Volume 2, Issue 2, February 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Amalgamation of Randomized Fuzziness and Uncertainty in Deception Detection using Coadunation Techniques

S.RAJKUMAR
Research Scholar/Bharathiar University,
AP&HOD/ CSE, NIET,
Coimbatore,India

V.NARAYANI
Research Scholar,
Dept. of Computer Science
St.Xavier's College, Tirunelveli

DR.S.P.VICTOR
Associate.Prof. &HOD
Dept. of Computer Science
St.Xavier's College, Tirunelveli

Abstract:In the modern era of Computer communication, safe surfing is a gifted process to ensure the protection of our belongings. Phising is the recent technique in the fraudulent mode of communication strategies. Only expertise knowledge people are the exceptional to secure themselves in these intelligent attacks .Manual or casual handling of deception detection in the case of Phising is a tedious process to provide maximum security. In this research paper we provide the techniques of Randomized approach, Fuzzified approach and Uncertainty approach with its individual and combined implementation strategies. We perform the efficiency comparison among these three techniques including all such combinations with an experiment. The results are compared and discussed for future developments.
Keywords: Deception, detection, uncertainty, Randomness, Fuzziness

I.INTRODUCTION:

Detection of Deception is useful for managers, employers, and for anyone to use in everyday situations where telling the truth from a lie can help prevent you from being a victim of fraud/scams and other deceptions. This is just a basic run down of physical gestures and verbal cues that may indicate someone is being untruthful [1].

A. Fuzzification

Fuzzy sets have movable boundaries, i.e., the elements of such sets not only represent true or false values but also represent the degree of truth or degree of falseness for each input.

Fuzzy logic is the part of artificial intelligence or machine learning which interprets a human's actions. Computers can interpret only true or false values but a human being can reason the degree of truth or degree of falseness. Fuzzy models interpret the human actions and are also called intelligent systems. Fuzzy logic has mostly been applied to control systems. Fuzzy control systems interpret the expert human and replace them for performing certain tasks such as control of a power plant [3]. Fuzzy controllers apply decision rules (if-then rules) by making use of critical variables to interpolate the output between the crisp boundaries. Some typical examples where fuzzy logic has been implemented are

1. Railway (Sendai Railways in Japan)
2. Automobile industries (transmission and braking)
3. Heating and cooling systems
4. Copy machines
5. Washing machines

Fuzzification is the process of changing a real scalar value into a fuzzy value. This is achieved with the differenttypes of fuzzifiers. Fuzzification of a real-valued variable is done with intuition, experience and analysis of the set of rules and conditions associated with the input data variables.[4]

B. Randomness

The Oxford English Dictionary defines 'random' as "Having no definite aim or purpose; not sent or guided in a particular direction; made, done, occurring, etc., without method or conscious choice; haphazard." This concept of randomness suggests a non-order or non-coherence in a sequence of symbols or steps, such that there is no intelligible pattern or combination.

C. Pseudorandom

Pseudorandom variable is a variable which is created by a deterministic procedure (often a computer program or subroutine) which (generally) takes random bits as input. The pseudorandom string will typically be longer than the original random string, but less random (lessen tropic, in the information theory sense). This can be useful for randomized algorithms [5].

D. Randomized Algorithm

A randomized algorithm is an algorithm which employs a degree of randomness as part of its logic. The algorithm typically uses uniformly random bits as an auxiliary input to guide its behavior, in the hope of achieving good performance in the "average case" over all possible choices of random bits. Formally, the algorithm's performance will be a random variable determined by the random bits; thus either the running time, or the output (or both) are random variables. [6].

II. PROPOSED RESEARCH MODEL

The following diagram illustrates the implementation of Randomized approach, Fuzzylogic and Uncertainty in the deceptive datum with three stages. Stage 1 accomplishes

the implementation of individual approaches of Randomized, Fuzzy and Uncertainty, then Stage 2 accomplishes the implementation of two approaches at a time, finally Stage 3 comprises the combined implementation of Randomized, Fuzzylogic and Uncertainty towards the deception datum. For comparison and performance analysis each individual Stage is independent (output of one stage will not be sent as input for other stage).

