A Diagram Method to Analyze Illogical Thinking

Modeling Typical Logical Mistakes concerning Information Security and Industrial Safety

Toru Nakata
Research Institute for Secure Systems
AIST
Tsukuba, 305-8568, Japan.
toru-nakata@aist.go.jp

Hajime Watanabe
Research Institute for Secure Systems
AIST
Tsukuba, 305-8568, Japan.
h-watanabe@aist.go.jp

***Abstract*—we propose a graphical method to express and analyze logical failure of human operators. Human factor is the weakest link on information security and industrial safety, and mare human mistakes have been making many severe accidents. Human thinking process is unstable and not always correct. There exist certain tendencies of cognitive misunderstanding, and people easily commit mistakes of such patterns. In the scene of computer crime, many attackers use those cognitive biases to deceive their victims. The cognitive biases are also the generator of critical misjudgments that brought severe industrial accidents. We nominate the most frequent patterns of irrational processing in order to identify the mechanisms of accidents on computer security and industrial safety. We can decompose processes of misjudgments into easy steps that reflect popular patterns of the cognitive biases by using the graphical method. In this paper, we try to reason patterns of thinking failures that took place in real cyber-attacks and an industrial accident.**

Keywords-logical fallacy; cyber attack; social engineering; information security; human factors

I. Introduction

Human operators in industries are the weakest link on safety. People often commit mistakes when the situation is confusing. It is true in cyber security too. Cyber attackers can easily deceive people by using *social engineering* strategies [1][2].

Human operators can cause serious wrong decisions that result in the worst accidents. Compared to machines, human operators have broader and stronger rights to control the situation, so wrong judgments made by the human operators can be critical.

In this paper, we propose a graph method for step-by-step analysis of human logical mistakes.

In traditional safety engineering, researchers consider illogical thinking as an important factor. However, there remain some difficulties to study the origin of mistakes. Some researches try to consider human operators as logical processing units [3]. Such researches explain human mistakes are results from overload on human thinking process. However, the excess of mental loads is not the sole cause of cognitive mistakes since cool-headed people can also commit mistakes on logical thinking.

Other researchers explain the origin of the human misconceptions by proposing cognitive models, such as category-based induction [4], fuzzy logic [5], and logical fallacy phenomena discussed in behavioral economy [6]. In the field of child education, Van Lehn has listed up children's common logical mistakes in solving calculus problems [7]. Using those lists of mistake patterns, we can predict human mistakes and estimate the risk of accident on human-machine systems.

However, it is quite difficult to use them to explain real accidents caused by human logical failure. The processes of wrong thinking are complicated and hard to describe with ordinary words.

The aim of this research is to present a clear and consistent method to explain human logical fallacies.

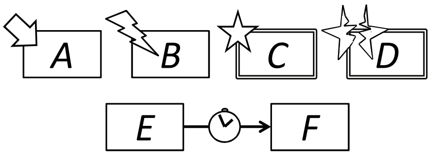
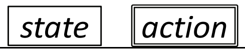
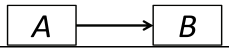
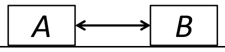
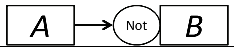
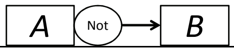
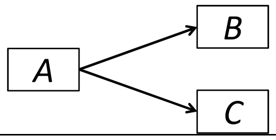
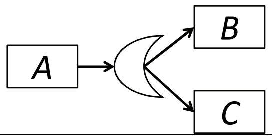


Figure 1. Auxiliary expressions standing for the following meanings: Proposition *A* is found true. Reliability of *B* is fabricated. Action *C* is carried out. Action *D* is not carried out. *E* affects *F* after a while.

TABLE I. PROPOSED GRAPH EXPRESSIONS OF LOGIC.

Formula expression	Graphic expression
State and action	
$A \Rightarrow B$	
$A \Leftrightarrow B$	
$A \Rightarrow \neg B$	
$\neg A \Rightarrow B$	
$(A \Rightarrow B) \wedge (A \Rightarrow C)$	
$(A \Rightarrow B) \vee (A \Rightarrow C)$	

