A Game Theoretic Framework for E-Mail Detection and Forgery Analysis

Long Chen
College of Computer Science and Technology
Chongqing University of Posts and Telecommunications
Chongqing, China

Min Xiao
College of Computer Science and Technology
Chongqing University of Posts and Telecommunications
Chongqing, China

Yuan Lou
College of Computer Science and Technology
Chongqing University of Posts and Telecommunications
Chongqing, China

Zhen-Xing Dong
College of Computer Science and Technology
Chongqing University of Posts and Telecommunications
Chongqing, China

***Abstract*—In email forensic, the email detection and forgery conflict is an interdependent strategy selection process, and there exists complex dynamics between the detector and the forger, who have conflicting objectives and influence each other's performance and decisions. This paper aims to study their dynamics from the perspective of game theory .We firstly analyze the email basic structure and header information, then discuss the email detection and forgery technologies. In this paper, we propose a Detection-Forgery Game (DFG) model and make a classification of players' strategy with the Operation Complexity (OC). In the DFG model, we regard the interactions between the detector and the forger as a two-player, non-cooperative, non-zero-sum and finite strategic game, and formulate the Nash Equilibrium. The optimal detection and forgery strategies with minimizing cost and maximizing reward will be found by using the model. Finally, we perform empirical experiments to verify the effectiveness and feasibility of the model.**

Keywords—email detection; email forgery; game theoretic model; Nash Equilibrium; the optimal strategy

I. Introduction

E-mail is ubiquitous in the contemporary commercial environment. Because of its convenience, low cost and rich content, it becomes one of the most widely used applications for people to transmit information on the Internet. However, the widespread use of the email has made it a common tool and carrier for criminals to commit criminal activities. Meanwhile, the forensic investigators are more often to take the email as evidence of criminal cases. Therefore, technical appraisal of email plays an increasingly important role in solving cases and providing evidence in the court.

To verify the authenticity of email evidence, a scientific appraisal technology is of great concern. However, as email is complicated, with different protocols for receiving and sending, various email server software and client, research on email technical appraisal is almost a blank in nowadays. Genwei Liao proposed the basic ideas to identify the authenticity of email from following parts: email header, server log file, email sending environment, email content abnormal and the logic between emails[1]. M.T Banday and Hong Guo et al. studied the working principle of an email, discussed the construction mechanism of keywords that commonly used in the header field, and applied the analysis to email forensic[2, 3]. Based on the email header information, Preeti and Surekha et al. provided an algorithm to identify the data, time, and address spoofing[4, 5]. Email authenticate is challenging due to not only the flexibility of composing, editing, deleting of emails by using offline or online applications, but also the various fields that can be forged by hackers or malicious users. However, the current researches study the emails authentic identification only from the perspective of detecting but not forging.

To have a better performance on emails authentic identification, detectors can consider what method had falsifiers taken to fabricate emails. The game theoretical analysis is useful for analyzing, modeling and deciding for the interdependent and antagonistic relationship. The player can have a prediction of other players' action and strategy in the game theory model. Recently, many literatures on the forensic and anti-forensic with game theoretical framework have been proposed. Mauro Barni et al. used the game model to solve the optimum forensic and counter-forensic strategies in source identification with training data[6].The model is used to derive the Nash equilibrium and the condition under which the false negative error probability tends to zero. Matthew C Stamm et al. developed a theoretical understanding of interactions between a falsifier who uses anti-forensics and a forensic investigator, and decided the optimum decision rule by predicting the falsifier's best anti-forensic strength [7]. Xiangui Kang et al. defined a VIF (Video Inter-frame Forgery) game to analyze the interplay between the forensic investigator and the falsifier, and used the Nash equilibrium strategy to decide under which false alarm rate can the detection rate reach 100%[8]. These studies demonstrate the efficiency of game model in solving the optimal strategy. However, the previous studies are almost based on the multimedia forensic, and the similar research on email forensic is still a blank. And there is no related research had a discussion of how can a forensic investigator predict the forger's action by introducing the game theory into email forensic.

Our work is different from the state-of-the-art studies in several aspects. Firstly, we make a new classification and cost-benefit quantification for the existing email forgery methods and authenticity appraisal technologies. Before we make a prediction with someone's action, we need to have a basic understanding.

