Scamming the Scammers: Towards Automatic Detection of Persuasion in Advance Fee Frauds

Matthew Edwards
m.edwards7@lancaster.ac.uk

Claudia Peersman
c.peersman2@lancaster.ac.uk

Awais Rashid
a.rashid@lancaster.ac.uk

Security Lancaster
School of Computing and Communications
Lancaster University
United Kingdom
LA1 4YW

ABSTRACT

Advance fee fraud is a significant component of online criminal activity. Fraudsters can often make off with significant sums, and victims will usually find themselves plagued by follow-up scams. Previous studies of how fraudsters persuade their victims have been limited to the initial solicitation emails sent to a broad population of email users. In this paper, we use the lens of scam-baiting – a vigilante activity whereby members of the public intentionally waste the time of fraudsters – to move beyond this first contact and examine the persuasive tactics employed by a fraudster once their victim has responded to a scam. We find linguistic patterns in scammer and baiter communications that suggest that the mode of persuasion used by scammers shifts over a conversation, and describe a corresponding stage model of scammer persuasion strategy. We design and evaluate a number of classifiers for identifying scam-baiting conversations amidst regular email, and for separating scammer from baiter messages based on their textual content, achieving high classification accuracy for both tasks. This forms a crucial basis for automated intervention, with a tool for identifying victims and a model for understanding how they are currently being exploited.

Keywords

Scam-baiting, advance fee fraud, mass-marketing fraud, persuasion, cybercrime

1. INTRODUCTION

Advance fee fraud typically involves promising its victims wealth, gifts, prizes or employment in exchange for a small advance payment. If a victim goes along with the story and pays the fee, fraudsters either completely disappear or invent a series of difficulties which require further payment from the victim until the victim is out of money or stops paying. Advance fee frauds have been afflicting people's lives for hundreds of years. In the 16th century, businessmen were contacted by an individual claiming to be in correspondence with a wealthy person imprisoned in Spain – usually alleged to be an unknown/remote relative of the victim who was relying on this individual to arrange his release. In exchange for a small amount of money required to bribe the prison guards, the victim would receive a reward once this person was smuggled out of prison. Another variant of the fraud dates back to the 19th century, mentioning a casket containing 16,000 francs in gold and the diamonds of a late marchioness and showing a remarkably high resemblance to the suspicious emails people receive today [9]:

> "Sir, you will be doubtlessly be astonished to be receiving a letter from a person unknown to you, who is about to ask a favour from you [...]"

The proliferation of the Internet and easy access to email harvesting software gave rise to more modern variants of the Spanish prisoner scam, such as black money scams, lottery frauds, employment scams, online sales and rental frauds, work-at-home scams, romance scams, etc. A few examples:

- The victim has won a large cash prize, a lottery jackpot or an inheritance and in order to claim it they must send a small payment as a release fee.
- An implausible series of events has occurred and the victim's help is required in getting money out of the scammer's country, which will normally involve making a series of payments.
- The victim is promised a dream job, but has to make payments for taxes, visas, "anti-terrorism certificates" or any number of other formalities.

These messages are delivered in bulk by spam-delivery networks, operated by other cybercriminals who lease out their (stolen) equipment to send email. For a majority of the audience, such emails will go unread, consigned to a spam folder, or else manually identified as spam and deleted or ignored[1]. However, for the respondents, this email will be the start of multiple conversations, the end result of which

[1]For more precise statistics on the ratio of spam solicitation to response and conversion, we refer to the reader to Kanich et al. [5]

WWW 2017 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3053889

will be the loss of money, if not their identities or even their lives[2].

In the first part of this paper, we present our approach to automatically distinguish advance fee fraud-related email exchanges from regular professional and personal email conversations included in the ENRON dataset. Such a system allows for the detection of an ongoing exchange that relates to a fraudulent arrangement, instead of merely identifying the original scammer solicitation email. Additionally, because rapid intervention is essential to safeguard the respondents from further losses, we go on to investigate the feasibility of automatically distinguishing between scammer and respondent messages. This task can be expected to be more challenging than separating the texts authored by scammers from general email communications due to the close similarity in topic and language usage. Because there are no actual scammer-victim exchanges available (yet), we examine *scam-baiting* conversations.

Scam-baiting is something of a vigilante activity, wherein a scam-baiter responds to a scam email which solicits an advance fee or similar fraud, and engages with the fraudster in order to waste their time and inconvenience them with thankless tasks, all the while acting the part of a duped victim. This is often done for the scam-baiter's own amusement, but justified by a maxim that any time and energy spent by a fraudster in dealing with scam-baiters distracts them from engaging with and defrauding actual victims. Most pertinently, scam-baiters commonly publish transcripts of their engagements for public appraisal, giving the public a rare, if atypical, insight into the stages of advance fee fraud which follow on from the more broadly recognised solicitation letters[3].