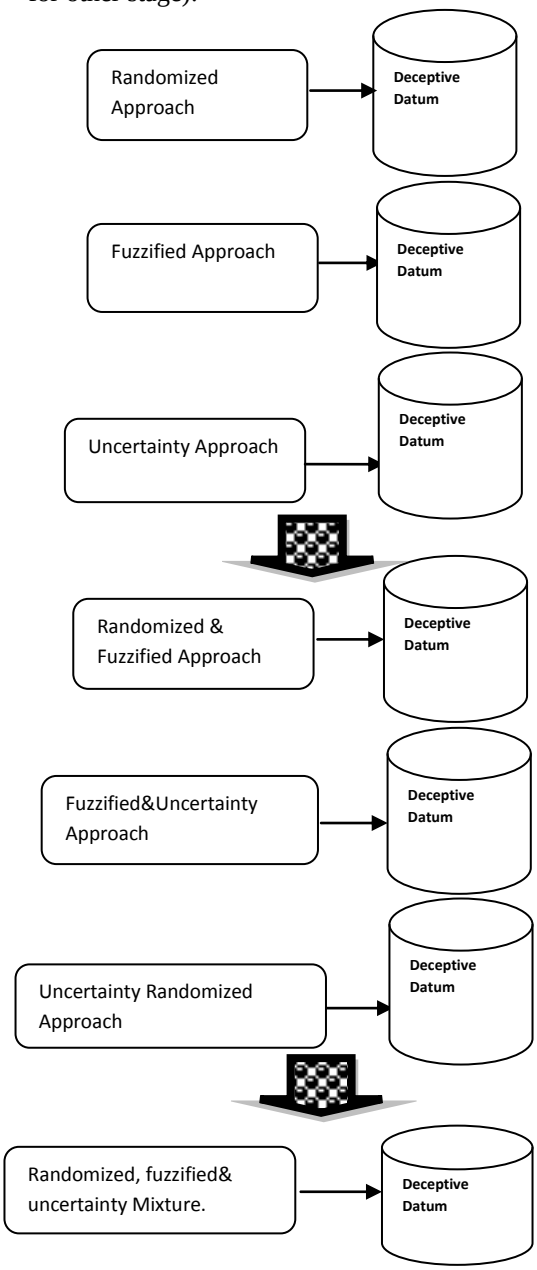


Fig 1:Proposed Model

III. RESEARCH METHODOLOGY

Consider the Deception Detection Factor as Z, which possibly revolves around the following conceptual schema. The value of Z attains any one of the following strategies, on the individual or combined applications.

r(X)-Represents Randomized calculation for Datum-X.
f(X)-Represents Fuzzified calculation for Datum-X.
u(X)-Represents Uncertainty calculation for Datum-X.
R-Randomized evaluation of Datum-X.
F-Fuzzified evaluation of Datum-X.
U-Uncertainty Evaluation of Datum-X.
RF-Randomized&Fuzzified evaluation of Datum-X.
FU-Fuzzified&Uncertainty evaluation of Datum-X.
RU-Randomized&Uncertainty evaluation of Datum-X.
RFU-Randomized,Fuzzified and Uncertainty evaluation of Datum-X.

$$Z = \begin{cases} R = (1-\alpha)*r(X) \text{ or } I\alpha I * r(X), \text{ if } \alpha > 0 \text{ or } \alpha < 0 \text{ respectively.} \\ F = (1-\beta)*(f(X) \text{ or } I\beta I * f(X), \text{ if } \beta > 0 \text{ or } \beta < 0 \text{ respectively.} \\ U = \gamma * u(X) \\ RF = Max(R, F) + \delta. \\ FU = Max(F, U) + \lambda. \\ RU = Max(R, U) + \epsilon. \\ RFU = Max(R, F, U, RF, FU, RU) + \pi. \end{cases}$$

Where α, β, γ, δ, λ, ϵ, and π is Real numbers and lies between 0 &1.