The strategies classification and cost-benefit quantification can help detectors know the factors that will affect forger's decision and action. Secondly, we take both email forensic and game theory into consideration simultaneously and propose a DFG game model to analyze the dynamics between email detection and forgery for the first time. The DFG model aims to help the detector and forger find out the trade-offs that depend upon the actions of another. Thirdly, we propose an algorithm to solve the Nash equilibrium of the DFG model. Then the optimal strategy with the maximum benefits and minimum risks can be found for the players.

The rest of paper is organized as follows. In section 2, we study the email header fields and information, and have a discussion on the email forensic and forgery technologies. In section 3, we will give a formalization definition to DFG model, make a strategy classification and cost-benefit quantification for the detector and forger, and introduce an optimal strategy selection algorithm. The experimental work is discussed in section 4. The conclusion and future work are discussed in section 5.

II. E-Mail Detection and Forgery Analysis

Electronic mail, often called email, is a method of exchanging digital messages from an editor to one or more recipients. The email relevant rules are defined by the RFC (Request for Comments), a series of number ranked memorandums issued by the IETF(Internet Engineering Task Force)[9].

A. E-mail Header Analysis

An Internet email messages consist of two major sections: header fields and body. The email header is divided into several fields and each field has a name and value. The email header contains the sender and recipient information, time and data information, email sever information, email transfer information and other relevant information, which plays an essential role to ensure the authenticity of an email. The basic header fields that have been defined in RFCS are show in Table I[3].

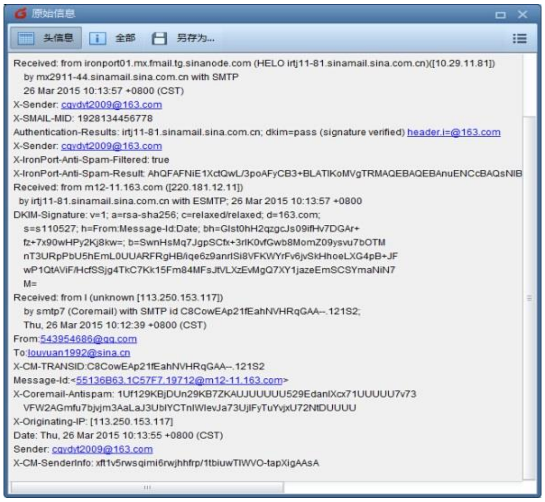
In addition to the basic fields, there are some non-standard, custom fields generated by different mail client, which begin with X-, such as *'X-Sender'*, *'X-Mailer'*, *'X-SMAIL-MID'*, *'X-Received'*, *'X-Originating-IP'* and so on. Expecting for these custom fields, some fields generated because of the security technology used by the mail server, such as *'DKIM-Signature'*, *'Received-SPF'*, *'Sender-ID'*. All these header fields are of great importance for email authenticity appraising.

TABLE I. BASIC FIELDS

Field name	**Description**
From	The email address, and optionally the name of the sender.
To	The email address, and optionally the name of the message recipient.
Data	The local time and date when the message was written.
Subject	A brief summary of the topic of the message.
Message-ID	An automatically generated field, it uniquely identifies this message.
Reply-To	Address that should be used to reply to the message.
Received	Tracking information generated by mail servers that have previously handled a message, in reverse order.
Content-Type	The type of the message content
Content-Transfer-Encoding	The transfer and encoding ways of message content.

