Backpropagation Learning Algorithms for Email Classification.

*David Ndumiyana and Tarirayi Mukabeta

Bindura University of Science Education
Faculty of Science, Computer Science Department
P. Bag 1020, Bindura, Zimbabwe.
*Email: ndumiyanad@gmail.com

Abstract - **Today email has become one the fastest and most effective form of communication. The popularity of this mode of transmitting goods, information and services has motivated spammers to perfect their technical skills to fool spam filters. This development has worsened the problems faced by Internet users as they have to deal with email congestion, email overload and unprioritised email messages. The result was an exponential increase in the number of email classification management tools for the past few decades. In this paper we propose a new spam classifier using a learning process of multi-layer neural network to implement back propagation technique. Our contribution to the body of knowledge is the use of an improved empirical analysis to choose an optimum, novel collection of attributes of a user's email contents that allows a quick detection of most important words in emails. We also demonstrate the effectiveness of two equal sets of emails training and testing data.**

Keywords: Back propagation algorithm, neural networks, machine learning, multilayer perceptron, false positives.

I. INTRODUCTION

The email has become the most powerful tool for use as a means of exchanging new ideas and information at a low cost and it guarantees of an efficient email delivery [16]. Its popularity and preference as a communication tool is largely due to its availability, reliability and user friendliness [19]. Spam is defined as an unsolicited and unwanted email usually from a stranger that is delivered in bulk large mailing list with some business intelligence objectives [18]. According to [9] spam can be grouped into the following categories:

- Health; such as fake pharmaceuticals
- Promotional products; such as fake fashion items like watches
- Adult content; pornography and prostitution
- Financial and refinancing; tax solution and loan package
- Education; online Diplomas and degrees
- Marketing: direct marketing material and sexual enhancement products

The exponential growth of spam has become a serious threat not only to the Internet but also to the business community, education and the society at large. The idea of email classification using message filtering systems cannot be overestimated but has to be tackled in a holistic manner by considering the source, network and end user [16]. A number of reports had been published by previous researchers on spam filtering which centred mainly on classification of spam email messages. Authors used an approach where a set of rules are created by either the user of the filter or by the software company providing a rule base spam filtering to classify email as spam or non-spam. This technique suffered one limitation in that the rules must constantly be updated and maintained which ended up de-regulating the system and waste time [2]. These rules could be updated in a centralised manner by the maintainer of the spam filtering tool and there is a peer-2-peer knowledge base solution, but when the rules are publicly available, the spammer has the ability to change the text of his message so that it would pass through a spam filter without detection. To alleviate these shortcomings many researchers are turning to machine learning for spam classification as the benefits of exploring teaching algorithms for spam classification are unavoidable [8]. In machine learning technique to spam filtering, a set of pre-classified email messages are used as training samples. An algorithm is then applied to learn the classification rules from the training samples[23]. In this paper we are going to use one of the popular machine learning algorithms called neural network which constitutes a back propagation algorithm. The subject of machine learning has been widely studied and there are lots of algorithms suitable for this task. To date the volume of unsolicited commercial email messages transmitted by the Internet has reached a large proportion of the total mail delivered on a daily basis. Spam messages have brought a lot of problems to both users and internet service providers. In the first place, spam occupies server storage space and consumes a lot of network bandwidth. On the second note, innocent users are forced to waste productive time identifying and removing spam from their computer systems. The simplest and most common solution for avoiding such discomfort is to use spam filters that screen messages based upon the presence of common key words or phrases common to junk emails.

II. Machine Learning Spam Filtering Technique

Machine learning is defined as a scientific discipline which is concerned with the design and development of algorithms that enables adaptation of computers to behaviour based on data. Machine learning is a subset from the broad field of Artificial Intelligence that aims at making machines able to learn like humans [3][19].

2.1 Artificial Neural Network for Spam Detection

Neural Network (NN) is a computational model based on biological neural network. It is an adaptive system that changes its structure based on information that flows through the artificial network during a learning phase and is based on the principle of learning.