α = No of Logical Factors-No of Non-Logical Factors / Total No of Factors.

β= No of Relational Categories-No of Non-Relational Categories / Total No of Categories.

γ=No of suspected Occurrences/Total No of Occurrences.

$\delta = (I\alpha I + I\beta I)/10.$

$\lambda = (I\beta I + I\gamma I)/10$

$\epsilon = (I\gamma I + I\alpha I)/10$

$\pi = (I\alpha I + I\beta I + \gamma + \delta + \lambda + \epsilon)/10.$

A. Randomized Approach r(X) Computation

Firewall/Antivirus/Internet Security warning- r1(X) =0.1
No secure protocol (http instead of https) - r2(X) =0.2
Hidden Address Bar- r3(X) =0.3
No padlock found- r4(X) =0.4
Invalid Headers (for Mail) /Sender IP (for open port) - r5(X) =0.5
Ordering component variance – r6(X) =0.6
Different or unusual appearance- r7(X) =0.7
(Including design/color/display)
Non supporting Site rating by browsers- r8(X) =0.8
Component Missing- r9(X) =0.9
Processed with wrong credentials- r10(X) =1.0
(Usernames/Passwords with purposive faultiness)

$$R(X) = \sum_{I=1}^{10} ri(X) / 10$$

B. Fuzzified Approach f(X) Computation
(Fuzzy Categorization for Phising)

1. Dual Phishing (fi(X) =0.9)
Acquire information such as usernames, passwords, and credit card details by masquerading as a trustworthy entity in an electronic communication.
2. Spear Phishing (fi(X) =0.8)
Targeted group members only based on attraction.
3. Clone Phishing (fi(X) =0.7)
A type of Phishing attack whereby a legitimate, and previously delivered, email containing an attachment or link has had its content and recipient address(es) taken and used to create an almost identical or cloned email. The attachment or Link within the email is replaced with a malicious version and then sent from an email address spoofed to appear to come from the original sender. It may claim to be a re-send of the original or an updated version to the original.
This technique could be used to pivot (indirectly) from a previously infected machine and gain a foothold on another machine, by exploiting the social trust associated with the inferred connection due to both parties receiving the original email.
4. Whaling (fi(X) =0.6)
Several recent Phishing attacks have been directed specifically at senior executives and other high profile targets within businesses, and the term whaling has been coined for these kinds of attacks
5. Link manipulation (fi(X) =0.5)
Most methods of Phishing use some form of technical deception designed to make a link in an e-mail (and the spoofed website it leads to) appear to belong to the spoofed organization.
6. Filter evasion (fi(X) =0.4)
Phishers have used images instead of text to make it harder for anti-Phishing filters to detect text commonly used in Phishing e-mails.
7. Website forgery (fi(X) =0.99)
Once a victim visits the Phishing website, the deception is not over. Some Phishing scams use JavaScript commands in order to alter the address bar. This is done either by placing a picture of a legitimate URL over the address bar, or by closing the original address bar and opening a new one with the legitimate URL.
An attacker can even use flaws in a trusted website's own scripts against the victim. These types of attacks (known as cross-site scripting) are particularly problematic.
8. Phone Phishing (fi(X) =0.1)
Not all Phishing attacks require a fake website. Messages that claimed to be from a bank told users to dial a phone number regarding problems with their bank accounts.[43] Once the phone number (owned by the phisher, and provided by a Voice over IP service) was dialed, prompts told users to enter their account numbers and PIN. Vishing (voice Phishing) sometimes uses fake caller-ID data to give the appearance that calls come from a trusted organization. [44]
9. Tab nabbing (fi(X) =0.2)
One of the latest Phishing techniques is tab nabbing. It takes advantage of the multiple tabs that users use and silently redirects a user to the affected site.
10. Evil twins (fi(X) =0.3)
It is a Phishing technique that is hard to detect. A phisher creates a fake wireless network that looks similar to a legitimate public network that may be found in public places such as airports, hotels or coffee shops. Whenever someone logs on to the bogus network, fraudsters try to capture their passwords and/or credit card information.

