cognitive roles are part of Incident Detection and other functions performed by the analysts. Triage analysis filters the large amount of data (e.g., IDS alerts, network or system log) to identify "suspicious activity", which feed to escalation analysis to investigate, interpret, and assemble data from multiple sources over a time period longer than that of triage analysis. Correlation analysis searches for patterns and trends in current or historic data. D'Amico and Whitley (2008) also described the workflow involving these three roles as an iterative process. Some of the details of these processes are still not well understood. For example, D'Amico and Whitley pointed out that analysts look for unexplained patterns during correlation analysis:

> An analyst might not know what patterns they are looking for in advance; instead, the analyst might "know it when they see it". When they encounter a pattern that they cannot explain, they form hypotheses about potential malicious intent, which they try to confirm or contradict via additional investigation.

How do cyber analysts actually perform this and other cognitive roles? What are the cognitive processes that tie these analysis roles together? To answer these questions, we need to capture and analyze the fine-grained cognitive reasoning processes of cyber analysts. In the following section, we describe the state-of-the-art in capturing fine-grained cognitive reasoning processes and the difficulties for applying it to the tasks performed by cyber analysts.

1.1 Fine-Grained Cognitive Reasoning Process Capture and Analysis

We use the term "fine-grained cognitive reasoning process" to refer to the detailed cognitive process that describes individual actions and reasoning steps performed by an analyst and the relationships between these actions and reasoning steps. For example, one or more hypotheses can be formulated by the analyst at a particular point of the reasoning process based on the observations the analyst has made up to that point. These hypotheses can be later refined, rejected, or confirmed by the analyst during his/her reasoning process. For cyber analysts, such detailed cognitive reasoning process can complement the "high-level cognitive processes" described in the previous section in four important ways. First, it will improve our understanding about the difference in the cognitive reasoning processes of the experts and less-experienced analysts. Such an understanding is critically important to facilitate the design of better training tools for cyber analysts. Second, the fine-grained cognitive reasoning process of cyber analysts can provide a unique basis for identifying the opportunities to improve the visualization support for cyber analysts (Erbacher et al. 2010a, b). Third, the analysis of fine-grained cognitive reasoning process can lead to the design of automated cognitive aid tools by reusing and/or aggregating the processes of analysts to enhance the performance of analysts. Finally, the automated capture of the fine-grained cognitive reasoning process of cyber analysts can facilitate the sharing of relevant information and knowledge between cyber analysts, whether they are in different work shifts or in different geographic locations.

Existing methods for capturing fine-grained cognitive reasoning process include (1) talk-aloud protocol, (2) think-aloud protocol, (3) retrospective reports protocol, (4) observational case study, and (5) behavior trace capture. The first three methods are also referred to as types of "verbal protocol analysis" (Ericsson and Simon 1980, 1993). In a verbal protocol analysis, a subject performs a given task while being monitored by experimenters and being recorded (audio or video). In a talk-aloud protocol, the subject is asked to verbally articulate anything that comes to their mind in performing a given mental task. In a think-aloud protocol, the subject is asked to verbally describe anything that comes to their mind as they think to solve a problem. In a retrospective reports protocol, the subject is asked to reflect and articulate their thinking after they solve the problem. Retrospective reports can be combined with one of the first two protocols to validate their completeness (Ericsson and Simon 1993). Protocol analysis is the basis for knowledge acquisition methods, which elicit expert knowledge and encode them in an artificial intelligence system (often referred to as "expert systems", "knowledge-based systems", or "intelligent agents") through interviews and case studies. Due to the complexity of these tasks, the verbal protocol analysis needs to be augmented with an "interviewer" (typically referred to as a "knowledge engineer" due to their familiarity of the target representation language to be used to encode the expertise), who guides the thinking aloud protocol by asking probing questions, and by providing information to simulate the outcome of an action (e.g., test result of a diagnostic task) performed by the subject (Durkin 1994). While this elicitation method is feasible for tasks whose actions generate a limited number of outcomes (e.g., result of a test is positive or negative), it is difficult to apply the method to cyber analysis task whose actions (e.g., filter alerts for a particular port number) can lead to a wide range of possible outcomes.

The fourth method for acquiring fine-grained cognitive reasoning process is observational case study, which observes the subject in performing a task (Bell and Hardiman 1989). This method can be combined with think aloud protocol and/or retrospective report protocol. A case or a scenario is used in observational study to provide a context and relevant information in response to the actions of the subject.

The fifth method for obtaining fine-grained cognitive reasoning process is behavior trace, which transforms the observational data gathered from the subject into a "behavior trace". Tools (such as MacSHAPA) have been developed to facilitate the generation of such behavior trace from observational data (Sanderson et al. 1994). For example, a knowledge/cognitive engineer can use MacSHAPA to encode actions and/or communications captured in the observational data as template or predicate. While this type of tool is useful, it cannot extract the cognitive process that the subject did not explicitly articulate in the think-aloud protocol.

In the rest of this chapter, we first provide a literature review about research related to capturing cognitive process. This is followed by a framework for non-intrusive capturing and analysis of fine-grained reasoning cognitive processes, which includes (1) the Action-Observation-Hypothesis (AOH) conceptual model, (2) the non-intrusive capturing of a cognitive trace of cyber analysts containing a temporal sequence of AOH objects and relationships, and (3) extracting the reasoning process of cyber analysts by analyzing the cognitive trace. Section 4 presents a case study of applying the framework to systematic capturing of the cognitive reasoning process from professional network analysts and the initial results of analyzing the cognitive traces. Finally, we summarize the key contributions of systematic capturing of the cognitive reasoning process of cyber analysts and its critical enabling role toward a more agile cyber defense.

2 Literature Reviews

2.1 *Cognitive Task Analysis*

A cognitive task analysis (CTA) (Crandall et al. 2006) derives the required tasks for highly analytical (cognitive) activities such as decision-making; network analyst determination of network event relevance, importance, and characterization is of particular relevance. More specifically, a CTA attempts to determine what tasks are required to be performed and how the target experts perform said tasks. A cognitive task analysis is critical for developing correct tools and capabilities to improve the effectiveness of the network analyst, such as advanced displays, recommender systems, etc. Three CTAs are particularly relevant to network analysis from existing literature.

- The first CTA (Foresti and Agutter n.d.) examined the tools used by network experts at the time of the CTA as well as the advanced displays that had been developed for use by network experts. The focus of the CTA was to acquire the fundamentals necessary for the development of advanced displays geared towards improving network administrator efficiency. Additionally, results of the CTA identified the temporal organization of decisions and event prioritization through semi-structured interviews.
- The second CTA (D'Amico et al. 2005, D'Amico and Whitley 2008) had three goals. First was to study the set of analyst goals. Second was to identify the needed analyst expertise and their depth. Third was to identify the viability of visual representations and how such visual representations might be used. This study was performed through subject interviews of seven different organizations.
- The third study (Erbacher et al. 2010a, b) performed interviews of individuals with different levels of decision-making responsibility within network operations at Pacific Northwest National Laboratory. In addition to a wide range of requirements, this study generated a cyber command and control task flow diagram with primary tasks including assessment, detailed assessment, response, audit, and big picture, which is shown in Fig. 2.

2.2 *Case-Based Reasoning*

The reuse of cyber analysts' analytical reasoning results has been investigated using case-based reasoning (CBR). Given a problem, a CBR system retrieves a similar problem from a case library (also referred to as case base or knowledge base), modifies its solution for the given problem, and retains the new problem and solution in the case library (Stahl 2004). The original concept of CBR derives from a cognitive model of dynamic memory by Schank (1982), which led to computer-based CBR systems (Kolodner 1983; Lebowitz 1983). The process model of CBR developed by

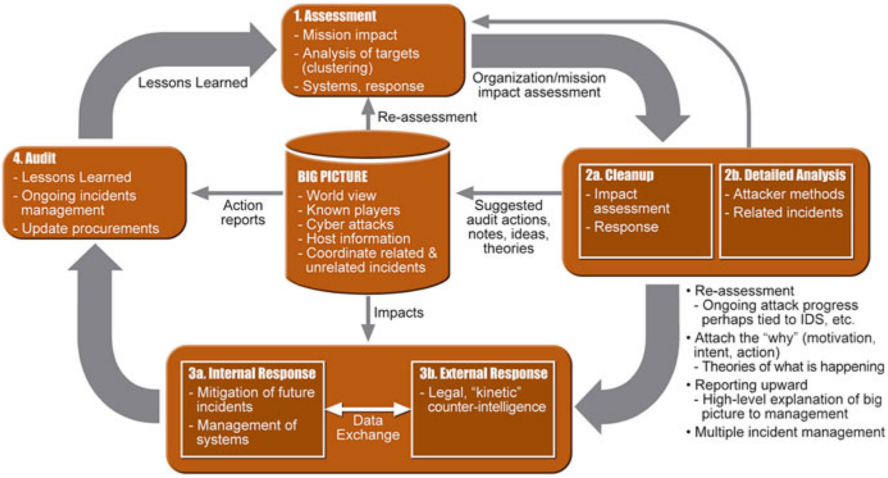

1. Assessment
- Mission impact
- Analysis of targets (clustering)
- Systems, response
Lessons Learned
Organization/mission impact assessment
Re-assessment
4. Audit
- Lessons Learned
- Ongoing incidents management
- Update procurements
Action reports
BIG PICTURE
- World view
- Known players
- Cyber attacks
- Host information
- Coordinate related & unrelated incidents
Suggested audit actions, notes, ideas, theories
2a. Cleanup
- Impact assessment
- Response
2b. Detailed Analysis
- Attacker methods
- Related incidents
Impacts
• Re-assessment
- Ongoing attack progress perhaps tied to IDS, etc.
• Attach the "why" (motivation, intent, action)
- Theories of what is happening
• Reporting upward
- High-level explanation of big picture to management
• Multiple incident management
3a. Internal Response
- Mitigation of future incidents
- Management of systems
Data Exchange
3b. External Response
- Legal, "kinetic" counter-intelligence

