Terminator for E-mail Spam - A Fuzzy Approach Revealed

P.SUDHAKAR[1], G.POONKUZHALI[2] , K.THIAGARAJAN [3] , K.SARUKESI[4]

[1]Vernalis systems Pvt Ltd, Chennai- 600116
[2] Department of Computer Science and Engineering, Rajalakshmi Engineering College,
Affiliated to Anna University- Chennai, Tamil Nadu
[3] Department of Science and Humanities, KCG College of Technology
Affiliated to Anna University-Chennai, Tamil Nadu
[4] Hindustan Institute of Technology and Science-Chennai,Tamil Nadu
INDIA
[1] sudhakar.asp@gmail.com, [2]poonkuzhali.s@rajalakshmi.edu.in,
[3] vidhyamannan@yahoo.com , [4]profsaru@gmail.com

Abstract - In this information technology world, the highest degree of communication happens through e-mails. Realistically most of the inboxes are flooded with spam e-mails as most of transactions through this internet is affected by Passive attacks and Active attacks. Several algorithms exist in the e-world to defend against spam e-mails. But the fulfilment of accuracy in deducting spam e-mail is still oscillating between 80-90%. This clearly shows the necessity for improvement in spam control algorithms on various projections. In this proposed work a new solvent was chosen in the fuzzy word to combat against spam e-mails. Various fuzzy rules are created for spam e-mails and every e-mail is enforced to pass through fuzzy rule filter for identifying spam. Results of the each fuzzy rule for the input e-mails are derived to classify the e-mail to be spam or consent.

Key-Words - E-mail, E-mail spam, Fuzzy, Fuzzy Control, Fuzzy logic, Spam, Spam deduction, User Attitude.

I. INTRODUCTION

E-mail spam, known as unsolicited bulk E-mail (UBE), junk mail, or unsolicited commercial e-mail (UCE), is the practice of sending unwanted e-mail messages, frequently with commercial content, in large quantities to an indiscriminate set of recipients. Spam in e-mail started to become a problem when the Internet was opened up to the general public in the mid-1990s. It grew exponentially over the following years, and today composes some 80 to 85% of all the e-mail in the world, by a "conservative estimate". Pressure to make e-mail spam illegal has been successful in some jurisdictions, but less so in others [1]. Spammers take advantage of this fact, and frequently outsource parts of their operations to countries where spamming will not get them into legal trouble. Though, e-mail is undoubtedly a very effective method of communication these days but at times it can be quite vexing when one is confronted with so many unwanted e-mails where the recipients miss their important e-mails just because their mailbox space is often eaten up by these unwanted e-mails.

The legal status of spam varies from one jurisdiction to another. Spammers collect e-mail addresses from chat rooms, websites, customer lists, newsgroups, and viruses which harvest users' address books, and are sold to other spammers. They also use a practice known as "e-mail appending" or "epending" in which they use known information about their target (such as a postal address) to search for the target's e-mail address. Much of spam is sent to invalid e-mail addresses. Spam averages 78% of all e-mail sent. According to the Message Anti-Abuse Working Group, the amount of spam e-mail was between 88-92% of e-mail messages sent in the first half of 2010. Most of the inbox is flooded with these Spams which occupies lot of memory space. There are several algorithms available for detecting and filtering spam e-mails. Among the existing algorithms, Bayesian filtering produces best result, still it does not detect all the spam e-mails. Most of the existing algorithms considers content alone for filtering the spam e-mails. To detect all the spam e-mails, existing spam filtering methods has to be enhanced. In this proposed work, a new algorithm is devised with various fuzzy rules and fuzzy variables. Each fuzzy rule will produce Attack Factor values which are consider for arriving result. Each rule Attack Factor value was arrived by comparing input parameter against Black list and White List. Black list contains predetermined spam content. White list contains acceptable contents. This final result from above calculated Attack Factor will decide the input e-mail content to be spam or ham or to be sent hold state. The final result of the algorithm was obtained by summing up each rule result value and decision was taken based on the result of the individual rules.

i. RELATED WORKS

Xavier Carreras et al.[2] proposed a Boosting algorithm for Anti Spam filtering. Even though Boosting algorithm delivers good result, possibility of misclassification costs persist inside the AdaBoost learning algorithm.

William W. Cohen et al.[3] suggested Speech act theory for e-mail filtering. The outcome of Speech act theory highly depend on the learning and this approach shows new projection for classifying e-mail spam content**.**

Harris Drucker et al.[7] developed support vector Machines for Spam Categorization. Even though support vector approach outperforms well, switching from training model need user intervention. Addition to that, reply e-mails are considered as no spam.

Joes M.Gomez Hidalgo et al.[8] presents a new dimension for spam e-mail classification.

Nikolos et al.[13] implemented new technique for spam categorization couple with header information and content information. However this system is under research in peer to peer networks. Even though the conceptualization is good, but the practical bottle neck will comes for identification of spam words from the global set. This will take large amount of time as it works with centralized architecture.

Peng et al.[9] Proprosed a new system for applying spam filter in distributed environment. The proposed techniques out performs well during implementation of spam filter in the distributed system. But Author fails to state the technique that can be used to identify spam based on content. The technique handled in this approach (copy rank) performs based on e-mail header rather than e-mail header and body content.

Wanli et al.[10] projected a new techniques for identifying spam e-mail of content type image. But the experimental results shows less confidence on their approach due to misclassification. From the misclassification list, image based classification got highest rank over other text, HTML and non English text classifications.

Sadegh et al.[12] follow through a new approach called Bayesian spanning tree with Likelihood function to identify the e-mail in the e-mail space. From the likelihood classification, Bayesian Spanning Tree outperforms well compared to Navie Bayesian approach by considering precision and F-measure as measurement. Nevertheless Bayesian approach produces high result, still there is a large space to reach 100% accuracy. Bayesian precision measure declares at the maximum of 85% efficiency can be obtained by using Bayesian spanning tree.

ii. OUTLINE OF THE DOCUMENT

Section 2 composes various fuzzy rules formation for the input e-mail parameters to identify the e-mail as a spam or consent. Section 3 Implements the fuzzy rules formed over input e-mail(s). Section 4 predicts the input e-mail and categorize into appropriate buckets. Section 5 proposes results and discussion on the results with future work.