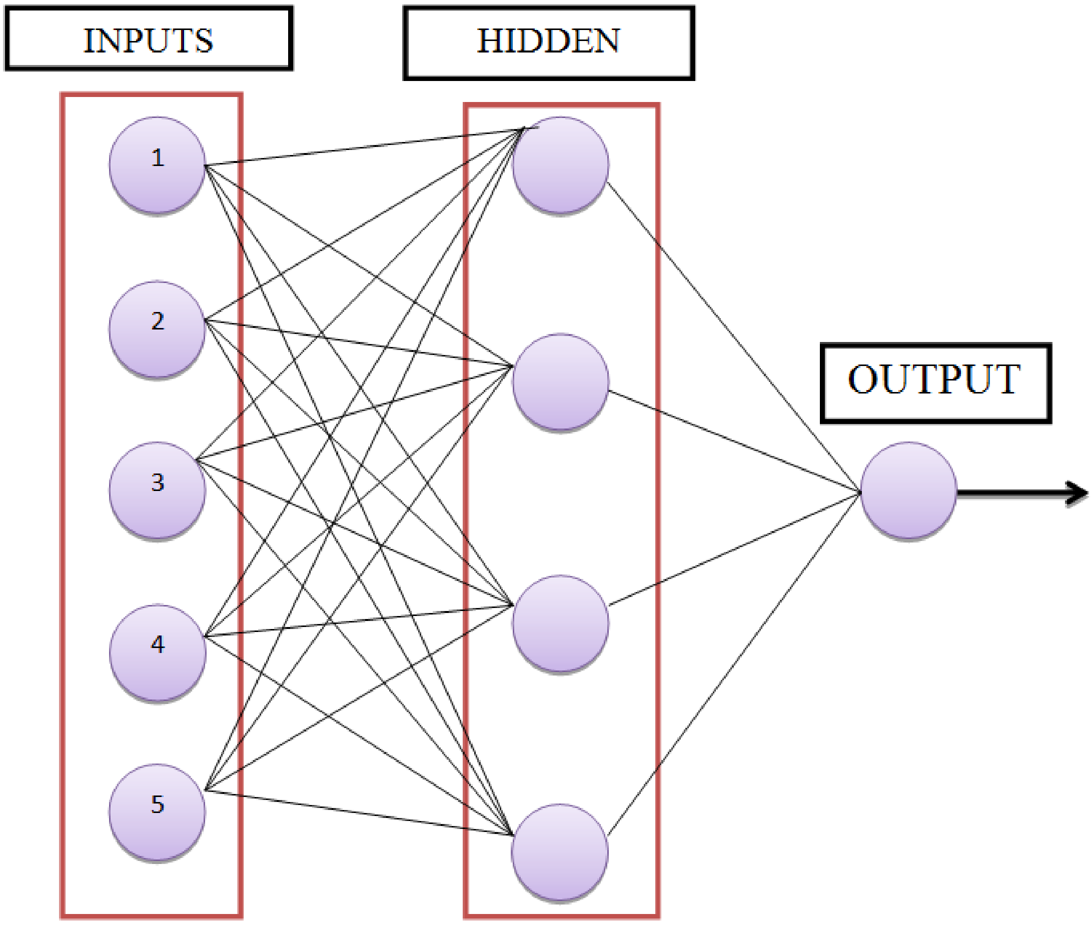

Figure 2.1Back-Propagation Neural Network Architecture

In their work [14] used perceptron algorithm to find a linear function of the feature vector

$f(x) = w^Tx + b$ such that $f(x) > 0$ for vectors of one class and $f(x) < 0$ for vectors of the other class. The perceptron learning is done with an iterative algorithm that starts with arbitrarily chosen parameters (w_0, b_0) of the decision and updates them repeatedly. A training sample

(x, c) is chosen on the n-th iteration of the algorithm such that the current decision function does not classify the outcome correctly. The parameter (x_n, b_n) are updated and the algorithm stops when a decision function is found that correctly classifies all the training samples. In his research [6] classified spam using LINGER, a neural network-based system which uses a multilayer perceptron. The results obtained from their report show that neural network based filters achieve better accuracy.

Our motivation in this proposed spam filtering system is to use a neural network to classify spam (unsolicited emails) and ham (wanted, personal messages) emails. The spam filtering problem is broken down into a simple classification problem so that a time-tested network s algorithm, Back propagation can be used effectively. This paper therefore describes the effectiveness of email classifiers based on the feed-forward back-propagation neural network. The results we obtained show that feed-forward back-propagation network algorithm classifier provides a relatively high accuracy that compares favourably with the best known classifiers.