Fig. 2 A cyber command and control task flow (Erbacher et al. 2010a, b)

Aamodt and Plaza (1994), consists of four components: retrieve, reuse, revise, and retain. The model has driven the majority of research and application development in CBR research. Research into each of the four component areas has been extensive resulting in numerous reviews and surveys (De Mantaras et al. 2005). An extension of CBR model, shown in Fig. 3, explicitly includes the generation of incident reports by analysts (Erbacher and Hutchinson 2012).

In operations, a new scenario is matched against the existing scenarios to find the most relevant match, which is then mapped, using a similarity metric, to the new scenario providing an updated solution. Such case-based reasoning has been applied to a wide range of domains including:

- Breathalyzers (Doyle 2005)
- Bronchiolitis (Doyle 2005)
- E-Clinic (Doyle 2005)
- Intelligent tutoring systems (Soh and Blank 2008)
- Help desk systems, i.e. diagnosis (Stahl 2004)
- Electronic commerce product recommendation systems (Stahl 2004)
- Classification, i.e., class membership (Stahl 2004)

The retrieval component of CBR requires a similarity metric between cases. A survey/taxonomy of similarity metrics can be found in Cunningham (2008). Examples of research in retrieval mechanisms include:

- Information theory approaches (Ranganathan and Ronen 2010). This research provides for the identification of similarities between instances in an ontology.
- User defined functions (Sterling and Ericson 2006). The associated patent also covers the representative database issues.

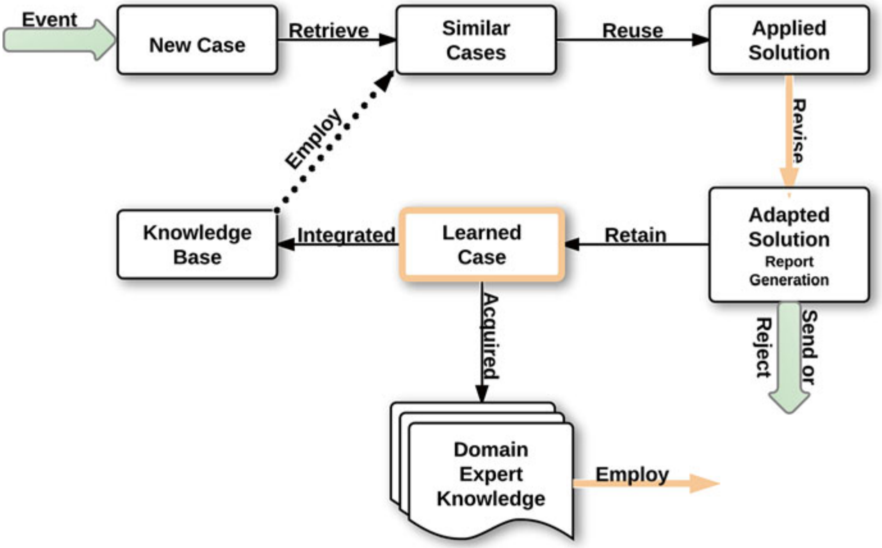


Fig. 3 An extended case-based reasoning process model

- Abduction versus deduction (Sun et al. 2005).
- Fuzzy similarity (Sun et al. 2005).
- Contextual probability (Wang and Dubitzky 2005). This metric integrates probability with distance-based neighborhood weighting and works for both ordinal and nominal data.
- Adaptive similarity (Long et al. 2004). This paradigm allows for specification of new similarity metrics and identification of the similarity metric to be applied in particular scenarios without the need for reprogramming.
- Semantic vs. syntactical similarities (Aamodt and Plaza 1994).
- Models of similarity (Osborne and Bridge 1997). The goal of this work was to identify the primary classes of similarity including absolute and relative similarity metrics.

Specific similarity metrics for categorical data include overlap, eskin, IOF, OF, Lin, Lin1, Goodall1, Goodall2, Goodall3, Goodall4, Smirnov, Burnaby, Anderberg, and Neighborhood Counting Metric (Boriah et al. 2008; Wang and Dubitzky 2005).

Case-based reasoning has been applied to support the reuse of an analyst's "report" that summarizes the analyst's analytical reasoning results regarding previous cyber-attacks so that the efforts of generating reports for a newly detected attack can be reduced significantly (Erbacher and Hutchinson 2012). However, CBR has not been used to capture and reuse the process of the analyst's analytical reasoning process. One of the challenges in applying CBR to retrieving and reusing analytical reasoning processes is a lack of non-intrusive way to capture them.

3 A Systematic Cognitive Reasoning Process Capture and Analysis Framework

To address the challenges of capturing a detailed cognitive process of a cyber analyst, we have developed a framework and associated cognitive trace tool for capturing the cognitive reasoning process of a cyber analyst. The framework not only integrates observational study and behavior trace methods described in Sect. 1.1, but also extends the previous approaches by enabling analysts to record their thinking (as "hypotheses"), and linking them to observations of interests during the observational study. In a way, the framework transforms "think aloud" to "type aloud"— instead of verbally articulating their thinking, analysts record each step of their cognitive reasoning process in a naturalistic way (not necessarily monitored) in the context of solving a given case involving cyber-attacks.

In the rest of this section, we first describe the conceptual model of the framework, which we will refer to as the A-O-H model, named after the three main objects in the framework: **Actions** performed by a subject, **Observations** of interest to the subject, and **Hypothesis** generated by the subject based on the observations. We then introduce the relationship between these objects that forms the analytical reasoning process of cyber analysts. Section 3.3 describes the AOH objects and relationship captured in a non-intrusive way. Finally, we discuss how the reasoning process can be extracted from the cognitive trace to provide the basis for systematic analysis of the cognitive reasoning processes at the individual level as well as across multiple analysts.

3.1 The A-O-H Conceptual Model of an Analytical Reasoning Process

A conceptual model of the analytical process of cyber analysis is informed by cognitive science theories including sense making theory and naturalistic decision making. The sense making theory builds on three key cognitive constructs: **Action**, **Observation**, and **Hypothesis**. Actions refer to analysts' evidence

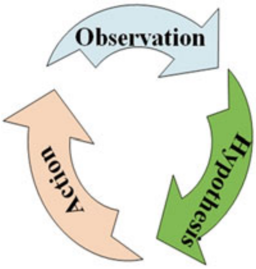


Fig. 4 The iterative analytical reasoning process involving action, observation and hypothesis (A-O-H Model) (Zhong et al. 2013)

exploration activities; Observations refer to the observed data/alerts considered relevant by the analysts; Hypotheses represent the analysts' awareness and assumptions in a certain situation. These three constructs iterate and form reasoning cycles. Actions can lead to new or updated observations, which result in new or updated hypotheses, and later subsequent Actions. Not surprisingly, these three constructs, being part of the general sense making theory, naturally map to cognitive activities of cyber analysts. While Actions and Observations in cyber analysis are obvious, Hypotheses are not explicit (i.e., they are "tacit" knowledge) and cannot be fully anticipated due to new attack behaviors (hence needs to be entered by the analyst in a semi-formal representation). Often, a Hypothesis is not known for certain until further evidence (e.g., presence of relevant vulnerability on a node) is gathered to confirm or disconfirm. All the Hypotheses maintained by an analyst are called "**Working Hypotheses**". We call the instances of Action, Observation and Hypothesis as "**AOH Objects**".