II. FUZZY SYSTEM AND FUZZY RULES GENERATION

Fuzzy Logic (FL) is a problem-solving control system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, multi-channel PC or workstation-based data acquisition and control systems. It can be implemented in hardware, software, or a combination of both. FL provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. FL's approach to control problems mimics how a person would make decisions, only much faster.
Fuzzy rules have been advocated as a key tool for expressing pieces of knowledge in “fuzzy logic”

i. FUZZYFICATION

Input variable : {Sender’sAddress, Sender_IP, Subject_Words, ContentWords, Attachment}
Fuzzy set : {positive, Zero, Negative}
Linguistic set : (highpositive, highNegative, Zero}

Rule 1:

a: IF ∃ SenderAddress ∈ spammer list → AttackFactor=-0.25;
b: IF ∃ SenderAddress ∈ to Ham list →AttackFactor=0.25;
c : IF ∃ Sender Address ∉ Spammerlist & ∃ Sender address ∉ Ham addresslist →AttackFactor=0;

Explanation:

Rule 1.a : If there exist a sender address belongs to spammer list, then Attack Factor of this rule should be set to -0.25;
Rule 1.b : If there exist a sender address belongs to Ham list then, Attack Factor of this rule should be set to 0.25;
Rule 1.c : If there exist a sender address that doesn’t belongs to spammer list and Ham list then, Attack Factor of this rule should be set to 0;

Rule 2 :

a: IF ∃ Sender_IP ∈ SpammerIPlist →AttackFactor= -0.25;
b: IF ∃ Sender_IP ∈ HamIPlist →AttackFactor=0.25;
c: IF ∃ Sender_IP ∉ SpammerIPlist & HamIPlist →AttackFactot=0;

Explanation:

Rule 2.a : If there exists a sender IP address belongs to Spammer list, then Attack Factor of this rule should be set to -0.25;
Rule 2.b : if there exists a sender IP address belongs to Ham list, then Attack Factor of this rule was set to 0.25;
Rule 2.c : If there exists a sender IP address doesn't belongs to Spammer list and Ham List then Attack Factor of this rule was set to 0;

Rule 3:

a: IF ∀ Subject words ∈ Spam words →AttackFactor= -0.50;
b: IF ∃ Subjectword ∈Spamwords → -0.50<AttackFactor< 0.50

Explanation:

Rule 3.a: If all Subject words belongs to Spam words then, Attack Factor of this rule should be set to -0.50;
Rule 3.b : If there exists a subject word that belongs to spam word then Attack Factor of this rule is varies from -0.50 to +0.50;

Rule 4:

a: IF ∀ Content words ∈Spamwordlist →AttackFactor= -0.50;
b: IF ∃ Content words ∈Spamwordslist →-0.50<AttackFactor< 0.50;

Explanation:

Rule 4.a : If all e-mail content words belongs to Spam words then, Attack Factor of this rule should be set to -0.50;
Rule 4.b : If there exists an e-mail content word that belongs to spam word then Attack Factor of this rule is varies from -0.50 to +0.50;

Rule 5 :

a: IF ∀ Attachment ∉VirusList →AttackFactor=1.0;
b: IF ∃ Attachment ∈Visuslist →AttackFactor= -1.0;

Explanation:

Rule 5.a : If all attachment doesn't belong to virus list then, Attack Factor of this rule is set to 1.0;
Rule 5.b : If there exist an attachment belongs to virus list, then Attack Factor of this rule is set to -1.0;

III. FUZZY RULE IMPLEMENTATION

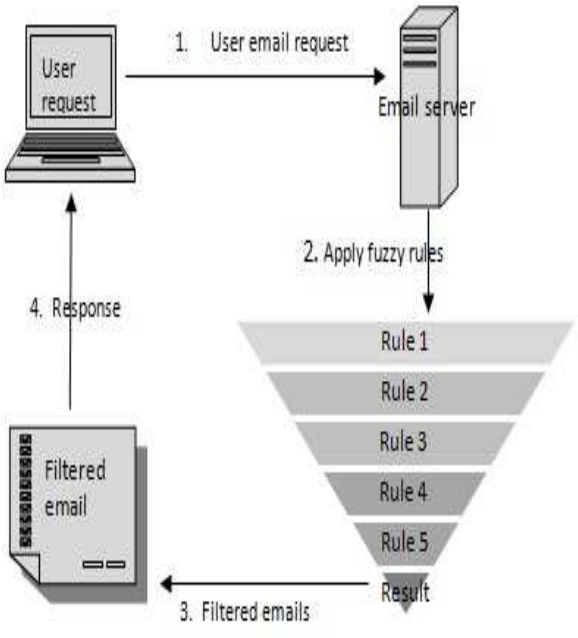

Figure 1. Architecture of proposed system

When an e-mail is arrived, identified fuzzy input parameters are extracted and it is passed to fuzzy system for identification as per Figure1. After Fuzzyfication and Defuzzyfication categorized e-mails are send back to user. Detailed internal follow was shown in Figure 2.

Rule 1 was applied on Fuzzy input parameter- Sender address. Based on Rule 1, Sender address was extracted from e-mail and compared against the Black list which has spammer e-mail address list. If any match was found then, Attack Factor for this rule was set to -0.25. If sender address was not found in the black list, then it was compared against the White list which contains all good and acceptable e-mail addresses. If match was found, then attack factor for this rule was set to 0.25. If sender address was not found in both Black and White list, then attack factor for this rule was set to 0. Set this rule result in R1.

Rule 2 was applied on Fuzzy Input parameter- Sender IP. IP Address of the sender was compared against the IP Address Black List. If match was found, then Rule 2 Attack Factor was set to -0.25. If not found, then Sender IP Address was compared against White List IP Address.

If match found then attack factor of Rule 2 was set to 0.25. If not found then Attack Factor of the Rule 2 was set to 0. Assign resultant value in R2.

Rule 3 was applied on Fuzzy input parameter- Subject words. An E-mail may contain one or more words in subject line. All subject word and Content words are pre-processed. The pre-process contains the following steps i.e. stemming, stop words elimination and tokenization. Stemming is the process of comparing the root forms of the searched terms to the documents in its database. Stop words elimination is the process of not considering certain words which will not affect the final result. Tokenization is defined as splitting of the words into small meaning full constituents