III. THE IMPACT OF UNSOLICITED EMAILS

The users, Internet service providers and organisations as potential targets of spam find it virtually impossible to tell exactly the first person to come upon a simple idea that if you send out an advertisement to millions of people, then at least one person will react to it no matter what the proposal may be. Email provides the perfect platform to send these millions of advertisements at no cost for the sender, an unfortunate fact that is extensively taken advantage of by several organizations. The end result, email inboxes of millions of people get clustered with unsolicited email messages. Since it is incredibly cheap to send, spam causes a lot of trouble to the Internet community and society at large. Some of the most notable problems are described in this section.

3.1Spam Costs Money

The [26] report for 2009 highlights spam levels reaching 87.7%, with compromised computers issuing 83.4% of the 107 billion spam messages distributed globally per day on average [26]. A 2009 study by Ferris Research estimated an increase in spam cost to a total of $130 billion dollars worldwide. The study indicated that the main cost occurs due to the following reasons:

- Productivity loss due to inspecting and deleting spam that gets missed by spam control products (false negatives).
- Productivity loss from searching for legitimate email deleted by error in by spam control products (false positives).
- Operations and helpdesk running cost[11]

3.2 Spam Wastes Storage Space

The perpetual flow of spam floods up users' inboxes thus forcing users to waste productive time deleting these unsolicited email messages, resulting in displacement of critical or legitimate emails. Moreover spam also causes the loss of Internet performance and bandwidth due to increased payload on the network [10] and it clogs up email servers to the point where the server is forced to crash.

3.3 Spam Spreads Malware

There has been a marked increase in the transmission of malware and viruses thus widening threats to network security and personal privacy [7]. According to a report by MessageLabs, 2009, emails infected with viruses for 2009 was 1 in 286.4 emails and more than 73.1 million malware infected emails containing over 2500 different malware strains, were detected and blocked [25]

3.4 Spam and Identity Theft

Authors of spam have gone a gear up in their quest to target as many Internet users as they can by deploying unsolicited email messages to gain personal information about the user for fraudulent purposes. Phishing activities related to identity theft and other forms of Internet frauds (Nigeria 419) have become the latest concerns for the Internet users. A technical report by MessageLabs researchers showed that the proportion of phishing attacks in emails traffic was 1 in 325.2 constituting 0.31% emails and estimated 161 billion email phishing attacks were in circulation in 2009. Thus researchers, business community and internet service providers all unanimously agree to the fact that the growing threats of spam definitely require desperate and drastic control measures.

IV. PREVIOUS LITERATURE ON MACHINE LEARNING

The exponential growth of unsolicited email messages in recent times has resulted in the necessity for more accurate and effective spam filtering systems. Machine learning is a very effective technique that has been successfully used in email classification. Allowing machines to classify email into spam and non-spam messages relieves human intervention thus reducing the cost of monitoring spam.

The author [5] classified spam using LINGER, a neural network-based system which uses a multilayer perceptron. LINGER consists of two feature selectors namely information gain (IG) and variance (V). Their results show that neural network-based spam filters achieve better accuracy in the training phase but has unstable portability across different corpora[5].

The authors[13] described the results of an empirical study on two spam detection methods namely Support Vector Machines (SVMs) and Naïve Bayesian Classifier (NBC). They used both term frequency (TF) and term frequency with inverse document frequency (TF-IDF) for features vector construction. Their results showed that Naïve Bayesian has a consistent performance for all ranges of data sets.

Researchers [22] used the integration of two linear classifiers, Perceptron and Winnow. The results produced showed that Winnow performed slightly better than Perceptron although both classifiers performed very well and in the process outperformed Naïve Bayesian classifier.

Researchers[12] used binary classification based on an extension of Bayes point machines. By using the Bayesian approach with inference expectation propagation (EP) they produced results that outperformed [23]SVM. [23] used a hybrid method of rule-based processing and back-propagation neural network for spam filtering and the system produced very low false positives and negative rates and with better results compared with content-based classification [11]

In their contribution to the body of world knowledge, [20] presented a new technique for filtering image spam using gradient histogram as a key technique in feed-forward back-propagation algorithm. Experimental results indicated that gradient histogram based image spam classification provided good results.