3.2 The AOH Objects and Their Relationships Can Represent the Analytical Reasoning Processes

In the iterative cycles of analytical reasoning processes, Hypotheses in different sense making cycles be related in important ways. One set of Actions and Observations can lead to a set of disjunctive hypotheses. Therefore, the AOH Objects are connected to each other in an analytical reasoning process. This is illustrated in Fig. 5. Since an Action always results in an Observation, we put Action and Observation in a unit, called "**AO**". The Hypotheses ("**H**"s) being the children of an AO indicates these Hypotheses are generated based on the AO. An AO being a child of an H indicates that this AO is triggered by the H. We can also consider the Hypotheses only. If an H_1 has an AO unit as its child and another H H_2 is a child of AO, we say that H_2 is a child

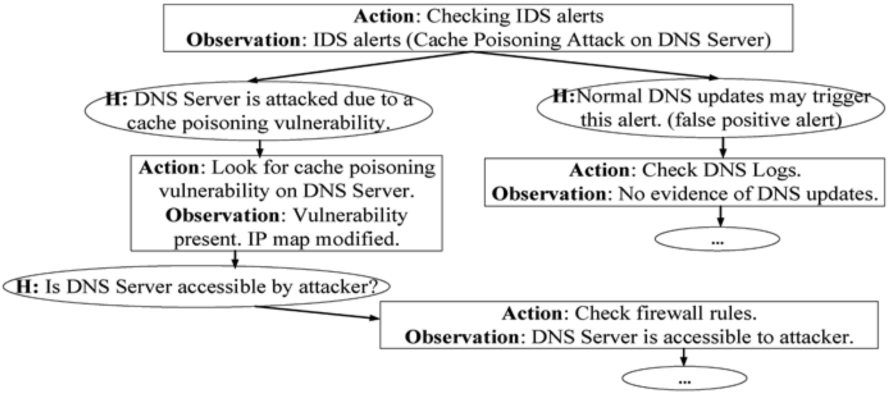

Action: Checking IDS alerts
Observation: IDS alerts (Cache Poisoning Attack on DNS Server)
H: DNS Server is attacked due to a cache poisoning vulnerability.
H:Normal DNS updates may trigger this alert. (false positive alert)
Action: Look for cache poisoning vulnerability on DNS Server.
Observation: Vulnerability present. IP map modified.
Action: Check DNS Logs.
Observation: No evidence of DNS updates.
...
H: Is DNS Server accessible by attacker?
Action: Check firewall rules.
Observation: DNS Server is accessible to attacker.
...

Fig. 5 An analytical reasoning process represented by the AOH objects

hypothesis of $H_{1.}$ A parent H is connected to its immediate children H showing a conjunctive AND relationship (i.e., refined sub-hypotheses). If an H H_1 and an H_2 have the same AO as their parent, we say that H_2 is a sibling hypothesis of $H_{1.}$ The sibling Hypotheses have disjunctive OR relationships (i.e., alternative hypotheses). Therefore, the AOH Objects in an analytical reasoning process are interconnected.

3.3 Capturing the Analytical Reasoning Processes

3.3.1 The Representation Indicates What Should Be Captured

We have proposed a model of analytical reasoning processes, which includes the AOH Objects and their relationships. The proposed model supports both a semi-structured representation of interconnected sense making constructs: Actions, Observations, and Hypothesis as well as an AND-OR organization of the Hypothesis. Actions and Observations can be captured in a structured representation, because the analysts' data exploration behaviors and the selected data can be automatically recorded. The Hypothesis constructs can be recorded in free text, which enables a flexible and analyst-friendly representation of analysts' thoughts.

An analyst could conduct various operations on the AOH Objects: the operations on Action could include filtering, searching, inquiring and data selecting; the operations on Observation could be selecting data entries and linking the data; the operations on Hypothesis could be creating a new Hypothesis, modifying an existing Hypothesis, switching the context and confirming/denying an existing Hypothesis. We will describe the operations in detail in Sect. 4. Therefore, we should also record the sequence of an analyst's operations on the AOH Objects in a temporal order.

3.3.2 Non-intrusive Capture

Regarding the importance of tacit knowledge and expertise, we capture the analytical reasoning processes of cyber analysis in a non-intrusive way. A monitoring system is developed to support the construction of AOH Objects, investigation and refinement of Hypotheses. This system audits the analysts' behaviors (e.g. data manipulation, hypothesis creation and refinement) and records them in traces, called "Cognitive Traces". This system would never interrupt the analysts. The Actions and related Observation are automatically tracked as the analysts selected data sources and specific entries of interest from each data source. When the analyst wishes to create a Hypothesis, the previously tracked Observations are automatically included in an initial list to be included as AO (i.e., the action-observation unit). The analyst can choose to modify the list to exclude data entries he/she looked at, but not relevant to the created Hypothesis. After the analyst confirms the captured AO to be associated with a Hypothesis, she/he is presented with a GUI interface to enter a short free-text description of the Hypothesis. Once the analyst completes the entering of Hypothesis description, the newly created AO and Hypothesis and their relationships are recorded to capture the analytic process of the cyber analyst. When the analyst wants to confirm a Hypothesis, he/she can mark the Hypothesis as “True”. Alternatively, the analyst can reject a Hypothesis by marking it as “False”.

3.4 Reasoning Processes in AOH Representation Can Be Extracted from the Cognitive Traces

Using the proposed representation, an analytical reasoning process is a process of evolving construction of AOH Objects, investigation and refinement of Hypotheses. Since the monitoring system has recorded the analysts’ behaviors of construction of AOH Objects, investigation and refinement of the Hypotheses, we can extract the analytical reasoning processes given the captured cognitive traces.

Figure 6 shows the framework of the proposed cognitive tracing analysis. The conceptual AOH model lays the cognitive foundation of our representation of analytical reasoning processes. This representation helps us to capture the analytical reasoning processes in a non-intrusive way. We can then extract the reasoning processes by analyzing the cognitive traces.

By analyzing the cognitive traces, which are generated by cyber analysts and gathered in a non-intrusive way, we can identify gaps and opportunities that lay the foundation for the next generation of cyber defense training, education, and development. More specifically, the results of analyzing analytical reasoning traces of cyber analysts will provide key insights about the differences of analytical reasoning between highly experienced analysts and less-experienced analysts so that opportunities to improve the training of analysts can be identified. Furthermore, the results of the trace analysis will demonstrate the feasibility and the opportunities about leveraging the experience of experienced analysts to support the analytical reasoning of less-experienced analysts. Another important benefit of the results of trace analysis is to

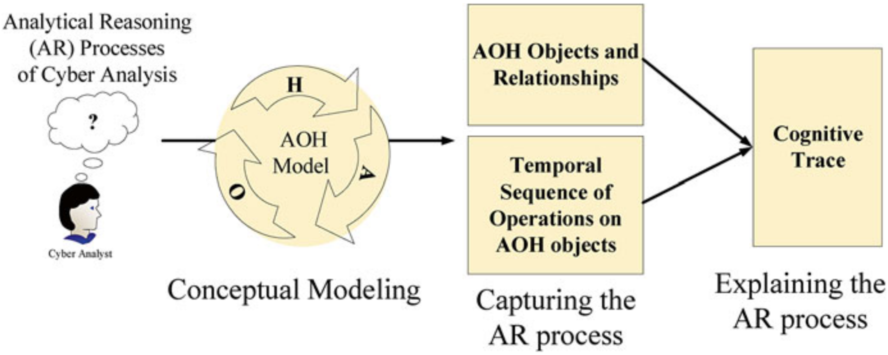

Analytical Reasoning (AR) Processes of Cyber Analysis
?
Cyber Analyst
H
AOH Model
O
A
Conceptual Modeling
AOH Objects and Relationships
Temporal Sequence of Operations on AOH objects
Capturing the AR process
Cognitive Trace
Explaining the AR process

Fig. 6 The framework of the cognitive tracing analysis

demonstrate the opportunities to improve the sharing and communication of knowledge regarding cyber attacks to the decision makers through a systematic construction of "story telling" using the traces. Finally, the results of the trace analysis involving multi-step attacks will identify opportunities for multiple analysts to collaborate and share forensic-sound information to facilitate near real-time cyber forensics to support "fight through" under an asymmetric information environment.

4 A Case Study about Professional Network Analysts

4.1 A Tool for Capturing the Cognitive Traces

We developed ARSCA (Analytical Reasoning Support Tool for Cyber Analysis) toolkit to track the traces of analysts' analytical reasoning processes while they are doing cyber analysis tasks. Figure 7 shows the architecture of ARSCA. ARSCA provides analysts with two main views: Data View and Analysis View. Data View integrates the monitoring data sources, for example, network topology, IDS alerts and firewall logs in this case. The Analysis View enables analysts to create instances of Action, Observation and Hypothesis (i.e. AOH Objects).

Figure 8 shows the interfaces of ARSCA. While an analyst is exploring the monitoring data, the tool automatically captures the activities of data manipulation (for example, searching and filtering) in an emerging Action instance, and also captures the selected data and other information resulting from the previous activities in the emerging Observation instance. ARSCA also enables an analyst to write down their thoughts as a Hypothesis instance and relate it to its corresponding Actions and Observations (Zhong et al. 2013).