B. E-mail Detection and Forgery Analysis

Email spoofing is one of the biggest challenges that threats email security, and the main important forms of email spoofing are data and time spoofing, address spoofing and content spoofing. Generally, an email may be required to be appraised in following three conditions: firstly, the sender or recipient does not recognize they sent or received the email; Secondly, the sender and recipient have objections on the email date and time. Thirdly, the litigants don't reach an agreement on the email content. The Fig.1 shows the header message of an email which has a question on the sender address.



```
Received: from ironport01.mx.fmail.tg.sinanode.com (HELO irtj11-81.sinamail.sina.com.cn)([10.29.11.81])
  by mx2911-44.sinamail.sina.com.cn with SMTP
  26 Mar 2015 10:13:57 +0800 (CST)
X-Sender: cqydyt2009@163.com
X-SMAIL-MID: 1928134456778
Authentication-Results: irtj11-81.sinamail.sina.com.cn; dkim=pass (signature verified) header.i=@163.com
X-Sender: cqydyt2009@163.com
X-IronPort-Anti-Spam-Filtered: true
X-IronPort-Anti-Spam-Result: AhQFAFNiE1XctQwL/3poAFyCB3+BLATIKoMVgTRMAQEBAQEBAnuENCcBAQsNIB
Received: from m12-11.163.com ([220.181.12.11])
 by irtj11-81.sinamail.sina.com.cn with ESMTP; 26 Mar 2015 10:13:57 +0800
DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed; d=163.com;
  s=s110527; h=From:Message-Id:Date; bh=Glst0hH2qzgcJs09ifHv7DGAr+
  fz+7x90wHPy2Kj8kw=; b=SwnHsMq7JgpSCfx+3rlK0vfGwb8MomZ09ysvu7bOTM
  nT3URpPbU5hEmL0UUARFRgHB/iqe6z9anrlSi8VFKWYrFv6jvSkHhoeLXG4pB+JF
  wP1QtAViF/HcfSSjg4TkC7Kk15Fm84MFsJtVLXzEvMgQ7XY1jazeEmSCSYmaNiN7
  M=
Received: from l (unknown [113.250.153.117])
  by smtp7 (Coremail) with SMTP id C8CowEAp21fEahNVHRqGAA--.121S2;
  Thu, 26 Mar 2015 10:12:39 +0800 (CST)
From:543954686@qq.com
To:louyuan1992@sina.cn
X-CM-TRANSID:C8CowEAp21fEahNVHRqGAA--.121S2
Message-Id:<55136B63.1C57F7.19712@m12-11.163.com>
X-Coremail-Antispam: 1Uf129KBjDUn29KB7ZKAUJUUUUU529EdanIXcx71UUUUU7v73
  VFW2AGmfu7bjvjm3AaLaJ3UbIYCTnIWIevJa73UjIFyTuYvjxU72NtDUUUU
X-Originating-IP: [113.250.153.117]
Date: Thu, 26 Mar 2015 10:13:55 +0800 (CST)
Sender: cqydyt2009@163.com
X-CM-SenderInfo: xft1v5rwsqimi6rwjhhfrp/1tbiuwTlWVO-tapXigAAsA
```

Fig. 1. An email's header message, there are two different senders 543954686@qq.com and cqydyt2009@163.com on the header message while the sender on the email envelop is 543954686@qq.com

To find out the real sender, we can analyze the email header information. There are many fields include sender information, such as *'X-Sender'*, *'Authentication-Results'*, *'From'*, *'Message-id'* and *'Sender'*. Among these fields, only the '*From'* field refers to 543954686@qq.com, and the *'From'* field is created by the author. Meanwhile, there are four fields referring to cqydyt2009@163.com, especially the '*Message-id'*, it was an automatically generated field and not easy to be changed.

After the multi-fields correlation analysis of sender, we can appraise the email sender 543954686@qq.com is forged and the real sender is cqydyt2009@163.com.

In fact, the email forgery and detection methods are various, and the example above only represents one situation. For example, modifying the system properties is the most convenient way to falsify an email, and we can falsify the email data by modifying the system time. The Simple Mail Transfer Protocol (SMTP) is an email transfer protocol, and we can use Telnet command to tamper the email address by logging in the SMTP server. We can also use the off-the-shelf software and website to forge emails. The most complex method is to steal someone's email password and imitate him to send emails, but it is not easy to know others' password because of the Encryption software and algorithms. Most people fabricate emails with a certain purpose, may be just a joke but the more is for some profits.

To protect people's profits from the email forgery, various detection strategies need to be taken. Viewing the email header information is the simplest method to detect an email. The multi-fields correlation analysis denotes to analyze a series of fields including one message. For example, the '*Received', 'Data', 'Message-ID',* and *'Boundary'* filed are all including the email data and time. And we appraise the email by contrasting the times of these fields.

We can also use the sender related fields to identify the true address. Making use of external resources means we can take use of off-the-shelf software like 'nslookup' to analyze the IP and DNS, or other information like login and server files to identify the email. Multi-emails correlation analysis indicates that we can identify if the emails are authentic by analyzing the logical relationship among emails, comparing the client, writing habits, IP, address and so on.

Since the methods are so various, how can the detector know which detection strategy is the most effective, and how can the forger know which forgery strategy can bring him the maximum benefits and minimum risk?