The author [4]proposed a new email classification model using a linear neural network trained by Perceptron Learning Algorithm (PLA) and a non-linear neural network trained by Back Propagation neural network (BPNN). A Semantic Feature Space (SFS) technique was also introduced for the first time in this classification model. In addition, a rule-based system as reported by Schuff et al [2] can provide straight forward method to semi automate email classification and such a system requires a user to define a set of instructions for the email application to sort incoming messages into folders and order them by priority.

In a related issue, [6] also proposed a new approach by automatically assessing incoming messages and making some recommendations before emails reach the user's inbox, hence the priority system classifies each message as of either high or low importance based on its expected importance to the user.

This literature survey could not be adequate if we leave out [17] who further elaborated about email classification saying depending on the mechanism used, email classification systems can be broadly classified into, Rule based classification, Information Retrieval based classification and Machine Learning based classification techniques.

The author [1] considered the classification of email messages by using Back-Propagation method. They used the process of cross validation measurement n different times, applying the model to predict the classification of email messages and applied neural network approach in their classification. The results of their findings indicated that if Back-Propagation were appropriate for a few received email messages, then it would achieve a 98% success compared with human judgement. Not to be outdone was the work done by [21] who introduced Optical Back-Propagation which is a type of Back-Propagation algorithm. One of the noted characteristic of this algorithm was the escape from the local minimum during the course of training with high speed. The authors used two different structures for Optical Back-Propagation namely OBP structure-1 and OBP structure-2. Their finding was that the first structure performed better than the second structure.

V. RESEARCH METHOD

The technique we implemented for our email classification into classes is a supervised learning. Our dominant classes in this paper are: Education, Fraud, Finance, Internet and Adult. This solution is based on a heuristic approach and on the fact that if an email is about:

- Obtaining an online academic Certificate, Diploma, Degree or Job training, then our classifier should classify it as Education.
- Getting a USA work permit, winning very big prize money in a lottery or latest information about investing on stock exchange then our system should classify this as Fraud.
- A request to send your credit cards numbers, money for service charges so as to process a transaction for you as a beneficiary, invitation to apply for a loan or taking an insurance policy, then our classifier should classify it as Finance.
- Web hosting or e-marketing, then our classifier should classify such email content spam as Internet.
- Dating services and pornography is classified by our classifier as Adult.

We went on to implement a neural network [14] based system for automatic email classification into user defined 'word classes' and our experiment was based on email content. The classes were words with meaning (Education, Fraud, Finance, Internet and Adult). In addition we went on to investigate the impact of various feature selection and the use of Back Propagation for the email datasets.

5.1 The Learning Process

Our learning technique is trained by giving it input and matching output patterns. The input-output pairs can be provided by the system that contains the neural network. The learning process described in this section is implemented in this work. Figure 5.1 below shows the sample learning process developed for our system. A neural network (NN) which has the ability to learn by example was implemented. According to Habra [15] Back propagation is a popular type of network that can be trained to recognise different patterns such as image, text and signals. When the user has created the data, this data is used to run a multilayer neural network. The inputs of NN are the important words and the output is the decision (spam/non-spam). Each word in the email message represents an input node in the neural network and the number of neurons in the first layer equals the number of words in the input vector.