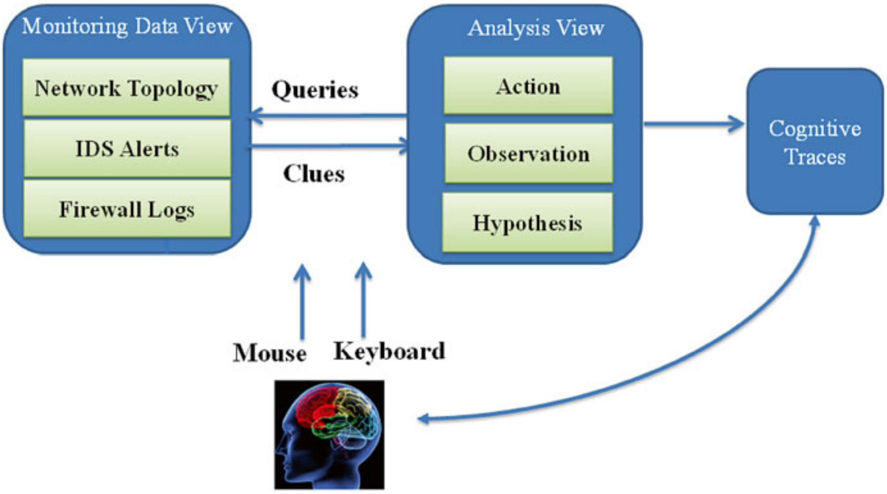
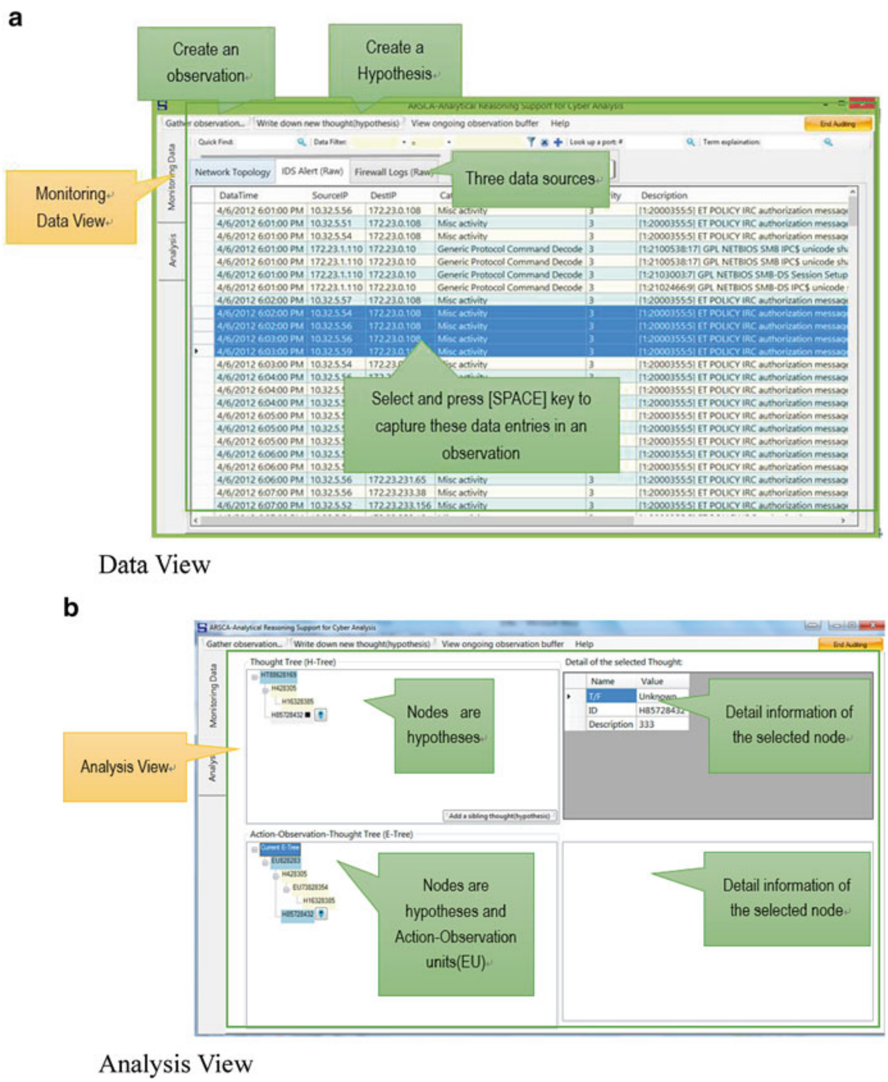


Fig. 7 The architecture of a cognitive trace capture tool


a
Create an observation
Create a Hypothesis
Monitoring Data View
Three data sources
Select and press [SPACE] key to capture these data entries in an observation
ARSCA-Analytical Reasoning Support for Cyber Analysis
Gather observation... | Write down new thought(hypothesis) | View ongoing observation buffer | Help
End Auditing
Quick Find: | Data Filter: | Look up a port # | Term explaination:
Network Topology | IDS Alert (Raw) | Firewall Logs (Raw)
Monitoring Data
Analysis

DataTime	SourceIP	DestIP	Cat	ity	Description
4/6/2012 6:01:00 PM	10.32.5.56	172.23.0.108	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:01:00 PM	10.32.5.51	172.23.0.108	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:01:00 PM	10.32.5.54	172.23.0.108	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:01:00 PM	172.23.1.110	172.23.0.10	Generic Protocol Command Decode	3	[1:2100538:17] GPL NETBIOS SMB IPC$ unicode sh
4/6/2012 6:01:00 PM	172.23.1.110	172.23.0.10	Generic Protocol Command Decode	3	[1:2100538:17] GPL NETBIOS SMB IPC$ unicode sh
4/6/2012 6:01:00 PM	172.23.1.110	172.23.0.10	Generic Protocol Command Decode	3	[1:2103003:7] GPL NETBIOS SMB-DS Session Setup
4/6/2012 6:01:00 PM	172.23.1.110	172.23.0.10	Generic Protocol Command Decode	3	[1:2102466:9] GPL NETBIOS SMB-DS IPC$ unicode
4/6/2012 6:02:00 PM	10.32.5.57	172.23.0.108	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:02:00 PM	10.32.5.54	172.23.0.108	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:02:00 PM	10.32.5.56	172.23.0.108	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:03:00 PM	10.32.5.56	172.23.0.108	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:03:00 PM	10.32.5.59	172.23.0.1	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:03:00 PM	10.32.5.54	172.23.0	sc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:04:00 PM	10.32.5.5	[illegible]	[illegible]	[illegible]	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:04:00 PM	10.32.5.5	[illegible]	[illegible]	[illegible]	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:04:00 PM	10.32.5.5	[illegible]	[illegible]	[illegible]	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:05:00 PM	10.32.5.5	[illegible]	[illegible]	[illegible]	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:05:00 PM	10.32.5.5	[illegible]	[illegible]	[illegible]	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:05:00 PM	10.32.5.5	[illegible]	[illegible]	[illegible]	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:06:00 PM	10.32.5.5	[illegible]	[illegible]	[illegible]	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:06:00 PM	10.32.5.5	[illegible]	[illegible]	[illegible]	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:06:00 PM	10.32.5.56	172.23.231.65	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:07:00 PM	10.32.5.56	172.23.233.38	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag
4/6/2012 6:07:00 PM	10.32.5.52	172.23.233.156	Misc activity	3	[1:2000355:5] ET POLICY IRC authorization messag

Data View

b
Analysis View
Nodes are hypotheses
Detail information of the selected node
Nodes are hypotheses and Action-Observation units(EU)
Detail information of the selected node
ARSCA-Analytical Reasoning Support for Cyber Analysis
Gather observation... | Write down new thought(hypothesis) | View ongoing observation buffer | Help
End Auditing
Monitoring Data
Analys
Thought Tree (H-Tree)
HT88628169
H428305
H16328385
H85728432
Add a sibling thought(hypothesis)
Action-Observation-Thought Tree (E-Tree)
Current E-Tree
EU828283
H428305
EU73828354
H16328385
H85728432
Detail of the selected Thought:

Name	Value
T/F	Unknown
ID	H85728432
Description	333

Analysis View

Fig. 8 The interface of ARSCA (Zhong et al. 2013) (**a**) Data view (**b**)Analysis view

4.2 *Conducting Human Studies for Collecting Cognitive Traces from Professional Network Analysts*

We conducted human studies with professional cyber analysts to gather their cognitive traces of the analytical reasoning processes. First of all, we needed to prepare the network monitoring data and the attack scenarios. We adopted the cyber analysis data of VAST 2012 Challenge Mini-challenge 2 (VAST Challenge 2012), including about 35,000 IDS alerts and 26,000,000 Firewall logs. Figure 9 shows the network

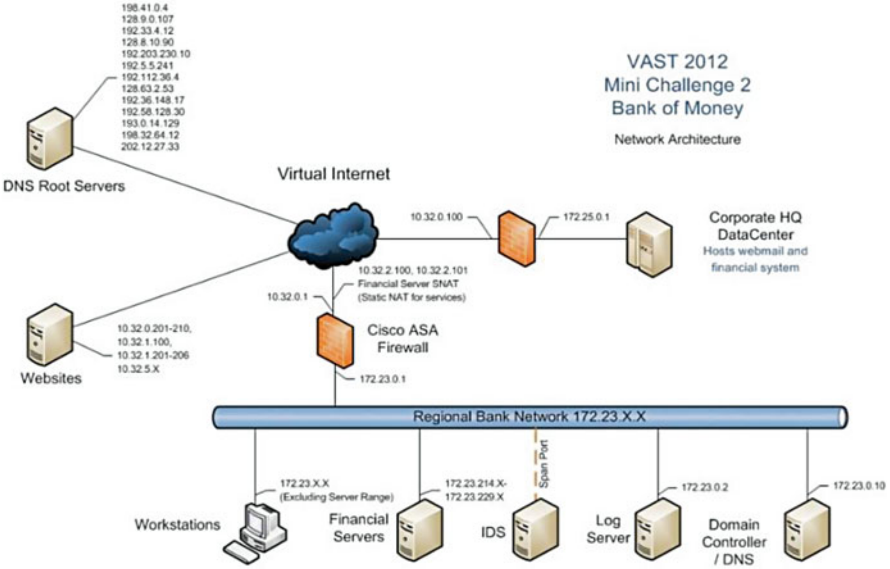


Fig. 9 The network topology of VAST 2012 Mini Challenge 2 (VAST 2012)