III. Detection-Forgery Game Model

Game theory is a study of strategic decision making. Specifically, it is "the study of mathematical models of conflict and cooperation between intelligent rational decision-makers". It attempts to determine mathematically and logically the actions that "players" should take to secure the best outcomes for themselves in a wide array of "games"[10].

A. Detection-Forgery Game Model Definition

A game theoretical model includes three basic elements: Player, Strategy set and Payoff function. The strategic form of a detection-forgery game is a 3-tuple *DFG=(N,S,U)*[11]:

- $N = (N_1, N_2 \cdots N_n)$ is a set of players. Players are the decision-makers who decide the action and strategy to maximize their own interests. And in this game model, the players are detector $\mathrm{N_d}$ and forger $\mathrm{N_f}$.
- $S = (S_1, S_2 \cdots S_n)$ is a set of players' strategies. $\forall i \in n, S_i \neq \emptyset, S_i = (S_1^i, S_2^i, \dots, S_m^i)$ is the strategy set of plyer i. And here we define the strategy sets as $S_d = (S_1^d, S_2^d, \dots, S_n^d)$ and $S_f = (S_1^f, S_2^f, \dots, S_m^f)$.
- $U = (U_1, U_2 \cdots U_n)$ is the payoff function set of the players. It reflects the gain and utility the players can gain from the game. We define the detector's payoff as $\mathrm{U_d}$, and the forger's payoff as $\mathrm{U_f}$.

Definition1: Nash Equilibrium (NE) is a solution concept of a non-cooperative game, it means each player gains the maximum benefits. In $DFG = (N_d, N_f), (S_d, S_f), (U_d, U_f))$, the strategy group $(\mathrm{S_*^d}, \mathrm{S_*^f})$ is a Nash equilibrium if and only if for $\forall S^d \in S_d$, $U_d(S_*^d, S_*^f) > U_d(S^d, S_*^f)$, and for $\forall S^f \in S_f$, $U_f(S_*^d, S_*^f) > U_f(S_*^d, S^f)$.

In a complete information game model, we can use the definition 1 to solve all the possible Nash equilibrium. In the DFG model, $\forall s_i^d \in S_d, \forall s_i^f \in S_f, U_d(s_i^d, s_j^f), U_f(s_i^d, s_j^f)$ represents the detector and forger's payoff while the detector selects the strategy i to detect the email which is forged by strategy j. The Fig.2 shows the corresponding strategy game, where each row represents the detector's strategy and each column represents the forgers' strategy, and the values in the matrix are the payoffs associate to the players.

$$U = \begin{array}{c} \\ S_1^d \\ S_2^d \\ S_3^d \\ S_4^d \end{array} \begin{array}{c} \begin{array}{cccc} S_1^f & S_2^f & S_3^f & S_4^f \end{array} \\ \begin{pmatrix} U_{d11}, U_{f11} & U_{d12}, U_{f12} & U_{d13}, U_{f13} & U_{d14}, U_{f14} \\ U_{d21}, U_{f21} & U_{d22}, U_{f22} & U_{d23}, U_{f23} & U_{d24}, U_{f24} \\ U_{d31}, U_{f31} & U_{d32}, U_{f32} & U_{d33}, U_{f33} & U_{d34}, U_{f34} \\ U_{d41}, U_{f41} & U_{d42}, U_{f42} & U_{d43}, U_{f43} & U_{d44}, U_{f44} \end{pmatrix} \end{array}$$

Fig. 2. The DFG payoff matrix

B. The Classification of Strategies

In the DFG model, the players' strategy set is a necessary component. In this paper, we mainly discuss the forgers' and detectors' strategies based on the email header, and classify the strategies according to Operation Complexity (OC). Richard E Overill[12] used the Operation Complexity(OC) to enable the complexity of both the cognitive and the computational components of a process, and the more complex a process is, the less likely it is to occur accidentally, unintentionally or spontaneously. Similarly, we use the operation complexity to measure the complexity or difficulty of a detection or forgery strategy. Generally, the more complex a strategy is, the higher costs it takes. This can be evaluated according to the amount of extra resources or the steps the players take. For any detection or forgery strategy i, the operational complexity of that strategy can be given by:

$$OC_i = KLM_i + R_i \tag{1}$$

Where OC_i comprises a cognitive complexity component KLM_i and an extra resource component R_i. The KLM is specified by the GOMS-KLM model for measuring the human involvement in the operational process[13] and the R represents the size of files for sending an email. The basic unit of the GOMS-KLM characterization of cognitive information processing is taken to be the mouse button press or release;

similarly, the basic unit of information processing used in characterizing the resource is the byte. In addition, the cognitive component should be scaled by the ratio of the processing rates of the human and computer, typically$\approx 10^6$. Table II shows the KLM operators and normal values[12] and Table III shows an example of the frequent KLM actions and values of modifying the system time to send a false email, and the total value is 62.6. Since this strategy needs no extra resource expect an email client or login an email website, such as Foxmail7.2, then the R is 15,624,827, so the OC=78,224,827.

TABLE II. KLM OPERATORS AND NORMAL VALUES

KLM operators	normal values(sec)
K(key press & release)	0.2
P (point mouse)	1.1
B (button press/ release)	0.1
H (hand to/from keyboard)	0.4
M (mental preparation)	1.2

TABLE III. THE FREQUENT KLM ACTIONS AND VALUES

Action	M	P	B	K	H	Total
1.point with mouse to the target	1	1	2	0	0	2.5
2. single click	4	4	8	0	0	10
3. Modify the time	1	1	2	12	2	5.7
4. open the foxmail	1	1	4	0	0	2.7
5. log in(username and pw)	4	2	4	28	4	14.6
6. write the email (the email content is the least)	6	3	6	66	6	26.7

In order to have a better strategy classification, we can divide the operation complexity into three relative levels:

- L1: The cost is very small and the OC<10^9. For example, modifying the system time needs only an email client, and the operation needs simple steps;
- L2: The operation needs some time and resource and the OC<10^{10}. For example, using telnet command to falsify the email address needs little resources, and the operational step is complex and time-consuming;
- L3: The operation needs much more resource and the OC<10^{11}. For example, stealing other's password not only needs many resources but the operational step is more complex and time-consuming.

Based on the email detection and forgery analysis and the definition of three level operation complexity above, the email detection- and forgery- strategy taxonomy according to the operation complexity are show in Table IV and Table V.

TABLE IV. DETECTION STRATEGY TAXONOMY

Category	Example	OC
No detect	*	0
View the email header	Received, X-Sender	L1
Multi-fields correlation analysis	Time analysis:Received, Date, Message-ID, Boundary	L2
Make use of external resources	nslookup, login files, server information	L3
Multi-emails correlation analysis	Logic, client, writing habit, IP, DNS...	L3

TABLE V. FORGERY STRATEGY TAXONOMY

Category	Example	OC
No forge	*	0
Modify the system properties	Modify the system time	L1
Make use of transfer protocol	Use telnet to falsify the address	L2
Make use of external resources	Email forgery software or website	L2
Steal password	Implant Trojan to the target computer	L3

C. Cost-Benefit Quantification

In order to make the payoff function more exactly and actually, we need to quantify the costs, risks and benefits. The relevant cost factors are defined as follows[14]:

Definition2: Detect Cost (DC) characterizes the amount of resources of implanting a detect action, such as hardware and software resource, expertise, time and so on.

Definition3:Detect Damage (DD) characterizes the amount of damage or risks inflicted by the detector if he can't identify the forged email or treat the forged email as the authentic one. (Expressed in negative values)

Definition4:Detect Benefit (DB) characterizes the amount of benefits inflicted by the detector or the extra-reward if he successfully detected the forged email.

Definition5:Forge Cost (FC) characterizes the amount of resource of implanting a fake action, such as hardware and software resource, expertise, time and so on.

Definition6:Forge Damage (FD) characterizes the amount of damage or the legal penalties to the forger which is inflicted by the detector if he can identify the forged email successfully (Expressed in negative values).

Definition7:Forge Benefit (FB) characterizes the amount of benefits if the forger escaping from the forgery detection.