$$F(X) = \sum_{I=1}^{10} fi(X) / 10$$

C. Uncertainty Evaluation u(X) Computation

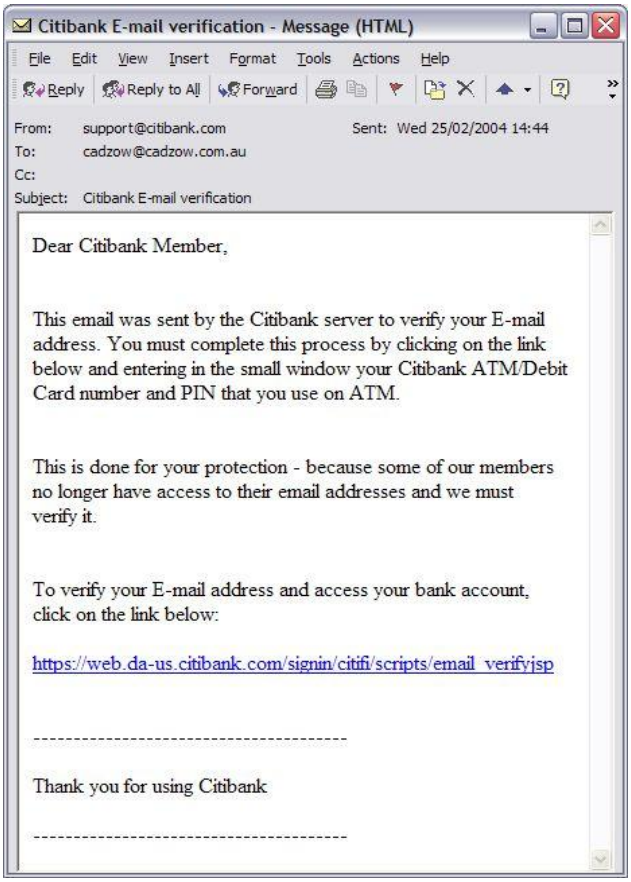
GUM (Guide to the Expression of Uncertainty in Measurement) approach is that it is not possible to state how well the true value of the measurand is known, but only how well it is believed to be known. Measurement uncertainty can therefore be described as a measure of how well one believes one knows the true value of the measurand. This uncertainty reflects the incomplete knowledge of the measurand.
The notion of "belief" is an important one, since it moves metrology into a realm where results of measurement need to be considered and quantified in terms of probabilities that express degrees of belief.

Components	Rating
1. Genuine Source: Belief/Disbelief	0.1
2. Genuine IP Address: Belief/DisBelief	0.2
3. Genuine Protocol: Belief/DisBelief	0.3
4. Genuine Design: Belief/DisBelief	0.4
5. Genuine Descriptions: Belief/DisBelief	0.5
6. Genuine Alert/Warning: Belief/Disbelief (Firewall/Internet Security)	0.6
7. Not Blacklisted: Belief/DisBelief	0.7
8. Genuine webpageRating: Belief/DisBelief	0.8
9. Genuine Headers/Links:Belief/Disbelief	0.9
10. Known userResponse: Belief/DisBelief	0.95

$$u(X) = \sum_{I=1}^{10} ui(X) / 10$$

IV.EXPERIMENT

We perform googling for the selection of websites and reached with a Citibank website and a Google Gmail website. [7]Then we proceed with our proposed model computations for the attainment of results. The computations are as follows,
The following examples are real phish attacks and the web addresses shown were real. Although, at the time of writing, these sites had been shut down, do not attempt to visit these sites. They are shown for illustration only.