Table 1 Time period and size of dataset of the tasks

Task	Time period	Raw data size
1	4/5 20:18–20:30 (12 min)	IDS: 214 Firewall: 123,133
2	4/5 22:15–22:26 (11 min)	IDS: 239 Firewall: 115,524
3	4/6 0:00–0:10 (10 min)	IDS: 296 Firewall: 112, 766
4	4/6 18:01–18:15 (14 min)	IDS: 252 Firewall: 85,463

topology of the VAST 2012. This dataset implies a multi-step attack that took place over two days (about 40 h). Considering the fact that it is impossible for humans to process such large amounts of data without the help of external data analysis tools in a limited time, we cut out four pieces of the dataset which includes some key attack events and made four tasks using each of them. We made the tasks containing the same number of key attack events occurring in a similar amount of time, and containing a similar amount of network data. Table 1 shows the detailed information about the time period and dataset size of each task. Therefore, we can assume the tasks are at the same level of difficulty.

In collaboration with the U.S. Army Research Laboratory (ARL), the study recruited participants from professional network analysts working at ARL. In each task, analysts were asked to analyze the prepared network monitoring data with the goal of detecting the attack events. We also requested that the analyst use this tool to accomplish the analysis. Therefore, the tool would capture their analytical reasoning traces while they were doing the tasks.

Since the analysts are asked to use our tool, we provided a training session before each task and designed a quiz to test an analyst's proficiency of working with ARSCA. Each subject had to pass the quiz before he/she performs the task.

As a part of the experiment, we also ask subjects to respond to a pre-task questionnaire and a post-task questionnaire. The pre-task questionnaire contains questions about the demographic of the analyst, reasoning style, and the level of knowledge and skills regarding cyber analysis. The post-task questionnaire includes the analyst's retrospective summarization of the key findings and conclusions, as well as their assessment about the usefulness of the tool.

4.3 The Cognitive Traces

4.3.1 What Is in a Cognitive Trace?

Once an analyst completes his/her task, ARSCA generates the analyst's cognitive trace. In the rest of the chapter, we will use one of the subjects, S1, as an example to demonstrate in further detail the cognitive trace captured by ARSCA.

Figure 10 shows the AOH Objects created by subject S1 and their relationships. The ovals are the AO units and the rectangles are Hs. The text in an oval or a rectangle is the ID number for the AOH Object. We refer to the set of Hypotheses that are linked to the same AO (i.e. Action-Observation unit) as Alternative Hypotheses. For example, the Hypotheses in the dotted box in Fig. 10 are Alternative Hypotheses.

The operations on the AOH Objects are recorded in the cognitive trace in the temporal order they were performed by the analyst. Each item in a trace contains a timestamp and an operation on the AOH Objects. These operations can be grouped into three categories: (1) the operations related to Action (i.e. "AOP_Inquring", "AOP_Filtering", "AOP_Searching", and "AOP_Selecting"), (2) the operations

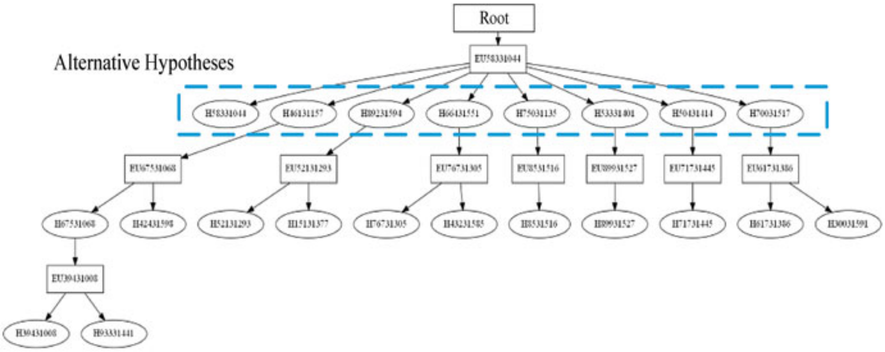


Fig. 10 The AOH objects and their relationships in S1's cognitive trace

Table 2 The description of operations

	Operation	Description
Operation on action	AOP_Filtering	Filter a data source
	AOP_Searching	Search a keyword in a data source
	AOP_Selecting	Select some data entries in a data source
	AOP_Inquiring	Inquire about a port or a term
Operation on observation	OOP_Selected	Generate an observation based on the selected data
	OOP_Linking	Link the selected data
Operation on hypothesis	HOP_New	Create a new hypothesis
	HOP_Add_Sibling	Add an alternative hypothesis
	HOP_SwitchContext	Switch the current focus of attention from one hypothesis to another hypothesis
	HOP_Modify	Modify the content of a hypothesis
	HOP_Confirm/Deny	Confirm/deny a hypothesis

related to Observation (i.e. "OOP_Selected" and "OOP_Linking"), and (3) the operations related to Hypothesis (i.e. "HOP_Confirm/Deny", "HOP_Modify", "HOP_SwitchContext", "HOP_Add_Sibling", and "HOP_New"). Table 2 summarizes these operations.

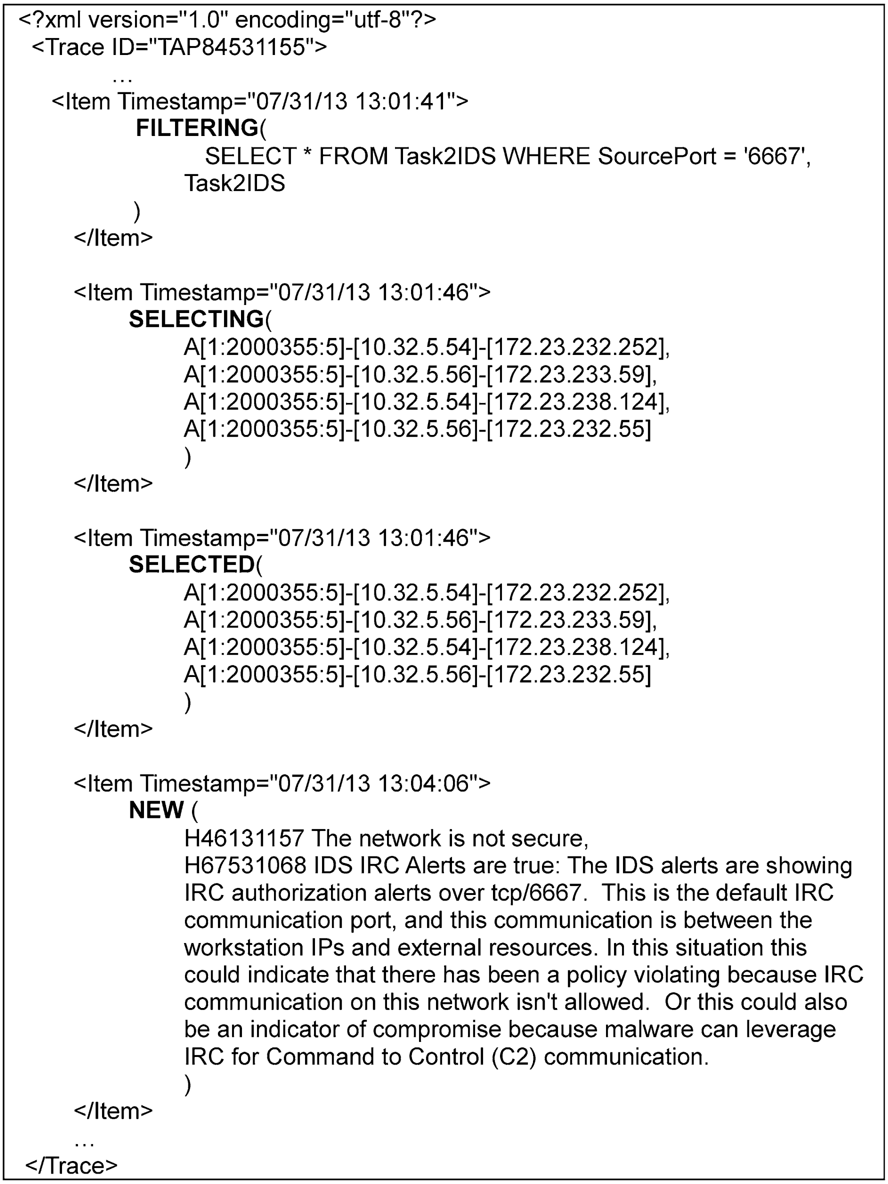
Figure 11 shows a portion of the file that records the cognitive trace generated by subject S1. Each item in the trace includes a timestamp and an operation. The operations in the trace items shown in Fig. 11 can be explained as follows.