Definition8:Detection Rate P indicated the possibility that a detect method can successfully identify the forged email as forgery. If there is an email E that has been manipulated using an editing operation m(*), then we assume the null hypothesis H_1 is that E is unaltered and authentic; The alternated hypothesis H_2 is that E is a manipulated version of another email E_1 and E is forged, i.e.

$$\begin{aligned} H_1 &: E \neq m(E_1) \\ H_2 &: E = m(E_1) \end{aligned} \qquad □□□(2)$$

Then we do experiments on a large number of emails that include forged and authentic, and the detection rate P defined as follows:

$$P = P\left(\left(H = H_2\right) \big/ E = m\left(E_1\right) \right) \qquad □(3)$$

Where $\left(H = H_2\right)$ means we appraise the email is forged, and $E = m\left(E_1\right)$ means the email is actually forged. Based on the definitions above, we can define the detailed rewards of detector and forger as follows:

$$U_d = DB*P + DD*(1-P) - DC \tag{4}$$

$$U_f = FB*(1-P) + FD*P - FC \tag{5}$$

From the (4) and (5), we can find that if the detector and forger want to maximize their rewards, the detector needs to maximize the detection rate P while the forger needs to minimize it. However, the detection rate depends not only on the cost of detecting but on the cost of forging. The more detection cost, the higher P, and the more forgery cost, the lower P. So we need to find out an optimal P to maximize both the forger's and detector's rewards.

D. Optimal Strategy Selection

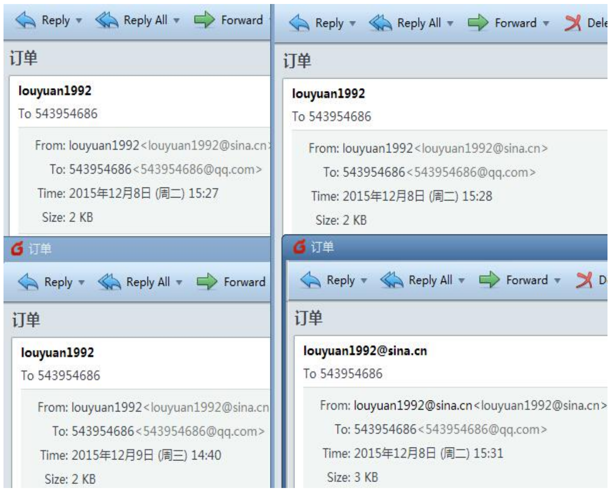
The detection strategy with highest possibility and maximum rewards to appraise emails, and the forgery strategy with maximum possible and rewards to falsify emails can be selected from the candidate sets through *DFG* model. And the detail of the optimal strategy selection process is presented as follows:

Input: detect and forge strategy set
Output: optimal strategy
Algorithm:

```
1) Construct the detector's strategy set $S_d = (S_1^d, S_2^d, \dots, S_n^d)$;
2) Construct the forger's strategy set $S_f = (S_1^f, S_2^f, \dots, S_m^f)$;
3) Initialize the DFG=(($N_d, N_f$),($S_d, S_f$),($U_d, U_f$));
4) For all $s_i^d \in S_d, s_j^f \in S_f$, compute the defection rate P
   according to the (2) and (3),and compute the rewards
   of detector and forger according to the (4) and (5),
   and get the payoff matrix $U_d, U_f$;
5) Set the utility matrix $U = U_d - U_f$;
6) Compute the Nash Equilibrium of DFG. Processes as follows:
   a) Test for the saddle point in the utility matrix;
   b) If there is a saddle point, the saddle point is the Nash Equilibrium;
   c) If there is no saddle point, solve it by linear program. Processes as follows:
      i). maximize Z ,minimize W;
      ii). subject to
      iii). for all $s_i^d \in S_d, s_j^f \in S_f$;
      iv). $\sum_{i=1}^{m} p_i^d U_d(S_i^d, S_i^f) \geq Z$;
      v). $\sum_{j=1}^{n} p_j^f U_f(S_i^d, S_i^f) \leq W$;
      vi). $\sum_{i=1}^{m} P_i^d = 1$, $p_i^d \geq 0$;
      vii). $\sum_{j=1}^{m} P_j^f = 1, p_j^f \geq 0$;
      viii). Get the Nash equilibrium$(p_d^*, p_f^*)$;
7) Decide the optimal strategy.
```

IV. Numerical Analysis

In order to verify the effectiveness and feasibility of the DFG model, we need to introduce the model into the actual email authenticity identification cases. Since the email address spoofing case is the most widely happened in civil disputes, so we consider the counterwork between the detector and forger as follows: A and B are business partners, and A used to order goods from B through emails. However, one day B received an email and the sender on the envelope is A, but A denied to have sent the email and received the goods. We consider the envelop of all the emails shown as Fig.3.