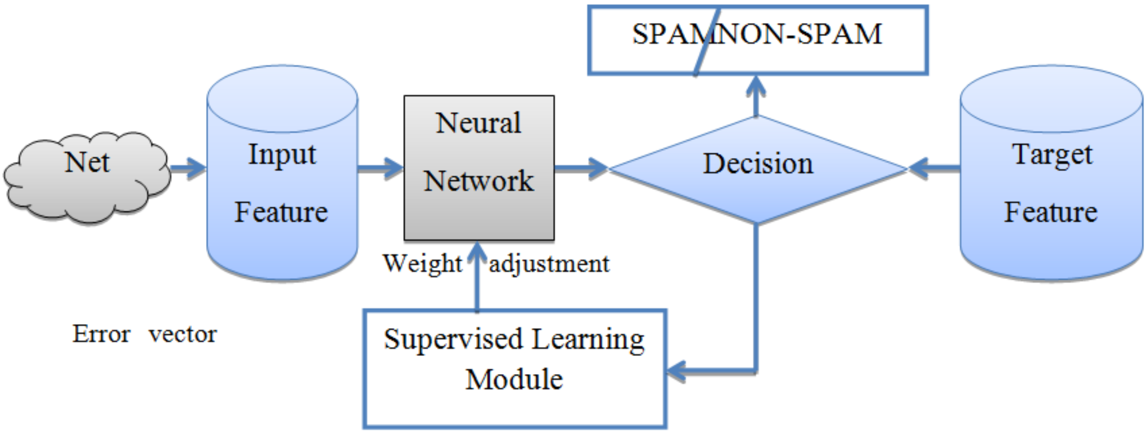

Figure 5.1: The Supervised Learning Process

For the output layer, there are five nodes. If the email is finance, the first output node gets value 1and the rest 0, and if the email is fraud, the first output gets 1, and the rest 0. The first class picked by the classifier has its output node set to 1 and the rest 0. The preference for 1 is given to the first class detected, and the rest 0 and so on. What matters most here is to have input-output data and the data comes from the email message (important words in the email content). Word extraction from email messages is based on information retrieval method according to the report described by Ramos [6]. The most important words which determine the success factor of our classifier were selected using the formula tf-idf.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the performance of our email classifier, we measured the number of accurate and inaccurate classifications and the total number of incorrect classification when compared with human judgement. We took the rate of success as the average of accurate predictions over all data used. Emails from a given user were separated into two equal datasets of training and testing. The training data was used to train the neural network (NN) and our email classifier knew the patterns that made emails to be classified as Education, Fraud, Finance, Internet and Adult. We discovered that our solution learned well when the training error was decreasing. The state and quality of results from experimental activities are shown in Table 6.1 below. In the table the numbers shown as accuracy are the numbers correctly classified emails and inaccurately classified mails over all test data.

Table 6.0: Email Classification Results

N^o of NN	HITS	MISSES	TOTAL MAILS	CORRECTLY CLASSIFIED %
1	223	519	742	30.05%
2	451	209	660	68.33%
3	443	134	577	76.77%
4	403	92	495	81.41%
5	361	51	412	87.62%
6	295	35	330	89.39%
7	222	26	248	89.51%
8	149	16	165	90.30%
9	80	3	83	96.38%
10	82	1	83	98.80%

Table 6.0 shows the experimental results of our proposed neural network solution for email classification into meaningful words. We used the following formula to compute our results.

$$Email\ Class = \frac{New\ Neural\ Network\ Correctly\ Classified}{Total\ Number\ of\ Emails}$$

A 98% success was achieved with email categories and this neural network algorithm was compared with human participants, our algorithm’s performance worked very well and was even better than the existing techniques.

Many more testing and evaluations were implemented on thousands of email messages to determine the class they belonged to using Back Propagation algorithm (BPA). A critical analysis of the results in table 6.2 shows that the more email messages the user had the more difficulty it became to determine the correct classes as a percentage of accuracy decreased when more email messages were grouped in a bulky dataset.

In our quest to compare different techniques of email classifications, our email classifier was compared with a human classifier and the results are presented in Table 6.2 below. The results show that when BPA was applied to a small number of incoming unsolicited email messages into the inbox, it was able to achieve 98% accuracy on email classification as compared to human prediction system.

Table 6.2: Human Predictions versus BPA Classifications

Email Counter	**Human Prediction**	**BPA Prediction**	**Accurate Prediction, %**
1	1 000	986	98.6%
2	2 000	1 884	94.4%
3	3 000	2 790	93.0%
4	4 000	3 699	92.4%
5	5 000	4 600	92.0%
6	6 000	5 461	91.0%
7	7 000	6 358	90.8%
8	8 000	7 146	89.3%
9	9 000	8 005	88.9%
10	10 000	8 710	87.1%

The proposed technique’s accuracy began to deteriorate in performance when BPA was applied to a huge number of email messages of above 6 000 with time kept at the same level. However, with above 1000 email messages, we were able to achieve 87% classification accuracy with our BPA system.

VII. CONCLUSION

We were able to show that neural networks using Back Propagation approach can be successfully used for email classification into meaningful words. The Back Propagation is based on learning by example and outperforms many other previously reported algorithms in terms of email classification. Many more experiments will be done in future.

VIII. REFERENCES

[1] Ayodele, T., Zhou, S., & Khusaino, R. (2010). Email classification using back propagation technique. International Journal of Intelligent Computing Research (IJICR), 1(1/2), 3-9.