- "FILTERING" (AOP_Filtering): Filtering the data source "Task2IDS" by the condition "SourcePort=6667".
- "SELECTING" (AOP_Selecting): Selecting the data entries in the filtered data set.
- "SELECTED" (OOP_Selected): The selected data entries. Such kind of operations always come in pairs with AOP_Selecting operations.
- "NEW" (HOP_NEW): Creating a new Hypothesis.

4.3.2 Cognitive Trace Analysis

We have conducted a preliminary analysis about the basic features of the collected cognitive traces from ten subjects, denoted by "S1", "S2", "S3", "S4", "S5", "S6", "S7", "S8", "S9", and "S10". Figure 12 shows the number of Action-Observation units and the number of hypotheses in the cognitive traces of these analysts, and the time they took to complete the cyber analysis task (based on VAST 2012). There is a significant differences among the analysts in terms of these three characteristics of their cognitive traces.

We further compared the number and the types of operations for the ten subjects in this case study. As shown in Fig. 13, there is a significant difference among the analysts both in terms of the number of operations and the type of operations



```
<?xml version="1.0" encoding="utf-8"?>
 <Trace ID="TAP84531155">
          ...
   <Item Timestamp="07/31/13 13:01:41">
           FILTERING(
                  SELECT * FROM Task2IDS WHERE SourcePort = '6667',
                Task2IDS
           )
     </Item>

     <Item Timestamp="07/31/13 13:01:46">
          SELECTING(
                A[1:2000355:5]-[10.32.5.54]-[172.23.232.252],
                A[1:2000355:5]-[10.32.5.56]-[172.23.233.59],
                A[1:2000355:5]-[10.32.5.54]-[172.23.238.124],
                A[1:2000355:5]-[10.32.5.56]-[172.23.232.55]
                )
     </Item>

     <Item Timestamp="07/31/13 13:01:46">
          SELECTED(
                A[1:2000355:5]-[10.32.5.54]-[172.23.232.252],
                A[1:2000355:5]-[10.32.5.56]-[172.23.233.59],
                A[1:2000355:5]-[10.32.5.54]-[172.23.238.124],
                A[1:2000355:5]-[10.32.5.56]-[172.23.232.55]
                )
     </Item>

     <Item Timestamp="07/31/13 13:04:06">
          NEW (
                H46131157 The network is not secure,
                H67531068 IDS IRC Alerts are true: The IDS alerts are showing
                IRC authorization alerts over tcp/6667.  This is the default IRC
                communication port, and this communication is between the
                workstation IPs and external resources. In this situation this
                could indicate that there has been a policy violating because IRC
                communication on this network isn't allowed.  Or this could also
                be an indicator of compromise because malware can leverage
                IRC for Command to Control (C2) communication.
                )
     </Item>
     ...
 </Trace>
```

Fig. 11 An example output file of S1's cognitive trace

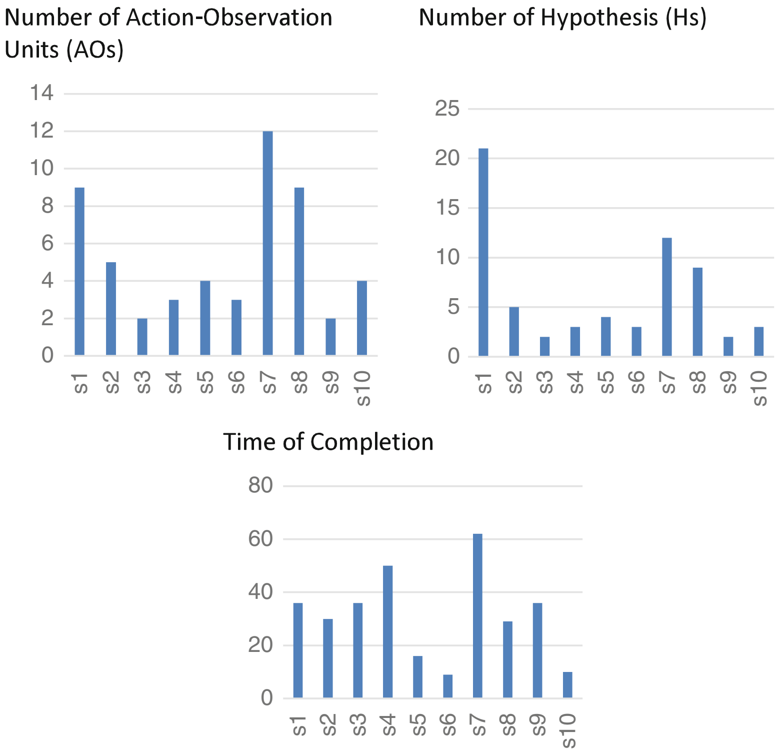


Fig. 12 The number of AOH's in the traces and task completion time

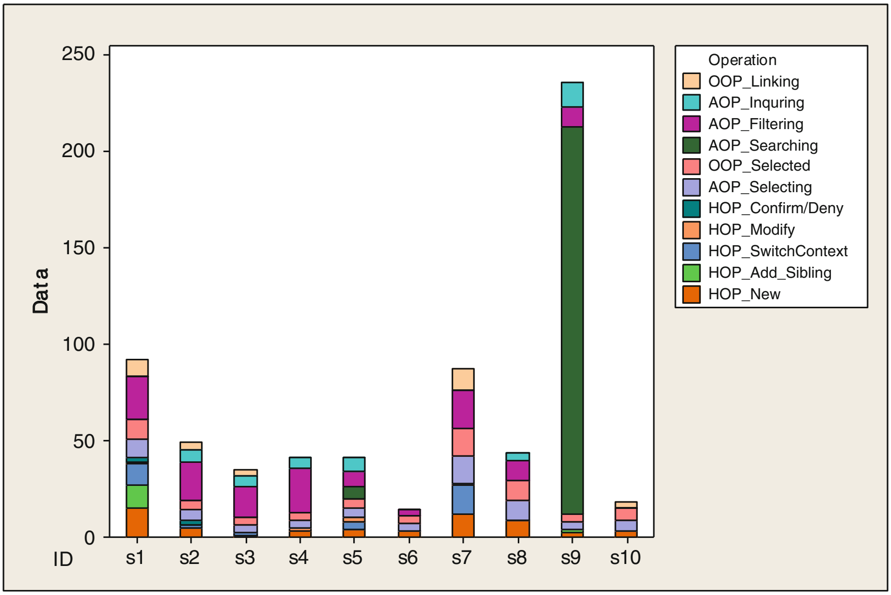


Fig. 13 Comparing the number of operations among ten Subjects (the operations are described in Table 2)

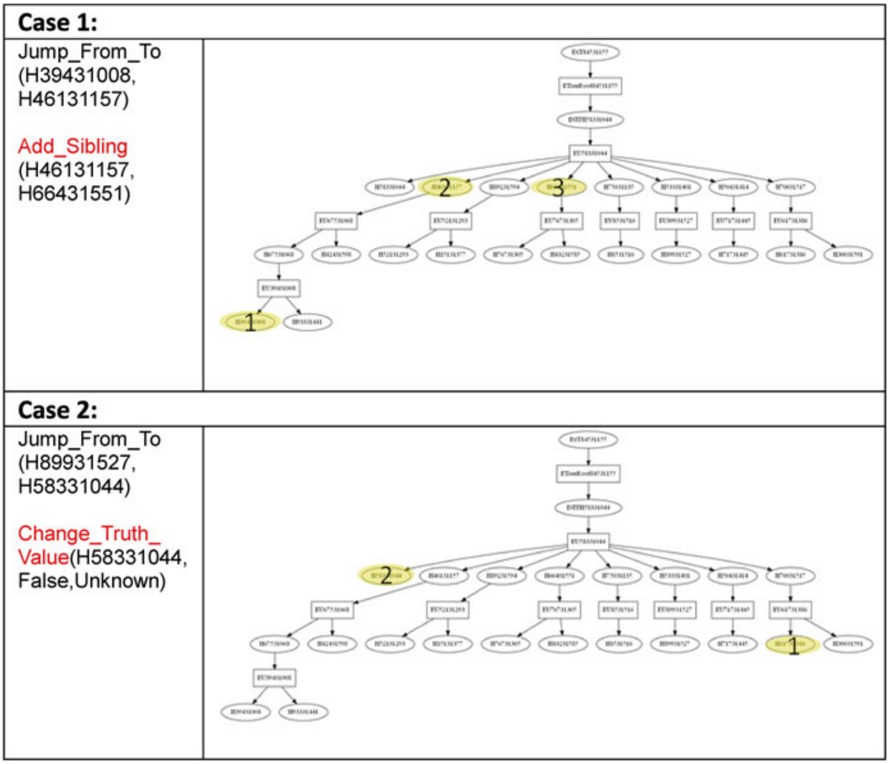


Fig. 14 Two cases of switching to previous hypothesis in S1's trace

they performed. This "heterogeneity" of the cognitive trace motivates us to further investigate to see whether there is any possible relationship between characteristics of cognitive traces and the performance of analysts. We will return to this point in the next section.