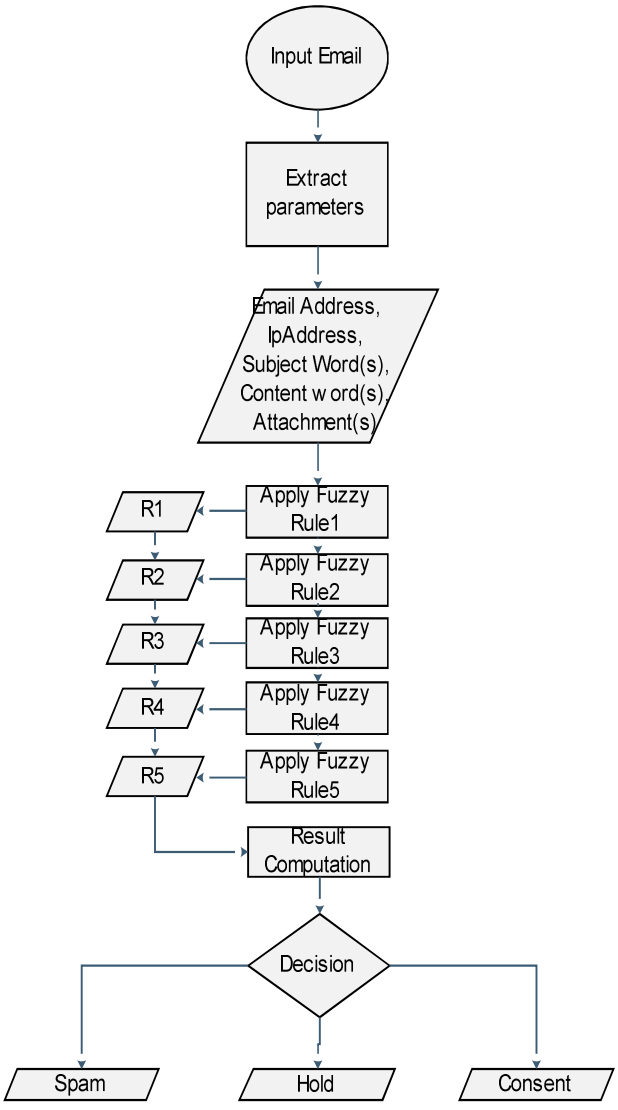

Figure 2. Detailed system flow

After pre processing all words are taken and compared against the Black list words. Every words impact (Attack Factor) on this subject line was calculated. Following are Algorithm to compute Attack Factor of Rule 3

Algorithm for E-mail Subject Attack Factor Calculation:

Step 1 :Split the Subject content into words say W_i where $n \geq i \geq 1$
Step 2 : assign to $T_{w=n}$
Step 3 :Calculate word Impact Factor W_f where $W_f = 0.5 / T_w$
Step 4 :Perform comparison for each word W_i in Black list
Step 5 :If match found then update the update $W_{fi} = -W_f$ else $W_{fi} = W_f$; where $i <= T_w$;
Step 6 : Calculate Attach Factor = $\sum Wfi$
Step 7 : Calculate R3 = $\sum Wfi$;

From the subject line after pre-processing total words are counted and each word impact on for this rule is calculate. i,e average impact. Now each word are compared against black and white list already available. If it is found in white list then the Attack factor for this word is set as positive. If it is found in black list then the Attack factor was set as negative.

Example :

Total words = 5
W_f = 0.5 / 5 = 0.1
If the word W_i is present in While list then the AttackFactor = + 0.1
If the word Wi is present in the Black list then the Attack Factor = - 0.1

Rule 4 was applied on Fuzzy Input variable-ContentWords after Pre-Processing. Every e-mail body may contain one or more words. Every words are taken and compared against the Block list words. Following are the Algorithm to compute Attack Factor of Rule 4.

Algorithm for E-mail Content Attack Factor Calculation:

Step 1 :Split the e-mail bodycontent to words say W_i where $i \geq 1$
Step 2: Count the total number of words in e-mail Bodyand assign to Tw
Step 3 : If $T_w > 0$ then continue Step 4.
Step 4 :Calculate word impact factor W_f where

$W_f = 0.5 / T_w$

Step 5 :Perform comparison for each word Wi in Black list
Step 6 :If match found then update the update $W_{fi} = -W_f$ else $W_{fi} = W_f$; where $i \leq Tw$;
Step 7 :Calculate Attach Factor = $\sum Wfi$
Step 8 : Calculate R4 = $\sum Wfi$;

Rule 5 was applied to calculate Attack Factor for e-mail containing attachment. If e-mail does not contain Attachment, then Attack Factor was set to zero. If any one of the attachment content was identified in virus list then Attack Factor was set to -1. If none of the content

was identified in virus list, then Attack Factor was set to 1. Rule 5 result was assigned to R5.

Defuzzification:

Result value of each e-mail was arrived by sum up previous rule results and these results are termed as decision making factors.
R1 = R1;
R2 = R2 + R1;
R3 = R3 + R2;
R4 = R4 + R3;
R5 = R5 + R4;

IV. Results based on User Attitude and Discussion

Every rule results are obtained and user attitude was taken consideration for categorizing input e-mails. User Attitude was initially configured to take decision based on fuzzy Linguistic set {High Positive, zero, high negative}. High positive users are type of user who strictly restricts spam emails. Zero level users are neutral user who does not have restriction. High negative users are more interested in receiving spam emails.

Following are the possible values of the Linguistic Set
High Positive ≥ 0.25;
Zero = 0;
High Negative ≤ -0.25

Following are the decision making process.

Decision making for High positive level users:

- If user's attitude was set as high positive and all applied rules values are > 0.25 then the e-mail is declared as consent.
- If user's attitude was set as high positive and any one of the rule result value various between 0.25 to 0 then the e-mail is declared as hold.
- If user's attitude was set as high positive and any one of the rule value is < 0 then the e-mail is set to Spam.

Decision making for Zero level user:

- If user's attitude level was set as Zero and all rule result value is ≥ 0 then the e-mail is declared as consent.
- If user's attitude level was set as Zero and any one of rule value is < 0 then the e-mail is set to Spam.

Decision making for High Negative level user:

- If user's attitude level was set High Negative and all rule result value is ≥ -0.25 then the e-mail is declared as consent.
- If user's attitude level was set as High Negative and any one the rule result value is < -0.25 then the e-mail is set to Hold in which user can take final decision.

All fuzzy rules are applied over 243 different kind of e-mails using fuzzy input variables: Sender's Address, Sender, Subject Words, Content Words and Attachment. Results of some sample e-mails are distributed in the following tables.

Table 1. Results based on Fuzzy Rules with High positive user Attitude

E-mail Source	Fuzzy Results					Result
	Rule1	Rule2	Rule3	Rule4	Rule5	
E1	0.25	0.5	1	1.5	2.5	Consent
E2	-0.25	0	0.5	1	2	Spam
E3	0	0.25	0.75	1.25	2.25	Hold
E4	0.25	0.25	0.75	1.25	2.25	Consent
E5	-0.25	-0.25	0.25	0.75	1.75	Spam
E6	0	0.25	0.25	0.75	1.75	Hold