The following email pretends to be from Citibank: Upon clicking the link the user is taken to the following authentic-looking page:

Table 1:
Computation table for Bank Webportal

α, =2-8/10= -0.6 -ve Value	β, =3-7/10=-0.4 -ve Value
γ, =8/10=0.8 (80 % Uncertainty)	δ, = α + β /10=0.1
λ, = β + γ /10=0.12	Є, = γ+ α/10=0.14
Π=(α + β+ γ+ δ+ λ +Є)/10 =0.216	r(X)=(0.1 +0.3 +0.4 + 0.7 +0.8 +0.9) /10 =0.34
f(X)=(0.5 +0.7+0.8+ 0.9+0.99)/10 = 0.389	u(X)=(0.1+0.2+0.3+ 0.5+0.6+0.7+0.8+0.9 +0.95+0.99) /10=0.604
R-0.204,	F-0.1556,
U-.48,	RF-.304 ,
FU-.50 ,	RU-0.52,
RFU-0.736	

Therefore 73.6% deceptive webpage.Morethan 50% represents the deceptive webpage so we kindly avoid browsing and confirm it with the corresponding authority. Then we verify the second web portal for Gmail as follows [9],

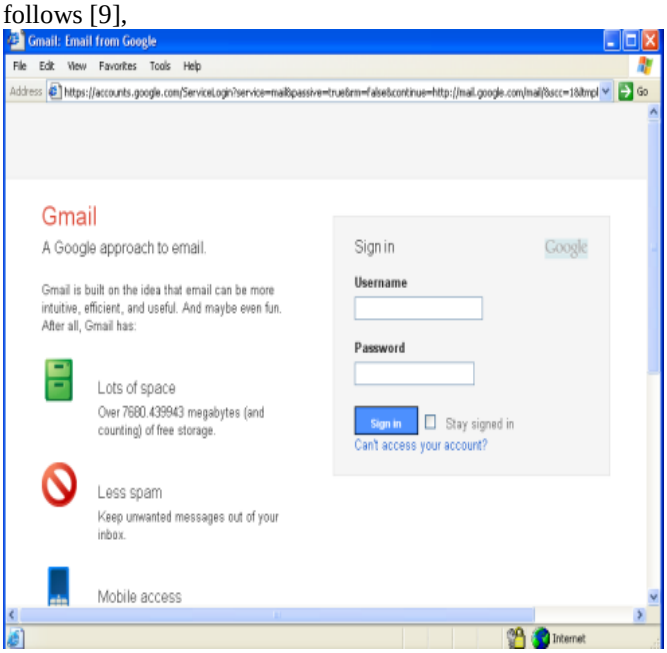


Table 2:
Computation table for GMail Webportal

α, =10-1/10= 0.9 +ve Value take (1- α) value=0.1	β, =10-1/10= 0.9 +ve Value take (1- β) value=0.1
γ, =1/10=0.1 (10 % Uncertainty)	δ, =0.18
λ, =0.1	Є, =0.1
Π=0.68/10=.068	r(X)=(0.1) /10=0.01
f(X)=(0.9)/10= 0.09	u(X)=0.1/10=0.01
R=.001,	F=0.009,
U=0.001,	RF-0.029,
FU-0.029,	RU-0.021,
RFU-0.097	

V.RESULTS AND DISCUSSION

While comparing the analysis results for the Citibank and Gmail web portals, we are in the situation of avoiding dangerous portal address for a genuine banking system (Citibank).Confirming the web portal address and following the Bank instructions definitely keep ourselves in safer surfing. The results are as follows,

For the Website: http://kb.cadzow.com.au:15384 /cadzow/details.aspx?ID=1422 [9]

Table 3:
Results table for Bank Webportal

Component Implementation	DeceptionDetection Level
Randomization	20.4 %
Fuzzification	15.56 %
Uncertainty	48 %
Randomization & Fuzzification	30.4 %
Fuzzification & Uncertainty	50 %
Randomization & Uncertainty	52 %
Randomization, Fuzzification & Uncertainty	73.6 %