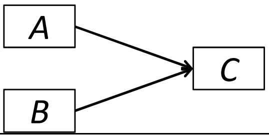
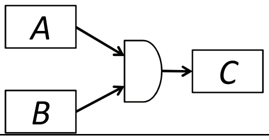
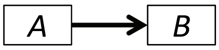
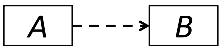
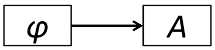
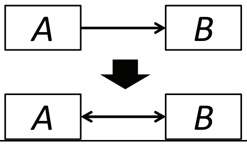
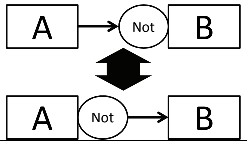
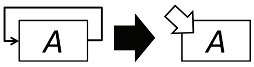
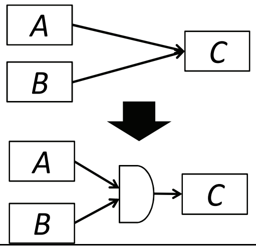
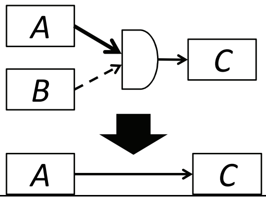
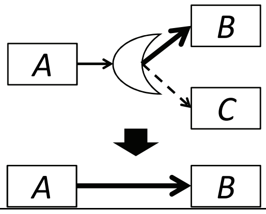
$(A \Rightarrow C) \wedge (B \Rightarrow C)$	
$(A \wedge B) \Rightarrow C$	
"$A \Rightarrow B$" is probably true.	
"$A \Rightarrow B$" is hardly true.	
A is an axiom. (A is unconditionally true.)	

TABLE II. TRANSFORMATION OF COMMON HUMAN LOGICAL FALLACIES.

Affirming the consequent	
Negation fallacy	
Misunderstanding a self-justifying tautology as true	
Conjunction fallacy	
Ignoring hidden factors on an unbalanced AND-gate	
Hasty affirming a disjunction on an unbalanced OR-gate	

II. Visual Expression of Logic and Misconception.

Graphical expressions are suitable for explaining and understanding complex process of human misconception. We adopt graphical expression of logic gates to explain human thinking (Table I) with auxiliary expressions (Fig. 1).

There are already several methods to express logic graphically. The *semantic tableau* (or *truth tree*) is the most ordinary way to organize and connect propositions as graphs. In the field of the software industry, we often need a large volume of arguments to verify the correctness of software, so the visualization methods are strongly required to understand the logic. The fault tree analysis (FTA) and its variations have been used for software verification [8][9].

The method that we propose expresses logical relationships to describe logical fallacy of human reasoning in graph structure.

Thanks to the precedent studies on human logical fallacy, we can list the most typical ways to convert proper logic into wrong thinking. We nominate the six most common patterns of the irrational transforms shown in Table II. The six patterns are not complete but practically enough to explain most of the real accidents of industries and information security.

A. Analysis of Phone fraud

The tremendously popular swindle in Japan is the phone fraud starting the conversation with the phrase of "Hello! It's me!" This trick is to pretend the real son. So many elderly people have mistaken the malicious speaker for real children. The swindlers deceive the elders to send a huge amount of money for them eventually.

At the beginning of this swindle, the victims can have four correct propositions. (Fig. 2 is the diagrammatic expression of the four propositions.)

P1: the speaker on the phone is the son or another person.
P2: the parent should respect the son's request.
P3: the son's voice is a male voice.
P4: other people may speak in a male voice.

Then, the victims tend to convert those propositions in wrong ways.

P1-transformed: the speaker on the phone is possibly the son, according to experience. (Transformation of *hasty affirming a disjunction.*)
P3-transformed: the son's voice is a male voice, so the voice is spoken by the son. (Transformation of *affirming the consequent.*)
P4-transformed: ignore the possibility that a stranger is speaking. (Transformation of *hasty affirming a disjunction.*)

After those wrong transformations, the victims obey the swindler since they have a wrong mindset described as Fig. 3.

We can say that this phone fraud mainly consists of tactics based on statistical unbalance. People have a psychological tendency named *normalcy bias*. We often prejudge the present event is normal and similar to uneventful situations in personal experience. This bias makes us reduce vigilance, so it will be comfortable for us if the occurring event is harmless. However, the swindlers can use it as a weapon against security.

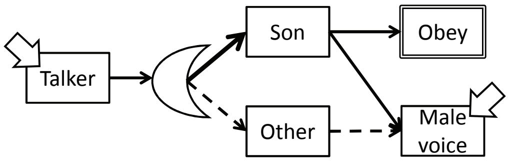


Figure 2. The correct mindset of the victims in the beginning of the phone fraud pretending the real son.

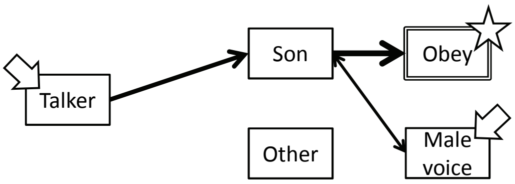


Figure 3. Typical but wrong transformation of the mindset of Fig. 2.