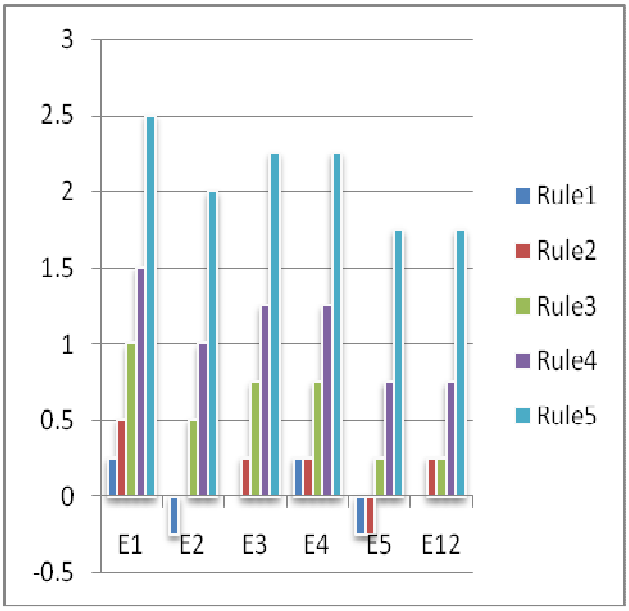

Figure 3. Graphical representation of Table 1

Table 2. Results based on Fuzzy user with Zero user Attitude

E-mail Source	Fuzzy Results					Result
	Rule1	Rule2	Rule3	Rule4	Rule5	
E1	0.25	0.5	1	1.5	2.5	Consent
E2	-0.25	0	0.5	1	2	Spam
E3	0	0.25	0.75	1.25	2.25	Consent
E4	0.25	0.25	0.75	1.25	2.25	Consent
E5	-0.25	-0.25	0.25	0.75	1.75	Spam
E6	0	0	0.5	1	2	Consent

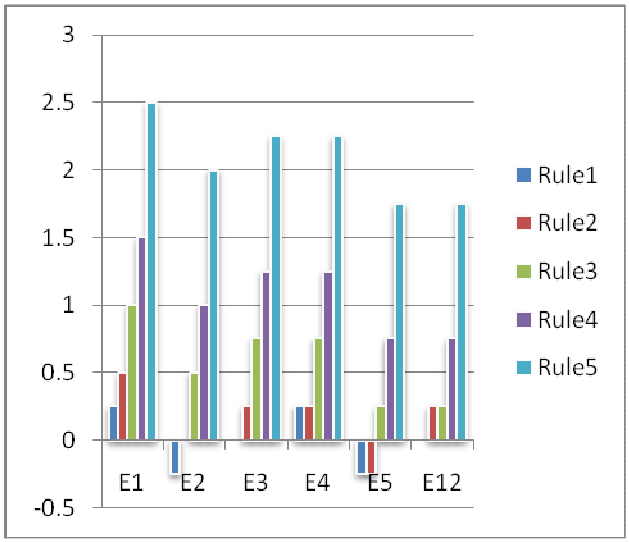

Figure 4. Graphical representation of Table2

From Fig 4 represents e-mail nature. If we see any e-mail that has negative region then the e-mail is set to spam. From the graph we can identify E2 and E5 are spam as it grows in negative region.

Table 3. Results based on Fuzzy Rules with high negative user Attitude

E-mail Source	Fuzzy Results					Result
	Rule1	Rule2	Rule3	Rule4	Rule5	
E44	-0.25	-0.5	-0.5	-0.5	0.5	Spam
E45	0	-0.25	-0.25	-0.25	0.75	Consent
E46	0.25	0.5	0	0	1	Consent
E47	-0.25	0	-0.5	-0.5	0.5	Spam
E48	0	0.25	-0.25	-0.25	0.75	Consent
E49	0.25	0.25	-0.25	-0.25	0.75	Consent

From table 3 the results can be easily predicted that the relaxation of user who intentionally wish to accept spam emails, user level was set to -0.25. So the range from -0.25 and above the e-mails are categorized as Consent. Below the level, e-mails are categorized as Hold. The same was represented in a graphical manner in Fig 4.

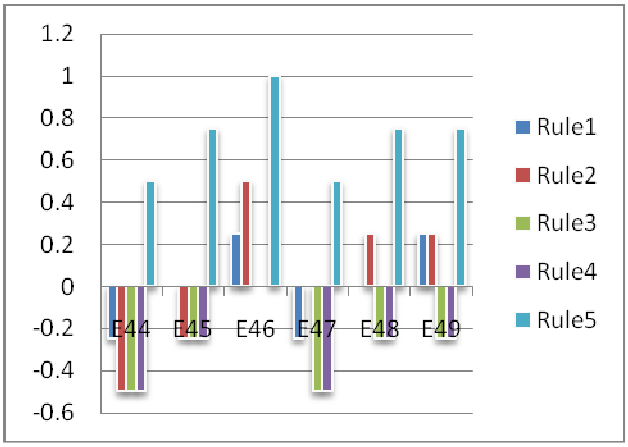

Figure 5. Graphical representation of Table 3

Table 4 : Results based on Fuzzy Rules with different user Attitudes

ES	R1	R2	R2	R4	R5	HP	Z	HN
E13	0.25	0.25	0.25	0.75	1.75	Consent	Consent	Consent
E15	0	0	0	0.5	1.5	Hold	Consent	Consent
E16	0.25	0	0	0.5	1.5	Hold	Consent	Consent
E17	-0.25	-0.5	-0.5	0	1	Spam	Spam	Spma
E18	0	-0.25	-0.25	0.25	1.25	Spam	Spam	Consent
E19	0.25	0.5	0	0.5	1.5	Hold	Consent	Consent

ES – E-mail Source
RX – Rule X where X varies from 1 to 5
HP – High Positive user Attitude
Z - Zero user Attitude
HN – High Negative user Attitude

Table 4 Consolidates different user projections on the same e-mail with e-mail samples. All possible e-mail combination results are provided in Appendix-1

243 different sets of emails are taken for evaluation and results are represented in following figures

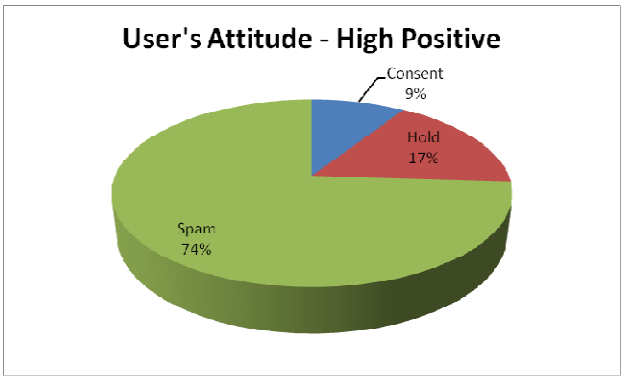

Figure 6. High positive user's attitude

Out of 243 e-mails based on high positive user's attitude, 22 e-malis are categorized as Consent, 41 e-malis are

categorized in Hold state and 180 emails are stamped as spam.

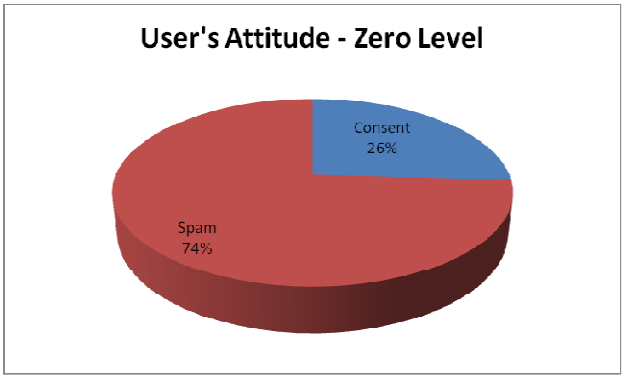

Figure 7. Zero level user's attitude

Out of 243 e-mails based on zero level user's attitude, 68 e-malis are categorized as Consent and 180 emails are stamped as spam.