For the Website: https://www.gmail.com

Table 4:
Results table for GMail Webportal

Component Implementation	Deception Detection Level
Randomization	0.1 %
Fuzzification	0.9 %
Uncertainty	0.1 %
Randomization & Fuzzification	2.9 %
Fuzzification & Uncertainty	2.9 %
Randomization & Uncertainty	2.1 %
Randomization, Fuzzification & Uncertainty	9.7 %

The proposed model identifies the deceptive web portal and genuine web portal with a variant level as 74% and 10 % deficiency approximately. When the final result of RFU-Randomized, FuzzyUncertainty Evaluation crosses 50%, it is the responsibility of the browser or user to skip the process immediately. Our proposed model produces only the maximum efficiency rate as 99 % with the concrete result as proceed or not to proceed further.

VI.CONCLUSION

Normal Detecting deception for the webpage portals are now being a tedious process due to the implementation of advanced techniques. But when we implement the tools of predictability from the unpredictable strategies such as Mathematicalrandomization, Fuzzylogic, Uncertainty, Genetic algorithm etc, it is possible to detect the deception level with some level of efficiency.

The individual application of predictable tools provide less efficiency than with the combined application In this research we identified that the individual application provides 30 % efficiency then the Combination of two applications provides 60 % efficiency finally the fusional application of three strategies provides 90% efficiency.

In near future we will try to implement Deception detection techniques with the combined approach of Mathematical Randomization, Fuzzylogic, Uncertainty, Genetic algorithm and artificial intelligence to attain 100 % efficiency.

VII.REFERENCES

[1] Steve Woznaik, Kevin D.Mitnick, William L.Simon, 2002. "The art of Deception: controlling the human element of security". Wiley; 1 Edition.
[2] Zuckerman, M., DePaulo, B.M. And Rosenthal, R."Verbal and Nonverbal Communication of Deception". In L.Berkowitz(Ed) (1981)
[3] Burgeon, J.K., and Qin, T. "The Dynamic Nature of Deceptive Verbal Communication". Journal of Language and Social Psychology, 2006, vol25 (1), 1-22.
[4] Bond,c.,F. "A world of lies: the global deception research team", Journal of Cross-culture Psychology, 2006, Vol.37 (1), 60-74.
[5] Pennebaker, J.W, Mehl, M.R. &Niederhoffer, K."Psychological Aspects of natural language use: our words, ourselves". Annual Review of Psychology, 2003, 54,547-577
[6] Whissell,C., Fournier,M.,Pelland,R., Weir, D.,& Makaree,K. "A Comparison of Classification methods for predicting deception in Computer - mediated communication". Journal of Management Information systems, 2004, 20(4), 139-165.
[7] http://www.google.com
[8]http://kb.cadzow.com.au:15384/cadzow/details.aspx?ID=1422
[9] www.gmail.com

AUTHORS PROFILE

Mr.S.Rajkumar completed his M.E–Computer Science &Engineering at Sathyabama University, Chennai and currently doing his Ph.D in the area of Computational Science. He is a Research Scholar of Bharathiar

University and working as a HOD/CSE at NIET Coimbatore.

Ms.V.Narayani completed her M.C.A in M.S University, Tirunelveli and M.Phil in Mother Teresa University, Kodaikanal. She submitted her Ph.D thesis in the area of Data Mining.

Dr. S. P. Victor earned his M.C.A. Degree from Bharathidasan University, Tiruchirappalli.The M. S. University, Tirunelveli, awarded him Ph.D degree in Computer Science for his research in Parallel Algorithms. He is the Head of the Department of Computer Science, and the Director of the Computer Science Research centre, St.Xavier's college (Autonomous), Palayamkottai,Tirunelveli.TheM.SUniversity, Tirunelveli and Bharathiar University, Coimbatore has recognized him as a research guide. He has published research papers in international, national journals and conference proceedings. He has organized Conferences and Seminars at national and state level.