If we analyze the process described above only with words or logical formula expression, it would become rather complicated. The graph expression allows us to analyze clearer.

B. *Analyses of Common Cyber Attack Strategies*

Let us see more examples of analysis with the graph expressions.

1) Pretexting technique

Adding pretexts in the story can gild false statements. The bias of *conjunction fallacy* makes us regard statements as true when many words about contexts and circumstance accompany the statements. *Pretexting* is one of such techniques, and it commonly used for fraud including cyber attacks (Fig. 4).

2) Tautological validation

Trap email claim itself as legitimate mail. Typically, such mail says that it is an alert issued by the security authority. It is a self-validating tautological message, and there is no evidence to support the correctness. However, it is enough to fool people in some cases.

3) Baiting technique

The attackers often make trap mail with messages suggesting that the attachment file provides attractive information for the recipient. It would drive the recipient to open the attachment file, and the malware in it can work.

4) Example of technique combination

We analyzed the scheme of trap email that contains an attachment of malware. In the scheme of this attack, the adversary makes the recipient open the attachment file by the following strategies (Fig. 5).

1. The adversary usually attaches a malware file of commonplace and non-executable format (such as Adobe PDF and Microsoft Word) to make the recipient less cautious.
2. The adversary writes messages of *baiting* or *tautological validation*.
3. The adversary may fabricate the senders address as safe senders, which imply safety of the mail. The cognitive bias of “*hasty affirming a disjunction*” makes the recipient mishandle the or-gate separating safe email and trap email, which is badly balanced. Eventually, they judge the mail safe.

C. *Analysis of Critical Human Machine Interaction*

The proposed graph expression technique is also useful to analyze industrial accidents. For example, we can arrange of the process of the airplane crash of Air France 447 on 2009 as Fig. 6. According to the accident report [10], the pilot controlled the pitch angles up and down inconsistently; meanwhile the airplane have been losing its speed and facing the danger of stall.

The system emitted the stall warning to the pilot, but the pilot did not strongly consider the risk of stall. It is curious, and there must have been unfortunate structure to fail the communication between the pilot and the system.

Because of the inertia of the airplane, there was some delay between the timings of pitch control and emitting the warning. It was terribly misleading that the warning appeared just after the pilot made right actions. Lack of synchronism makes logical relationships complex. The delayed alarms confused the pilot trying to grasp the logical relationship of the circumstance. If the pilot made the pitch angle flat and increased the propulsion power consistently, the airplane would have avoided stall.

We think that the cognitive bias of *ignoring hidden factors* was the basic mechanism of the pilot’s failure. During normal level flight, controlling airplane’s heading with the rudders is easy and commonplace, so it has priority over adjusting thrust with controlling of the engines. Therefore, the pilot may be reluctant to control the engine power, and the thrust issue becomes a hidden factor.

III. Conclusion and Future Work

We proposed a method to analyze human logical failure by using gate-logic-like graphical way. The merit of the proposed method is as the following:

- It can express complicated connections of propositions.
- We can deal with the typical patterns of logical mistakes as conversion of graph connection.
- The diagrammatic method is easy to analyze the weakest points of the logical structure. On the diagram, we can discuss the easiest transformation to bring wrong conclusion.

Our work is in progress, and those analysis examples were processed by hands. In the future work, we try to create automatic analysis system to check the cognitive flows. We will calculate probabilities of that people commit the typical logical fallacy patterns shown in Table II, and then we will be able to evaluate the weakest points on the logic and amount of the risks quantitatively.

References

[1] K. Mitnick and W. Simon, The Art of Deception, John Wiley & Sons, 2002.

[2] D. Kennedy, J. O'Gorman, D. Kearns, and M. Aharoni, Metasploit: The Penetration Tester's Guide, No Starch Press, 2011.

[3] C. Wickens, "Information processing, decision-making, and cognition," G. Salendy (Ed.), Handbook of Human Factors, John Wiley & Sons, 1987, pp.72-107.

[4] D. Osherson, E. Smith, O. Wilkie, A. Lopez, and E. Shafir, "Category-Based Induction," Psychological Review, vol. 97, No. 2, 1990, pp.185–200.

[5] L. Zadeh, et al. Fuzzy Sets, Fuzzy Logic, Fuzzy Systems, World Scientific Press, 1996.

[6] G. Belsky, and T. Gilovich, Why Smart People Make Big Money Mistakes, Simon & Schuster, 1999.