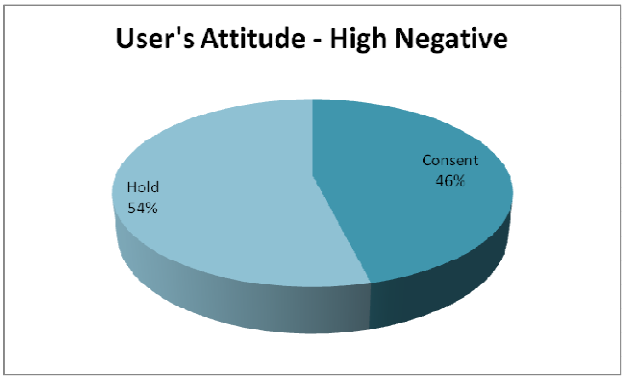

Figure 8. High Negative user's attitude

Out of 243 e-mails based on High Negative user's attitude, 112 e-malis are categorized as Consent and 131 emails are categorized as Hold.

CONCLUSION AND FUTURE WORK

In this proposed work, Fuzzy rules are constructed for 5 input parameters namely Sender's Address, Sender_IP, Subject_Words, Content Words and Attachment for common user to deduct the spam e-mails based on the attitude of the user. The proposed simplistic approach out performs in terms of accuracy in deducting spam e-mails than the existing approaches provided the Black list and White lists to be up to date. The proposed approach works only for e-mails having subject and body content as plain text. Future work aims at deducting spam emails having images and HTML also.

Acknowledgment

The authors would like to thank Dr. Ponnammal Natarajan worked as Former Director – Research , Anna University- Chennai,India and currently an Advisor, (Research and Development), Rajalakshmi Engineering College and Dr. K..Ravi, Associate Professor, Department of Mathematics, Sacred Heart College-Tirupattur, India for their intuitive ideas and fruitful discussions with respect to the paper's contribution.

REFERENCES

[1] E-mail Metrics report http://www.maawg.org/e-mail_metrics_report
[2] Carreras, X. and Mdrquez, L., "Boosting trees for anti-spare e-mail filtering", In Proc. of RANLP, 2001.
[3] Cohen, W.W., "Learning Rules that Classify E-Mail.", Proceedings. of the AAAI Spring Symposium on Machine Learning in Information Access, Stanford, California,1996.
[4] Cournane, A. and Hunt, R., "An Analysis of the Tools Used For the Generation and Prevention of Spam", Computer and Security, Vol. 23, pp 154-166, 2004.
[5] Cox, E., "The Fuzzy System Handbook", Academic Press, Second Edition, 1999.
[6] Daelemans, W., Z. Jakub, K. van der
[7] Sloot and A. van den Bosch, TiMBL: Tilburg Memory Based Learner, version 2.0, Reference Guide. ILK,Computational Linguistics, Tilburg University. http:/ilk.kub.nl/~ilk/papers/ilk9901.ps.gz, 1999.
[8] Drucker, H., Wu, D., & Vapnik, V., Support vector machines for Spam categorization. IEEE-NN, Vol. 10, No.5, pp. 1048–1054,1999.
[9] Graham, P., Better Baysian Filtering. In Proceedings of Spam Conference http://spamconference.org/proceedings2003.html, 2003.
[10] Peng Liu, Guangliang Chen, Liang Ye, Weiming Zhong, Proceedings of the 5th WSEAS Int. Conf. On Simulation, Modeling and Optimization, Corfu, Greece, August 17-19, 2005 (pp61-66).
[11] Wanli Ma, Dat Tran, Dharmendra Sharma, Sen Li, Proceedings of the 2007 WSEAS International Conference on Computer Engineering and Applications, Gold Coast, Australia, January 17-19, 2007 533.
[12] Hidalgo, J. G., Spez, M, and Sanz, E, Combining text and heuristicz for cost-sensitive spam filtering. In Proc. of CONL, 2000.
[13] Sadegh Kharazmi, Ali FarahmandNejad, Proceeding of the 9th WSEAS Int. Conference on Data Networks, Communications, Computers, Trinidad and Tobago, November 5-7, 2007.
[14] Nikolaos Korfiatis, Marios Poulosy, Sozon Papavlassopoulos, Proceeding of the WSEAS International Conference on Applied Mathematics, Greece, Aug 19, 2004 (488-429).
[15] Lee, J., "Spam: An escalating attack of the clones", The New York Times, 2002.
[16] Mayer, C., and Eunjung-Cha, A., "Making spam go splat: Sick of unsolicited e-mail,
[17] businesses are fighting back", The Washington Post, 2002.
[18] Norvig P. and Russell S., "Artificial Intelligence A Modern Approach", Prentice Hall, New Jersey, 2003.
[19] Nozaki, K., Ishibushi, H. and Tanaka, H., "Trainable Fuzzy classification systems based on Fuzzy If-Then-Rules," Proc. IEEE, vol. 1, pp. 498-502, 1994.
[20] RFC 822: Standard for the Format of Arpa Internet Text Messages, www.w3.org/Protocols/rfc822/, 1994.
[21] Sahami, M., Dumais, S., Heckerman, D. and Horvitz, E., "A Bayesian Approach to Filtering Junk E-Mail. In Learning for Text Categorization", AAA1 Workshop, pp. 55-62, Madison Wisconsin, 1998.
[22] SpamAssassin, www.spamassassin.org, 2004.

[23] Sudhakar.P, Poonkuzhali.S, Thiagarajan.K and Sarukesi.K., "Fuzzy Logic for E-mail Spam deduction", Proceedings of the WSEAS 10th International Conference on Applied Computer and Applied Computational Science, Venice, Italy, March 8-10, 2011 ISBN: 978-960-474-281-3

[24] Poonkuzhali.S, Thiagarajan.K, P.Sudhakar Kishore Kumar.R and Sarukesi.K., "Spam Filtering using Signed and Trust Reputation Management", Proceedings of the WSEAS 10th International Conference on Applied Computer and Applied Computational Science, Venice, Italy, March 8-10, 2011 ISBN: 978-960-474-281-3

P.Sudhakar received Bachelor of Engineering degree in Computer science from Anna University Chennai-India in 2006 and Master of Engineering degree in Computer Science from Anna University Chennai-India in 2008. He started his carrier as a Junior software programmer in Vernalis systems Pvt Ltd, Chennai India at 2008 and elevated to Associate software. He also presented various papers in National level conferences and published his research work in International Conferences and Journals.