订单

louyuan1992
To 543954686
From: louyuan1992<louyuan1992@sina.cn
To: 543954686<543954686@qq.com>
Time: 2015年12月8日 (周二) 15:27
Size: 2 KB

订单

louyuan1992
To 543954686
From: louyuan1992<louyuan1992@sina.cn>
To: 543954686<543954686@qq.com>
Time: 2015年12月8日 (周二) 15:28
Size: 2 KB

订单

louyuan1992
To 543954686
From: louyuan1992<louyuan1992@sina.cn
To: 543954686<543954686@qq.com>
Time: 2015年12月9日 (周三) 14:40
Size: 2 KB

订单

louyuan1992@sina.cn
To 543954686
From: louyuan1992@sina.cn<louyuan1992@sina.cn>
To: 543954686<543954686@qq.com>
Time: 2015年12月8日 (周二) 15:31
Size: 3 KB

Fig. 3. The envelop of emails,the emails above have the same sender and recipient shown in the envelop,but not all the emails are authentic, some of them are forged,the sender is not the real one

In this case, we considered there are different strategies for a forger to fabricate an email with false sender, and the methods to identify the email's real sender are also various for the detector. In real society, the mostly related money of E-mail appraisal cases is almost from ten thousand to ten million, and the influence of the appraisal results is from 0% to 100%. In this case, the influence of appraisal result is 100%, and the basic unit of the cost and benefit is one thousand dollar. Since the more complex, the higher cost, here we set the cost of L1 from 0-10, the cost of L2 from 10-40, the cost of L3 from 40-100. Table VI,VII shows the strategies, benefits, costs of the email forger and detector.

TABLE VI. Summary of Forgery Strategy

Forger	strategy	FB	FD	FC
S_1^f	Use telnet to login in SMTP for address tampered	100	-100	15
S_2^f	Send a forgery email by an email fake website	100	-100	15
S_3^f	Use off-the-shelf software to send a forgery email	100	-100	20
S_4^f	Build a local email server to send a forgery email	100	-100	30

TABLE VII. Summary of Detection Strategy

Detector	strategy	DB	DD	DC
S_1^d	No detect	0	-100	0
S_2^d	View the email header, X-sender	100	-100	5
S_3^d	Multi-fields correlation analysis, such as From, X-Sender, Received, Message-ID	100	-100	20
S_4^d	Multi-emails correlation analysis	100	-100	50

According to the tables above, we fabricated numbers of email with the forgery strategy set and appraised emails with the given detection strategies. Then we use the optimal strategy selection algorithm to solve the optimal solution. According to the (2),(3),(4),(5), we can get the detection rate P matrices and detection-forgery payoff matrices $\mathrm{U_d}$, $\mathrm{U_f}$ as follows:

$$P = \begin{array}{c} \\ S_1^d \\ S_2^d \\ S_3^d \\ S_4^d \end{array} \begin{array}{cccc} S_1^f & S_2^f & S_3^f & S_4^f \\ \left(\begin{matrix} 0 & 0 & 0 & 0 \\ 0.9 & 0.15 & 0.1 & 0.2 \\ 0.95 & 0.2 & 0.3 & 0.25 \\ 0.95 & 0.85 & 0.8 & 0.78 \end{matrix} \right) \end{array}$$

$$U_f = \begin{array}{c|cccc} & S_1^f & S_2^f & S_3^f & S_4^f \\ \hline S_1^d & 85 & 85 & 80 & 70 \\ S_2^d & -95 & 55 & 60 & 30 \\ S_3^d & -105 & 45 & 20 & 20 \\ S_4^d & -105 & -85 & -80 & -86 \end{array}$$

$$U_d = \begin{array}{c|cccc} & S_1^f & S_2^f & S_3^f & S_4^f \\ \hline S_1^d & -100 & -100 & -100 & -100 \\ S_2^d & 75 & -75 & -85 & -65 \\ S_3^d & 70 & -80 & -60 & -70 \\ S_4^d & 40 & 20 & 10 & 6 \end{array}$$

$$U = \begin{array}{c|cccc} & S_1^f & S_2^f & S_3^f & S_4^f \\ \hline S_1^d & -185 & -185 & -180 & -170 \\ S_2^d & 170 & -130 & -145 & -95 \\ S_3^d & -175 & -125 & -80 & -90 \\ S_4^d & 145 & 105 & 90 & 92 \end{array}$$