To gain a deeper understanding about the reasoning process of analysts, further analyses about the temporal ordering of these operations are also important. For example, switching context is an interesting aspect for trace analysis, because it may reveal the rationale and associated reasoning that enables the analyst to change focus of attention at a particular time in his/her reasoning process. We will use the trace of S1 to illustrate this: S1 switched context twice (shown in Fig. 14). The relevant trace segments are shown on the left of Fig. 14, and the AOH Objects (i.e. AOs and Hs) and their connections in S1's trace are shown on the right. In the first case of context switching, S1 jumped from Hypothesis "H39431008" (labelled "1") to "H46131157" (labelled "2"). Following this operation, S1 created a new Hypothesis "H666431551" (labelled "3") as a sibling Hypothesis of "H46131157". In the second case, S1 jumped from "H89931527" to "H58331044", and then change the truth value of "H58331044" from "Unknown" to "False" (i.e. rejecting it). Even though the analyst S1 switched contexts in both cases, the rationales are quite different. In the first case, S1 went back to a previous hypothesis to create an alternative hypothesis. In the latter case, he/she recalled a previous hypothesis to reject it. This example illustrates the importance in analyzing the temporal sequence of operations to obtain a richer understanding about the reasoning process of the analyst.

4.4 What Are the Characteristics of Cognitive Traces for Different Levels of Performance?

Since the pursuit of our research is to improve the analysts' performance in cyber analysis, we are interested in the analysts' performance in our tasks and the characteristics of cognitive traces for different levels of performance.

The ground truth of our tasks is known, which is the attack scenario of the VAST 2012 Challenge Mini Challenge 2. Therefore, we can evaluate the analyst's performance in a task based on how accurate his/her findings and conclusions are compared to the known ground truth. We conducted two rounds of evaluation to decide a final performance score for each subject, on a scale of 0–5 (with 5 being the best performance). Figure 15 shows the performance score of the ten subjects. Three analysts were rated highest (5 points), four analysts received 4 points, and three analysts were rated lowest (3 points).

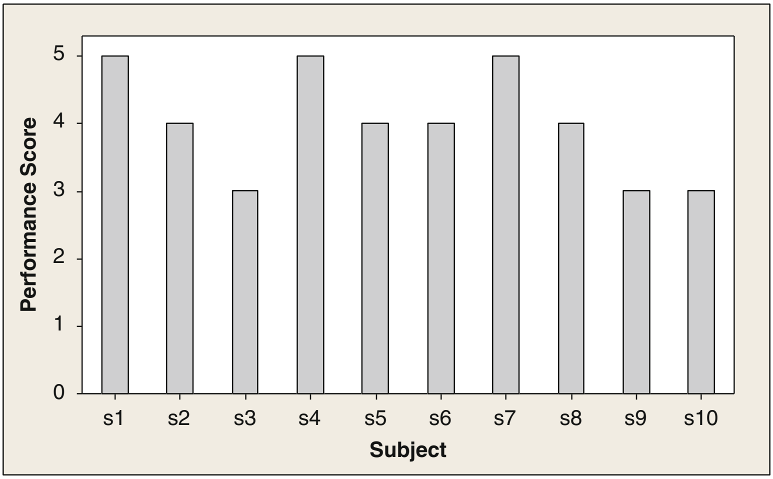


Fig. 15 The performance score of the ten subjects

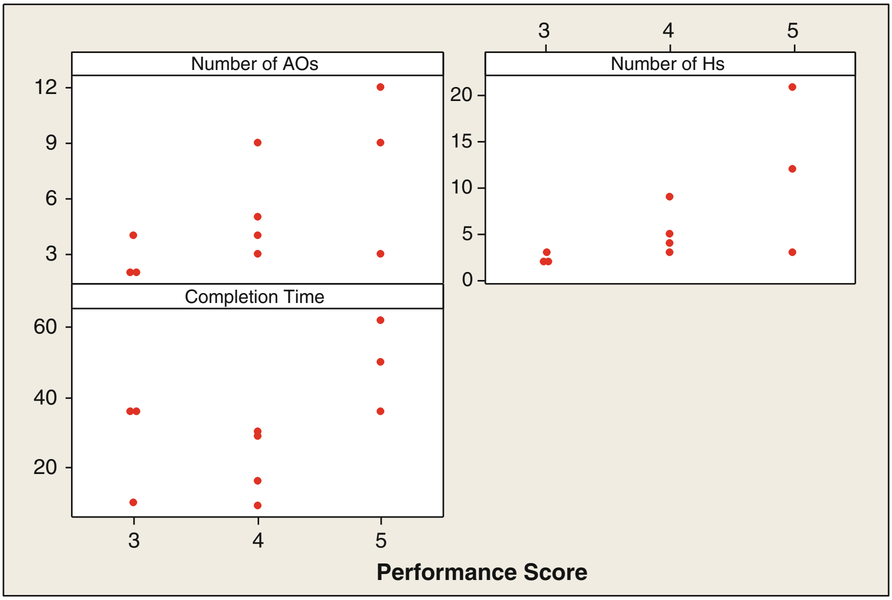


Fig. 16 The completion time and the number of A-O-H objects in the three groups of cognitive traces with different levels of performance scores

Next, we categorize the cognitive traces into three groups according to the performance score (that is, the traces with 3 points, 4 points and 5 points respectively), and investigate the characteristics of these traces in each group.

We first compare the completion time and the number of AO units and hypotheses for analysts with different levels of performance (i.e. 3 points, 4 points, and 5 points). Figure 16 shows that traces in the lowest performance group have, on average, the smallest number of AO units and hypotheses in their traces. The task completion time for the group with the best performance is also larger, on the average, than the completion time of those from the other two groups. While we are not able to arrive at conclusions about the relationship between analyst performance and the characteristics of their traces due to the small sample size of the analysts, these preliminary findings do suggest that further studies are warranted to further investigate potential relationships between analyst performance and the characteristics of their traces.

Using a similar strategy, we want to investigate whether the number of operations for each operation type is related to analyst performance in some way. Figure 17 shows the result of this comparison. The group of high performance analysts, on the average, uses more filtering operations (AOP) than the two other groups. They also tend to do more context switching (HOP SwitchContext) than the others. Finally, the high performance group performs more linking operations among selected observations (OOPLinking). As we mentioned before, more samples and further studies are needed to investigate whether these detailed trace characteristics are correlated with analyst performance in a statistical significant way. These preliminary results, however, do suggest that comparing the characteristics of cognitive traces of analysts with their performance is a promising direction of future research.

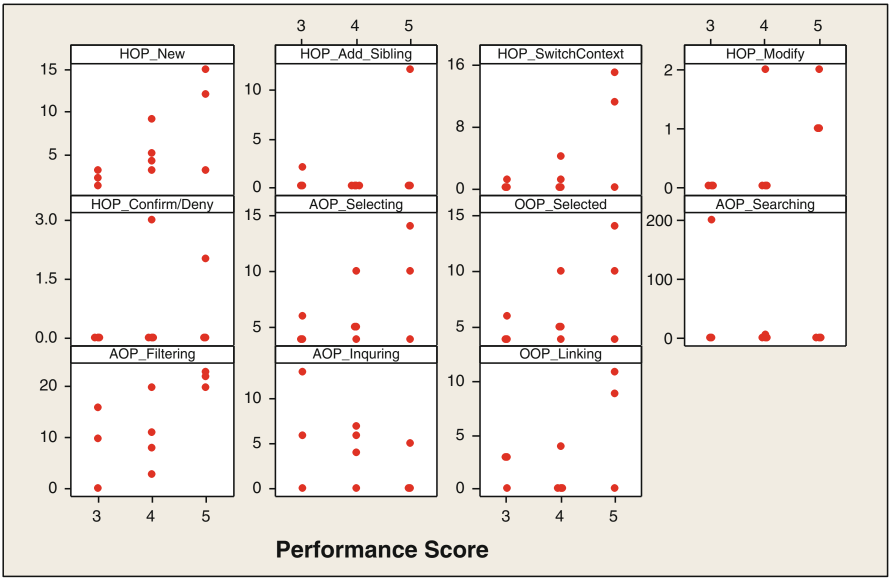


Fig. 17 The number of different types of operations in the cognitive traces at different levels of performance score

5 Summary

As computing devices connected to the internet explode for personal health monitoring and management, environment and physical security surveillance, smart home appliances, smart vehicles, smart energy grid, and ubiquitous computing (e.g., Google Glass), the complexity and the frequency of cyber-attacks faced by cyber defense analysts of governments and business enterprises continue to increase at a rapid speed. The ultimate goal of cyber defense is to increase its agility even for zero-day attacks (e.g., attacks leveraging vulnerabilities that are not known by the cyber defenders), so that the time from detection of attacks to creating automated support tool to enable early and effective detection of future similar attacks is as close to real-time as possible. A critical obstacle on the path to achieving this vision is lacking a systematic framework and supporting methods/tools for capturing the analytical reasoning process of professional cyber defense analysts.