G.Poonkuzhali received B.E degree in Computer Science and Engineering from University of Madras, Chennai, India, in 1998, and the M.E degree in Computer Science and Engineering from Sathyabama University, Chennai, India, in 2005. Currently she is pursuing Ph.D programme in the Department of Information and Communication Engineering at Anna University – Chennai, India. She has presented and published 10 research papers in international conferences & journals and authored 5 books. She is a life member of ISTE (Indian Society for Technical Education) ,IAENG (International Association of Engineers), and CSI (Computer Society of India).

K.Thiagarajan working as Senior Lecturer in the Department of Mathematics in KCG College of Technology - Chennai-India. He has totally 14 years of experience in teaching. He has attended and presented research articles in 33 National and International Conferences and published one national journal and 26 international journals. Currently he is working on web mining through automata and set theory. His area of specialization is coloring of graphs and DNA Computing.

Dr. K. Sarukesi has a very distinguished career spanning of nearly 40 years. He has a vast teaching experience in various universities in India and abroad. He was awarded a commonwealth scholarship by the association of common wealth universities, London for doing Ph.D in UK. He completed his Ph.D from the University of Warwick – U.K in the year 1982. His area of specializations is Technological Information System. He worked as expert in various foreign universities. He has executed number of consultancy projects. he has been honored and awarded commendations for his work in the field of information technology by the government of TamilNadu. He has published over 40 research papers in international conferences/journals and 40 National Conferences/journals.

Appendix -1

E-mail Source	Fuzzy Rules					High Positive			Zero		High Negative	
	Rule1	Rule2	Rule3	Rule4	Rule5	Consent	Hold	Spam	Consent	Spam	Consent	Hold
E1	0.25	0.5	1	1.5	2.5	YES			YES		YES	
E2	-0.25	0	0.5	1	2			YES		YES	YES	
E3	0	0.25	0.75	1.25	2.25		YES		YES		YES	
E4	0.25	0.25	0.75	1.25	2.25	YES			YES		YES	
E5	-0.25	-0.25	0.25	0.75	1.75			YES		YES	YES	
E6	0	0	0.5	1	2		YES		YES		YES	
E7	0.25	0	0.5	1	2		YES		YES		YES	
E8	-0.25	-0.5	0	0.5	1.5			YES		YES		YES
E9	0	-0.25	0.25	0.75	1.75			YES		YES	YES	
E10	0.25	0.5	0.5	1	2	YES			YES		YES	
E11	-0.25	0	0	0.5	1.5			YES		YES	YES	
E12	0	0.25	0.25	0.75	1.75		YES		YES		YES	
E13	0.25	0.25	0.25	0.75	1.75	YES			YES		YES	
E14	-0.25	-0.25	-0.25	0.25	1.25			YES		YES	YES	
E15	0	0	0	0.5	1.5		YES		YES		YES	
E16	0.25	0	0	0.5	1.5		YES		YES		YES	
E17	-0.25	-0.5	-0.5	0	1			YES		YES		YES
E18	0	-0.25	-0.25	0.25	1.25			YES		YES	YES	
E19	0.25	0.5	0	0.5	1.5		YES		YES		YES	
E20	-0.25	0	-0.5	0	1			YES		YES		YES
E21	0	0.25	-0.25	0.25	1.25			YES		YES	YES	
E22	0.25	0.25	-0.25	0.25	1.25			YES		YES	YES	
E23	-0.25	-0.25	-0.75	-0.25	0.75			YES		YES		YES
E24	0	0	-0.5	0	1			YES		YES		YES
E25	0.25	0	-0.5	0	1			YES		YES		YES
E26	-0.25	-0.5	-1	-0.5	0.5			YES		YES		YES
E27	0	-0.25	-0.75	-0.25	0.75			YES		YES		YES
E28	0.25	0.5	1	1	2	YES			YES		YES	
E29	-0.25	0	0.5	0.5	1.5			YES		YES	YES	
E30	0	0.25	0.75	0.75	1.75		YES		YES		YES	
E31	0.25	0.25	0.75	0.75	1.75	YES			YES		YES	
E32	-0.25	-0.25	0.25	0.25	1.25			YES		YES	YES	
E33	0	0	0.5	0.5	1.5		YES		YES		YES	
E34	0.25	0	0.5	0.5	1.5		YES		YES		YES	
E35	-0.25	-0.5	0	0	1			YES		YES		YES
E36	0	-0.25	0.25	0.25	1.25			YES		YES	YES	
E37	0.25	0.5	0.5	0.5	1.5	YES			YES		YES	
E38	-0.25	0	0	0	1			YES		YES	YES	
E39	0	0.25	0.25	0.25	1.25		YES		YES		YES	
E40	0.25	0.25	0.25	0.25	1.25	YES			YES		YES	
E41	-0.25	-0.25	-0.25	-0.25	0.75			YES		YES	YES	

E42	0	0	0	0	1		YES		YES		YES	
E43	0.25	0	0	0	1		YES		YES		YES	
E44	-0.25	-0.5	-0.5	-0.5	0.5			YES		YES		YES
E45	0	-0.25	-0.25	-0.25	0.75			YES		YES	YES	
E46	0.25	0.5	0	0	1		YES		YES		YES	
E47	-0.25	0	-0.5	-0.5	0.5			YES		YES		YES
E48	0	0.25	-0.25	-0.25	0.75			YES		YES	YES	
E49	0.25	0.25	-0.25	-0.25	0.75			YES		YES	YES	
E50	-0.25	-0.25	-0.75	-0.75	0.25			YES		YES		YES
E51	0	0	-0.5	-0.5	0.5			YES		YES		YES
E52	0.25	0	-0.5	-0.5	0.5			YES		YES		YES
E53	-0.25	-0.5	-1	-1	0			YES		YES		YES
E54	0	-0.25	-0.75	-0.75	0.25			YES		YES		YES
E55	0.25	0.5	1	0.5	1.5	YES			YES		YES	
E56	-0.25	0	0.5	0	1			YES		YES	YES	
E57	0	0.25	0.75	0.25	1.25		YES		YES		YES	
E58	0.25	0.25	0.75	0.25	1.25	YES			YES		YES	
E59	-0.25	-0.25	0.25	-0.25	0.75			YES		YES	YES	
E60	0	0	0.5	0	1		YES		YES		YES	
E61	0.25	0	0.5	0	1		YES		YES		YES	
E62	-0.25	-0.5	0	-0.5	0.5			YES		YES		YES
E63	0	-0.25	0.25	-0.25	0.75			YES		YES	YES	
E64	0.25	0.5	0.5	0	1		YES		YES		YES	
E65	-0.25	0	0	-0.5	0.5			YES		YES		YES
E66	0	0.25	0.25	-0.25	0.75			YES		YES	YES	
E67	0.25	0.25	0.25	-0.25	0.75			YES		YES	YES	
E68	-0.25	-0.25	-0.25	-0.75	0.25			YES		YES		YES
E69	0	0	0	-0.5	0.5			YES		YES		YES
E70	0.25	0	0	-0.5	0.5			YES		YES		YES
E71	-0.25	-0.5	-0.5	-1	0			YES		YES		YES
E72	0	-0.25	-0.25	-0.75	0.25			YES		YES		YES
E73	0.25	0.5	0	-0.5	0.5			YES		YES		YES
E74	-0.25	0	-0.5	-1	0			YES		YES		YES
E75	0	0.25	-0.25	-0.75	0.25			YES		YES		YES
E76	0.25	0.25	-0.25	-0.75	0.25			YES		YES		YES
E77	-0.25	-0.25	-0.75	-1.25	-0.25			YES		YES		YES
E78	0	0	-0.5	-1	0			YES		YES		YES
E79	0.25	0	-0.5	-1	0			YES		YES		YES
E80	-0.25	-0.5	-1	-1.5	-0.5			YES		YES		YES
E81	0	-0.25	-0.75	-1.25	-0.25			YES		YES		YES
E82	0.25	0.5	1	1.5	1.5	YES			YES		YES	
E83	-0.25	0	0.5	1	1			YES		YES	YES	
E84	0	0.25	0.75	1.25	1.25		YES		YES		YES	
E85	0.25	0.25	0.75	1.25	1.25	YES			YES		YES	
E86	-0.25	-0.25	0.25	0.75	0.75			YES		YES	YES	