An equilibrium $P_d = (p_1^d\ p_2^d, p_3^d, p_4^d)$=(0,0.0083,0,0.9917), Z=90.4545; $P_f = (p_1^f, p_2^f, p_3^f, p_4^f\ )$=(0,0,1,0), M=90 can be found by the optimal strategy selection algorithm. Therefore, the detector plays the strategy$\mathrm{S_2^d}$ with the possibility 0.0083 and the strategy $\mathrm{S_4^d}$ with the possibility 0.9917, and the strategy $\mathrm{S_3^f}$ is the optimal strategy for the forger. From the above results, we can find that the strategy set (S_4^d, S_3^f) is an optimal strategy for the example case. This result indicates that the forger is most likely to use off-the-shelf software to fabricate an email with false sender, and the forensic investigator will have a maximum reward by making correlation analysis with multi-emails when facing such cases.

V. Conclusion

In this paper, we have proposed a DFG game model for analyzing optimal detect and forge strategy decision in email authenticity identification. We regard the interactions between a forensic investigator and a forger as a two-player, non-cooperative, nonzero-sum game and formulated the DFG game model. Based on the strategies' cost-benefit quantification and DFG model, we selected the optimal strategy from the given sets. And finally, we used a practical case study to verify the effectiveness of the DFG model. Nevertheless, there are still some problems of the DFG model, such as the cost-benefits quantification and payoff functions we adopted in this paper is not very comprehensive. We will pay more attention to improving it in the future work.

Acknowledgment

This work was supported partially by the National Social Science Committee of China under Grant No. 14BFX156 and partially by Chongqing science and technology research project under Grant No. KJ1400428

References

[1] Genwei Liao, The Email Authenticity Identification, Criminology Review, (2009) 42-48.

[2] M.T. Banday, Technology Corner: Analysing E-mail Headers For Forensic Investigation, Journal of Digital Forensics, Security and Law, 6 (2011) 49-64.

[3] H. Guo, B. Jin, W. Qian, Analysis of Email Header for Forensics Purpose, Communication Systems and Network Technologies (CSNT), 2013 International Conference on, IEEE2013, pp. 340-344.

[4] P. Mishra, E.S. Pilli, R. Joshi, Forensic Analysis of E-mail Date and Time Spoofing, Computer and Communication Technology (ICCCT), 2012 Third International Conference on, IEEE2012, pp. 309-314.

[5] S. Gupta, E.S. Pilli, P. Mishra, S. Pundir, R. Joshi, Forensic analysis of E-mail address spoofing, Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference-, IEEE2014, pp. 898-904.

[6] M. Barni, B. Tondi, Optimum forensic and counter-forensic strategies for source identification with training data, Information Forensics and Security (WIFS), 2012 IEEE International Workshop on2012, pp. 199-204.

[7] M.C. Stamm, W.S. Lin, K.J.R. Liu, Temporal Forensics and Anti-Forensics for Motion Compensated Video, Information Forensics and Security, IEEE Transactions on, 7 (2012) 1315-1329.

[8] X. Kang, J. Liu, H. Liu, Z.J. Wang, Forensics and counter anti-forensics of video inter-frame forgery, Multimedia Tools and Applications, (2015) 1-21.

[9] P. Loshin, Essential email standards: RFCs and protocols made practical, John Wiley \\&amp; Sons, Inc.2000

[10] M.J. Osborne, An introduction to game theory, Oxford University Press New York2004.

[11] R.E. Overill, J.A. Silomon, K.-P. Chow, A complexity based model for quantifying forensic evidential probabilities, Availability, Reliability, and Security, 2010. ARES'10 International Conference on, IEEE2010, pp. 671-676.

[12] D. Kieras, Using the keystroke-level model to estimate execution times, University of Michigan, 555 (2001).

[13] W. Jiang, B.X. Fang, H.L. Zhang, Z.H. Tian, X.F. Song, Optimal Network Security Strengthening Using Attack-Defense Game Model, Information Technology: New Generations, 2009. ITNG &amp;#039;09. Sixth International Conference on2009, pp. 475 - 480.