In this chapter, we have described the current understanding about the high-level cognitive process of cyber analysts, based on Cognitive Task Analysis (CTA) conducted previously with professional cyber defense analysts, and the difficulty of capturing fine-grained cognitive reasoning process of analysts using existing methods. To address this difficulty in a way that is well-founded theoretically (for high generality) and, in the same time, practically feasible for being embedded into the work environment of analysts in a “non-intrusive” way, we present a paradigm, we believe, that has a potential to create transformational impacts toward a much more agile cyber defense in the near future. We summarize below the key features of the framework and how they can contribute to enhancing the agility of cyber defense.

First, the sense making cognitive theory foundation of the A-O-H conceptual model enables the framework to be general and broadly applicable to a wide range of tasks and domains. The notion of actions, observations, and hypotheses naturally map to the observable actions performed by the analysts, observations from immense data presented to the analysts, and their hypothesized attack step, sequence, and/or plan. Because the framework is built on the A-O-H model, it can be applied not only to intrusion detection at the tactical level as demonstrated in the case study, but also to other types of tactical cyber analysis tasks (e.g., forensics) as well as to cyber defense tasks at the strategic level. In fact, the framework can also be applied to other domains such as intelligence analysis.

Second, the non-intrusive nature of the framework enables the capturing of the cognitive process to be embedded in the work environment of the professional analysts. The system audits the analysts’ behaviors (e.g. data manipulation, hypothesis creation and refinement) and records them in “Cognitive Traces” without interrupting the analysts. The Actions and related Observation are automatically tracked as the analysts selected data sources and specific entries of interest from each data source. When the analyst wishes to create a Hypothesis, the previously tracked Observations are automatically included in an initial list to be included as AO (i.e., the action-observation unit). The non-intrusive capturing of cognitive trace is a key enabler toward a more agile defense because it enables the cognitive trace to be captured at the earliest possible time, and significantly reduce the time and the cost (e.g., due to extra efforts the analysts need to make) it may take to extract reasoning process from the analysts otherwise.

Third, the cognitive traces captured in non-intrusive way, as demonstrated by the case study, provide, for the first time, important characteristics of the reasoning process of analysts and their potential relationship to the performance of analysts. These characteristics and relationship offers promising indication that the analysis of the reasoning process (both at the individual level and at the aggregate level) can be beneficial to the design of training programs and cognitive aids for enhancing the performance of analysts (Zhong et al, 2014).

In summary, this book chapter presents a theoretically well-founded and practical non-intrusive framework for capturing and analyzing the cognitive reasoning processes of professional cyber analysts. It provides important basis for further studies regarding collaboration among analysts (e.g., in two adjacent work shifts), visualization needs and design for supporting analysts, cognitive aids, and training procedures that leverage the reasoning processes captured to assist analysts to perform the cyber defense analysis at hand with higher quality and more efficiency.

References

Aamodt, A. and Plaza, E. (1994) “Case-based reasoning: foundational issues, methodological variations, and system approaches.” *AI Commun.* 7, 1, 39–59.

Bell, J., and Hardiman, R. J. (1989) “The third role – the naturalistic knowledge engineer”, in *Knowledge elicitation: Principles, Techniques, and Applications*, Dan Diaper (ed.), John Wiley & Sons, New York.

Biros, D., and Eppich, T. (2001) Human Element Key to Intrustion Detection, Signal, p. 31, August.

Boriah, S., Chandola, V., Kumar, V.: (2008) Similarity measures for categorical data: A comparative evaluation. In: SDM, pp. 243–254. SIAM, Philadelphia.

Crandall, B., Klein, G., and Hoffman, R. (2006). *Working minds: A practitioner's guide to cognitive task analysis.* MIT Press.

Cunningham, P., (2008) “A Taxonomy of Similarity Mechanisms for Case-Based Reasoning,” University College Dublin, Technical Report UCD-CSI-20080-11, January 6.

D’Amico, A., Whitley, K., Tesone, D., O’Brien, B., and Roth, E., (2005) Achieving Cyber Defense Situational Awareness: A Cognitive Task Analysis of Information Assurance Analysts, in Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, 229–233.

D’Amico, A. and Whitley, K. (2008) “The Real Work of Computer Network Defense Analysts,” *VizSEC 2007: Proceedings of the Workshop on Visualization for Computer Security*, Springer-Verlag Berlin Heidelberg, pp. 19–37.

De Mantaras, R. L., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Faltings, B., Maher, M. L., Cox, M. T., Forbus, K., Keane, M., Aamodt, A., and Watson, I. (2005) Retrieval, reuse, revision and retention in case-based reasoning. *Knowl. Eng. Rev.* 20, 3 (September 2005), 215–240.

Doyle, D. (2005) “A Knowledge-Light Mechanism for Explanation in Case-Based Reasoning,” University of Dublin, Trinity College. Department of Computer Science, Doctoral Thesis TCD-CS-2005-71.

Durkin, J. (1994), “Expert Systems: Design and Development”, Mamillan, New York, NY.

Erbacher, R. F. and Hutchinson, S. E. (2012) “Extending Case-based Reasoning to Network Alert Reporting”, in *Proceedings of 2012 International Conference on Cyber Security*, pp. 187–194.

Erbacher, R. F., Frincke, D. A., Wong, P. C.,Moody, S. J., Fink, G. A. (2010a) A multi-phase network situational awareness cognitive task analysis, Information Visualization 9(3): 204–219.

Erbacher, R. F., Frincke, D. A., Wong, P. C.,Moody, S. J., Fink, G. A, (2010b) Cognitive task analysis of network analysts and managers for network situational awareness. VDA 2010: 75300

Ericsson, K. A. and Simon, H. A., (1980) “Verbal reports as data”, Psychological Review, 87 (3), pp. 215–251.

Ericsson, K. A. and Simon, H. A., (1993) “Protocol analysis”, MIT Press, Cambridge, MA.

Foresti, S. and Agutter, J., “Cognitive Task Analysis Report,” University of Utah, CROMDI. Funded by ARDA and DOD.

Kolodner, J. (1983) “Reconstructive Memory: A Computer Model,” *Cognitive Science* 7 (4), pp. 281–328.

Lebowitz, M. (1983) “Memory-based parsing,” *Artificial Intelligence* 21, 4, pp. 363–404.

Long, J., Stoecklin, S., Schwartz, D. G., and Patel, M., (2004) “Adaptive Similarity Metrics in Case-based Reasoning,” *The 6th IASTED International Conference on Intelligent Systems and Control* (ISC 2004), August 23–25, Honolulu, Hawaii, pp. 260–265.

Osborne, H. and Bridge, D., (1997) “Models of Similarity for Case-Based Reasoning,” *Proc. Interdisciplinary Workshop Similarity and Categorisation*, pp. 173–179.

Ranganathan A., and Ronen, R. (2010) “Information-Theory Based Measure of Similarity Between Instances in Ontology,” International Business Machines Corporation, United States Patent #7,792,838 B2.

Sanderson, P., Scott, J., Johnston, T., Mainzer, J., Watanabe, L., and James, J., (1994) “MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA)”, *Int. J. Human-Computer Studies*, 41, pp. 633–681.

Schank, R., (1982) *Dynamic Memory: A Theory of Learning in Computers and People* (New York: Cambridge University Press.

Soh, L. K., and Blank, T. (2008) "Integrating Case-Based Reasoning and Meta-Learning for a Self-Improving Intelligent Tutoring System. *Int. J. Artif. Intell. Ed.* 18, 1, 27–58.

Stahl, A. (2004) Learning of Knowledge-Intensive Similarity Measures in Case-Based Reasoning. PHD-Thesis, dissertation.de, Technische Universität Kaiserslautern.

Sterling, W. M., and Ericson, B. J. (2006) "Case-Based Reasoning Similarity Metrics Implementation Using User Defined Functions," NCR Corp., United States Patent # 7,136,852 B1, Nov. 14.

Sun, Z., Finnie, G., and Weber, K. (2005) "Abductive Case Based Reasoning," *International Journal of Intelligent Systems*, 20(9), 957–983.

Wang, H. and Dubitzky, W., (2005) "A flexible and robust similarity measure based on contextual probability." In *Proceedings of the 19th international joint conference on Artificial intelligence* (IJCAI'05). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 27–32.

Zhong, C., Kirubakaran, D. S., Yen, J. and Liu, P., (2013) "How to Use Experience in Cyber Analysis: An Analyt-ical Reasoning Support System," in Proc. of IEEE Conf. on Intelligence and Security Informatics (ISI), pp. 263–265.

Zhong, C., Samuel, D., Yen, J., Liu, P., Erbacher, R., Hutchinson, S., Etoty, R., Cam, H., and Glodek, W. (2014) "RankAOH: Context-driven Similarity-based Retrieval of Experiences in Cyber Analysis," in Proceedings of IEEE International Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA 2014) pp. 230–236.

VAST Challenge 2012 http://www.vacommunity.org/VAST+Challenge+2012