[2] Schuff, D. O., Turetke, D. & Croson, F. (2007). ‘Managing Email Overload: Solutions and Future Challenges’, IEEE Computer Society, 40(2), pp.31-36

[3] Awad, W. A. & Eseuofi, S. M. (2011). Machine Learning methods for Spam email Classification. International Journal of Computer Science and Information Technology, 3(1), pp 173-184.

[4] Yukun, C., Xiaofeng, L., Yunfeng, L. (2007). An Email Filtering Approach Using Neural Network, Springer, Berlin, pp. 688-694.

[5] Clark, J., Koprisnka, I. & Poon, J. (2003). A Neural Network Based Approach to Automated Email Classification. IEEE International Conference on Web Intelligence, pp. 702-705.

[6] Ramos, J. (2002). Using TF-IDF to Determine Word Relevance in Document Queries, Department of computer science, Rutgers University, Piscataway, NJ, 08855.

[7] Lai, GH., Chen, CM., Laih, CS. & Chan, T.(2009). A collaborative anti-spam system. Expert Systems with Applications. Elsevier, 36: 6645-6653.

[8] Druker, H., Wu, D. & Vanpik, V.N. (1999). Support Vector Machines for Spam Categorization. IEEE Transactions on Neural Networks. 10(5), pp. 1048-1054.

[9] Ferris Research (2009). Spam, Spammers and Spam control A White Paper.

[10] Ferris Research (2010). Industry Statistics.

[11] Guzella, T. S. & Caminhas, V. M. (2009). A review of machine learning approaches to Spam filtering.Expert system with applications, 36: 10206-10222

[12] Matsumoto, R., Zhang, D. & Lu, M. (2004). Some Empirical Results on Two Spam Detection Methods. IEEE, pp. 198-203.

[13] Lobato, D. H. & Lobato, J. M. (2008). Bayes Machines for binary classification. PatternsRecognition Letters. Elsevier, 29: 1466-1473.

[14] Artificial Neural Networks (2008). Neural Networks.[Online] Artificial Intelligence Technologies tutorial. Available on http://www.learningartificialneuralnetworks.com/#Intro. Last visited 10 December, 2014

[15] Habra, A. (2005). Neural Networks-An Introduction. [Online] Technology Exponent. Available at http://www.tek271.com/ Last visited 30 November, 2014.

[16] Rafiqual, I. & Morshed, U. C. (2005). Spam Filtering using Machine Learning Algorithms. IADIS International Conference on www/internet .ISBN: 972-8924-02, pp. 419-426

[17] Aery, M. A. (2005). eMailSift: Email Classification Based on Structure and Content. In Proceedings of the 5th IEEE International Conference on Data Mining(pp. 18-25). Washington DC:IEEE Computer Society.

[18] Sebastian, F. (2002). Machine learning in automated text categorization. ACM computing surveys. 34(1), pp. 1-47

[19] Thamarai, S., Hamid, A. J. & Alaa, Y. T. (2010). Overview of textual anti-spam filtering Techniques. International Journal of the Physical Science, 5(12), pp. 1869-1882.

[20] Soranamageswari, M. & Meena, C. (2011). A Novel Approach towards Image Spam Classification. International Journal of computer Theory and Engineering, 3(1), pp. 1793-8201

[21] Hamid, S. M. & Mohammed, N. J. (2013). A content based spam filtering using optical back propagation technique. International Journal of Application or Innovation in Engineering and management, 2(7), 416-421

[22] Wang, B., Jones, G. J. F. & Pan, W. (2006). Using online linear classifiers to filter spam Emails. Patterns Analysis and Applications, 9: 339-351.

[23] Wu, C. H. (2009). Behaviour-based Spam detection using a hybrid method of rule-based techniques and neural networks. Expert Systems with Applications. Elsevier, pp. 4321-4330

[24] Zhang, J. (2003). Modified Logistic Regression. An Approximation to SVM and Applications in large-scale text categorization. Proceedings of the 20th international conference on machine learning. AAAI press, pp. 888-895

[25] Wood, P., Bleaken, D., Nisbet, M., Zhang, J., Johnston, N., Lee, M. & Lewis, D. (2010). MessageLabs intelligence: (2009) Annual Security Report. Retrieved on 10 December, 2014 at http://www.messagelabs.co.uk/intelligence.aspx.