E87	0	0	0.5	1	1		YES		YES		YES	
E88	0.25	0	0.5	1	1		YES		YES		YES	
E89	-0.25	-0.5	0	0.5	0.5			YES		YES		YES
E90	0	-0.25	0.25	0.75	0.75			YES		YES	YES	
E91	0.25	0.5	0.5	1	1	YES			YES		YES	
E92	-0.25	0	0	0.5	0.5			YES		YES	YES	
E93	0	0.25	0.25	0.75	0.75		YES		YES		YES	
E94	0.25	0.25	0.25	0.75	0.75	YES			YES		YES	
E95	-0.25	-0.25	-0.25	0.25	0.25			YES		YES	YES	
E96	0	0	0	0.5	0.5		YES		YES		YES	
E97	0.25	0	0	0.5	0.5		YES		YES		YES	
E98	-0.25	-0.5	-0.5	0	0			YES		YES		YES
E99	0	-0.25	-0.25	0.25	0.25			YES		YES	YES	
E100	0.25	0.5	0	0.5	0.5		YES		YES		YES	
E101	-0.25	0	-0.5	0	0			YES		YES		YES
E102	0	0.25	-0.25	0.25	0.25			YES		YES	YES	
E103	0.25	0.25	-0.25	0.25	0.25			YES		YES	YES	
E104	-0.25	-0.25	-0.75	-0.25	-0.25			YES		YES		YES
E105	0	0	-0.5	0	0			YES		YES		YES
E106	0.25	0	-0.5	0	0			YES		YES		YES
E107	-0.25	-0.5	-1	-0.5	-0.5			YES		YES		YES
E108	0	-0.25	-0.75	-0.25	-0.25			YES		YES		YES
E109	0.25	0.5	1	1	1	YES			YES		YES	
E110	-0.25	0	0.5	0.5	0.5			YES		YES	YES	
E111	0	0.25	0.75	0.75	0.75		YES		YES		YES	
E112	0.25	0.25	0.75	0.75	0.75	YES			YES		YES	
E113	-0.25	-0.25	0.25	0.25	0.25			YES		YES	YES	
E114	0	0	0.5	0.5	0.5		YES		YES		YES	
E115	0.25	0	0.5	0.5	0.5		YES		YES		YES	
E116	-0.25	-0.5	0	0	0			YES		YES		YES
E117	0	-0.25	0.25	0.25	0.25			YES		YES	YES	
E118	0.25	0.5	0.5	0.5	0.5	YES			YES		YES	
E119	-0.25	0	0	0	0			YES		YES	YES	
E120	0	0.25	0.25	0.25	0.25		YES		YES		YES	
E121	0.25	0.25	0.25	0.25	0.25	YES			YES		YES	
E122	-0.25	-0.25	-0.25	-0.25	-0.25			YES		YES	YES	
E123	0	0	0	0	0		YES		YES		YES	
E124	0.25	0	0	0	0		YES		YES		YES	
E125	-0.25	-0.5	-0.5	-0.5	-0.5			YES		YES		YES
E126	0	-0.25	-0.25	-0.25	-0.25			YES		YES	YES	
E127	0.25	0.5	0	0	0		YES		YES		YES	
E128	-0.25	0	-0.5	-0.5	-0.5			YES		YES		YES
E129	0	0.25	-0.25	-0.25	-0.25			YES		YES	YES	
E130	0.25	0.25	-0.25	-0.25	-0.25			YES		YES	YES	
E131	-0.25	-0.25	-0.75	-0.75	-0.75			YES		YES		YES

E132	0	0	-0.5	-0.5	-0.5			YES		YES		YES
E133	0.25	0	-0.5	-0.5	-0.5			YES		YES		YES
E134	-0.25	-0.5	-1	-1	-1			YES		YES		YES
E135	0	-0.25	-0.75	-0.75	-0.75			YES		YES		YES
E136	0.25	0.5	1	0.5	0.5	YES			YES		YES	
E137	-0.25	0	0.5	0	0			YES		YES	YES	
E138	0	0.25	0.75	0.25	0.25		YES		YES		YES	
E139	0.25	0.25	0.75	0.25	0.25	YES			YES		YES	
E140	-0.25	-0.25	0.25	-0.25	-0.25			YES		YES	YES	
E141	0	0	0.5	0	0		YES		YES		YES	
E142	0.25	0	0.5	0	0		YES		YES		YES	
E143	-0.25	-0.5	0	-0.5	-0.5			YES		YES		YES
E144	0	-0.25	0.25	-0.25	-0.25			YES		YES	YES	
E145	0.25	0.5	0.5	0	0		YES		YES		YES	
E146	-0.25	0	0	-0.5	-0.5			YES		YES		YES
E147	0	0.25	0.25	-0.25	-0.25			YES		YES	YES	
E148	0.25	0.25	0.25	-0.25	-0.25			YES		YES	YES	
E149	-0.25	-0.25	-0.25	-0.75	-0.75			YES		YES		YES
E150	0	0	0	-0.5	-0.5			YES		YES		YES
E151	0.25	0	0	-0.5	-0.5			YES		YES		YES
E152	-0.25	-0.5	-0.5	-1	-1			YES		YES		YES
E153	0	-0.25	-0.25	-0.75	-0.75			YES		YES		YES
E154	0.25	0.5	0	-0.5	-0.5			YES		YES		YES
E155	-0.25	0	-0.5	-1	-1			YES		YES		YES
E156	0	0.25	-0.25	-0.75	-0.75			YES		YES		YES
E157	0.25	0.25	-0.25	-0.75	-0.75			YES		YES		YES
E158	-0.25	-0.25	-0.75	-1.25	-1.25			YES		YES		YES
E159	0	0	-0.5	-1	-1			YES		YES		YES
E160	0.25	0	-0.5	-1	-1			YES		YES		YES
E161	-0.25	-0.5	-1	-1.5	-1.5			YES		YES		YES
E162	0	-0.25	-0.75	-1.25	-1.25			YES		YES		YES
E163	0.25	0.5	1	1.5	0.5	YES			YES		YES	
E164	-0.25	0	0.5	1	0			YES		YES	YES	
E165	0	0.25	0.75	1.25	0.25		YES		YES		YES	
E166	0.25	0.25	0.75	1.25	0.25	YES			YES		YES	
E167	-0.25	-0.25	0.25	0.75	-0.25			YES		YES	YES	
E168	0	0	0.5	1	0		YES		YES		YES	
E169	0.25	0	0.5	1	0		YES		YES		YES	
E170	-0.25	-0.5	0	0.5	-0.5			YES		YES		YES
E171	0	-0.25	0.25	0.75	-0.25			YES		YES	YES	
E172	0.25	0.5	0.5	1	0		YES		YES		YES	
E173	-0.25	0	0	0.5	-0.5			YES		YES		YES
E174	0	0.25	0.25	0.75	-0.25			YES		YES	YES	
E175	0.25	0.25	0.25	0.75	-0.25			YES		YES	YES	
E176	-0.25	-0.25	-0.25	0.25	-0.75			YES		YES		YES