[7] K. Van Lehn, Mind Bugs: The Origins of Procedural Misconceptions, MIT Press, 1990.

[8] R. Weaver, J. Fenn, and Tim Kelly, "A Pragmatic Approach to Reasoning about the Assurance of Safety Arguments," Proc. 8th Australian Workshop on Safety Critical System and Software, 2003.

[9] L. Wouters and M.-P. Gervais "xOWL an Executable Modeling Language for Domain Experts," IEEE International Enterprise Distributed Object Computing Conference, 2011.

[10] Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, Final Report On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro–Paris, 2012.

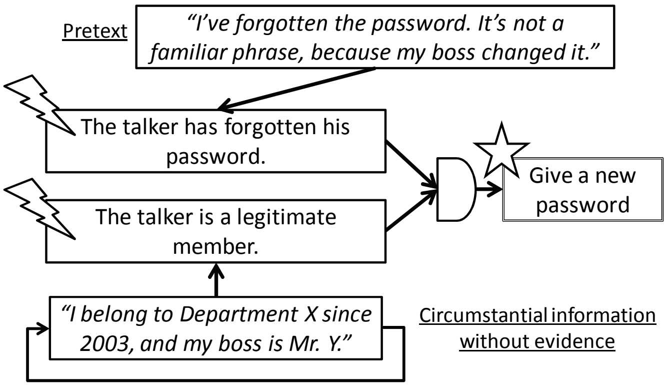

Pretext
"I've forgotten the password. It's not a familiar phrase, because my boss changed it."
The talker has forgotten his password.
The talker is a legitimate member.
Give a new password
"I belong to Department X since 2003, and my boss is Mr. Y."
Circumstantial information without evidence

Figure 4. A common scheme to steal password with *Pretexting* technique.

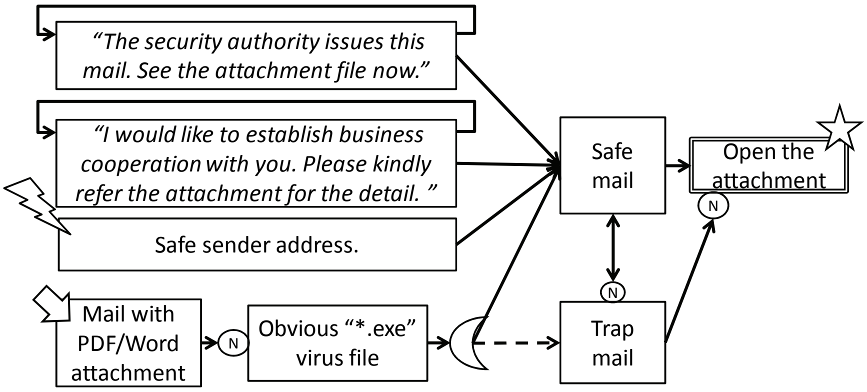

"The security authority issues this mail. See the attachment file now."
"I would like to establish business cooperation with you. Please kindly refer the attachment for the detail. "
Safe sender address.
Safe mail
Open the attachment
N
Mail with PDF/Word attachment
N
Obvious "*.exe" virus file
N
Trap mail

Figure 5. A typical mechanism of trap email.

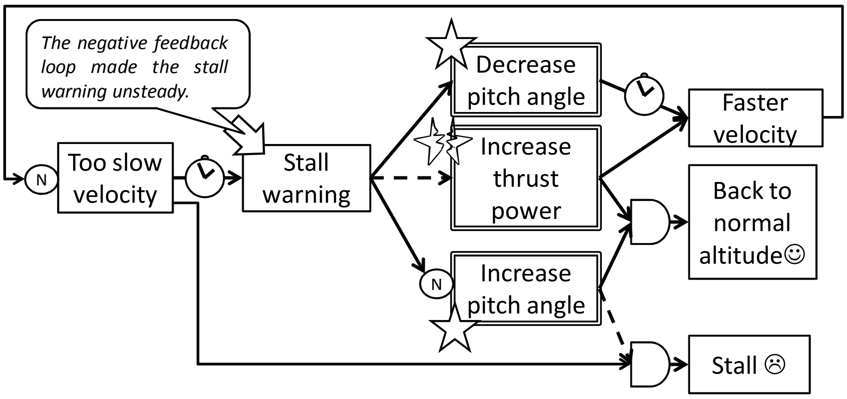


Figure 6. Estimated the pilot's mindset at the crash of Air France 447 on 2009. The broken arrows are causal relationships ignored by the pilot. This feedback loop with delays cased intermitted and unstable stall warning that resulted in contradicting control of the airplane's nose up and down.