E177	0	0	0	0.5	-0.5			YES		YES		YES
E178	0.25	0	0	0.5	-0.5			YES		YES		YES
E179	-0.25	-0.5	-0.5	0	-1			YES		YES		YES
E180	0	-0.25	-0.25	0.25	-0.75			YES		YES		YES
E181	0.25	0.5	0	0.5	-0.5			YES		YES		YES
E182	-0.25	0	-0.5	0	-1			YES		YES		YES
E183	0	0.25	-0.25	0.25	-0.75			YES		YES		YES
E184	0.25	0.25	-0.25	0.25	-0.75			YES		YES		YES
E185	-0.25	-0.25	-0.75	-0.25	-1.25			YES		YES		YES
E186	0	0	-0.5	0	-1			YES		YES		YES
E187	0.25	0	-0.5	0	-1			YES		YES		YES
E188	-0.25	-0.5	-1	-0.5	-1.5			YES		YES		YES
E189	0	-0.25	-0.75	-0.25	-1.25			YES		YES		YES
E190	0.25	0.5	1	1	0		YES		YES		YES	
E191	-0.25	0	0.5	0.5	-0.5			YES		YES		YES
E192	0	0.25	0.75	0.75	-0.25			YES		YES	YES	
E193	0.25	0.25	0.75	0.75	-0.25			YES		YES	YES	
E194	-0.25	-0.25	0.25	0.25	-0.75			YES		YES		YES
E195	0	0	0.5	0.5	-0.5			YES		YES		YES
E196	0.25	0	0.5	0.5	-0.5			YES		YES		YES
E197	-0.25	-0.5	0	0	-1			YES		YES		YES
E198	0	-0.25	0.25	0.25	-0.75			YES		YES		YES
E199	0.25	0.5	0.5	0.5	-0.5			YES		YES		YES
E200	-0.25	0	0	0	-1			YES		YES		YES
E201	0	0.25	0.25	0.25	-0.75			YES		YES		YES
E202	0.25	0.25	0.25	0.25	-0.75			YES		YES		YES
E203	-0.25	-0.25	-0.25	-0.25	-1.25			YES		YES		YES
E204	0	0	0	0	-1			YES		YES		YES
E205	0.25	0	0	0	-1			YES		YES		YES
E206	-0.25	-0.5	-0.5	-0.5	-1.5			YES		YES		YES
E207	0	-0.25	-0.25	-0.25	-1.25			YES		YES		YES
E208	0.25	0.5	0	0	-1			YES		YES		YES
E209	-0.25	0	-0.5	-0.5	-1.5			YES		YES		YES
E210	0	0.25	-0.25	-0.25	-1.25			YES		YES		YES
E211	0.25	0.25	-0.25	-0.25	-1.25			YES		YES		YES
E212	-0.25	-0.25	-0.75	-0.75	-1.75			YES		YES		YES
E213	0	0	-0.5	-0.5	-1.5			YES		YES		YES
E214	0.25	0	-0.5	-0.5	-1.5			YES		YES		YES
E215	-0.25	-0.5	-1	-1	-2			YES		YES		YES
E216	0	-0.25	-0.75	-0.75	-1.75			YES		YES		YES
E217	0.25	0.5	1	0.5	-0.5			YES		YES		YES
E218	-0.25	0	0.5	0	-1			YES		YES		YES
E219	0	0.25	0.75	0.25	-0.75			YES		YES		YES
E220	0.25	0.25	0.75	0.25	-0.75			YES		YES		YES
E221	-0.25	-0.25	0.25	-0.25	-1.25			YES		YES		YES

E222	0	0	0.5	0	-1			YES		YES		YES
E223	0.25	0	0.5	0	-1			YES		YES		YES
E224	-0.25	-0.5	0	-0.5	-1.5			YES		YES		YES
E225	0	-0.25	0.25	-0.25	-1.25			YES		YES		YES
E226	0.25	0.5	0.5	0	-1			YES		YES		YES
E227	-0.25	0	0	-0.5	-1.5			YES		YES		YES
E228	0	0.25	0.25	-0.25	-1.25			YES		YES		YES
E229	0.25	0.25	0.25	-0.25	-1.25			YES		YES		YES
E230	-0.25	-0.25	-0.25	-0.75	-1.75			YES		YES		YES
E231	0	0	0	-0.5	-1.5			YES		YES		YES
E232	0.25	0	0	-0.5	-1.5			YES		YES		YES
E233	-0.25	-0.5	-0.5	-1	-2			YES		YES		YES
E234	0	-0.25	-0.25	-0.75	-1.75			YES		YES		YES
E235	0.25	0.5	0	-0.5	-1.5			YES		YES		YES
E236	-0.25	0	-0.5	-1	-2			YES		YES		YES
E237	0	0.25	-0.25	-0.75	-1.75			YES		YES		YES
E238	0.25	0.25	-0.25	-0.75	-1.75			YES		YES		YES
E239	-0.25	-0.25	-0.75	-1.25	-2.25			YES		YES		YES
E240	0	0	-0.5	-1	-2			YES		YES		YES
E241	0.25	0	-0.5	-1	-2			YES		YES		YES
E242	-0.25	-0.5	-1	-1.5	-2.5			YES		YES		YES
E243	0	-0.25	-0.75	-1.25	-2.25			YES		YES		YES