Secondly, in this study we attempt to understand the different stages of persuasion in online scammer conversations. While such persuasion strategies have been studied before (e.g. [1, 8, 4]), the analysis of scammer strategies is usually restricted to the initial solicitation email. Hence, the existing literature fails to properly analyse the persuasive strategies which emerge later in the conversation, where a respondent is talked into becoming a victim. Therefore, in the second part of this paper we contribute to a more general linguistic understanding of the strategies of persuasion in advance fee email frauds by analysing entire scammer/scam-baiter conversations.

The novel contributions of this paper are as follows:

1. The description of a novel dataset of cleaned and labelled scammer/scam-baiter exchanges, the ADVANCE FEE SCAM-BAITING dataset, sourced from active scam-baiter communities.

2. The development of a highly accurate classification system for distinguishing scam-baiting exchanges from normal email traffic (F1=0.971) and scammer from scam-baiter messages (ACC=0.963) based upon both the ADVANCE FEE SCAM-BAITING and the ENRON dataset, comparing a range of classification approaches and feature sets.

3. A statistical analysis of scammer and scam-baiter communications throughout exchanges, examining correlations of linguistic categories with message sequence, and patterns which develop when these categories are aggregated.

4. The generation of a prospective linguistic outline of the persuasive process employed by scammers, including the notable result that persuasion used later in the exchanges differs from that visible in solicitation emails.

Our classifier results demonstrate that scammer communications can be automatically identified, and the parties in these exchanges accurately labelled to enable automatic intervention. Further, our analysis of the conversational patterns of scammers and scam-baiters reveals that persuasive strategies employed can shift, leading us to pose a model of scammer persuasion. Our results validate the necessity of studying full transcripts to understand persuasion.

The paper proceeds as follows: in Section 2 we present background research on scam-baiting and advance fee fraud. In Section 3 we describe the scam-baiting dataset underpinning our analyses. In Section 4 we present results of an experiment classifying scammer and scam-baiter messages based on their textual content, and in Section 5 we discuss the presentation of various linguistic features across the observed conversations, drawing out notable correlations and patterns. In Section 6 we outline a resulting understanding of scammer persuasive strategy, and discuss the implications and directions for future work.

2. RELATED WORK

Previous work on scam-baiting data has mainly focused on qualitative assessment and categorisation of scam-baiter activities [11]. Regarding advance fee fraud specifically, prior studies often focus their analysis on the components of the initial solicitation messages. Chang [1] provides one such analysis, tracing aspects of Cialdini's influence theory [2] in six examples of advance-fee fraud solicitation emails, finding presentations and assertions of authority, as well as a tendency towards urgency to pressure victims into taking decision-making short cuts.

Jakobsson [4] comes to similar conclusions in his investigation of the solicitation emails for a variety of scams. He bases his analysis on the principles laid out by Ferreira et al [3] as underlying phishing. Namely, these are the principles of *Authority*, *Social Proof*, *Liking*, *Similarity*, *Deception*, *Commitment*, *Reciprocation*, *Consistency* and *Distraction*. His conclusions are that Commitment, Reciprocation and Consistency combined with Distraction and Deception are the most frequently applied across email scam solicitation, with Authority playing a secondary role. A typical example of the application of these principles is the lottery scam, in which the victim feels obliged to respond to a request, such as paying a fee to obtain the prize, by focusing on what can be gained, instead of the fraudulent nature of the request.

While informative, these analyses do not go further than understanding the principles of persuasion that are used in

[2]Aside from losses due to suicide, victims who travel out to meet scammers are in some cases murdered. A US congressional report from 1998 estimated 17 such deaths from known cases [*Combating International African Crime: Hearing Before Subcomm. on Africa of House Comm. on Int'l Rel. (July 15, 1998)* `http://commdocs.house.gov/committees/intlrel/hfa50884.000/hfa50884_0.htm`].

[3]A more detailed description of the scam-baiting community is provided by Zingerle [11]

the initial solicitation email. A linguistic model that incorporates the different stages in the scammer's persuasive strategies throughout a series of email exchanges is – to our knowledge – lacking. With the analyses provided in this study, we aim to contribute to forming the first step towards such an understanding.

3. DATA SOURCES

3.1 The Advance Fee Scam-baiting Dataset

Our ADVANCE FEE SCAM-BAITING dataset is collected from public transcripts posted by members of the "419eater" scam-baiting community[4] in the 419eater archives and members' forum, supplemented by transcripts posted at the site "What's the Bloody Point?"[5]. The corpus currently consists of 57 complete exchanges, numbering 2,248 messages.

Each email in the dataset is annotated with the role of the author (*scammer* or *scam-baiter*). The ratio of messages is slightly in favour of scammers, at 1162 : 1086. Although the majority of transcripts begin with the initial solicitation message from the scammer, 5 open instead with a message from the baiter, after an explanation of the context. The corpus covers a spread of dates from 2003 to 2015, with an average of 38 messages per exchange. We show a sample of such an exchange in Appendix A.

3.2 The Enron Dataset

The ENRON Dataset contains around 0.5M professional and personal email messages from about 150 different people. Most of these people were part of the senior management of Enron at the time of the dataset's collection as part of a corruption investigation. A more complete description of the Enron corpus is provided by Klimt & Yang [6]. The dataset is freely available online[6]. To simulate a real-life data distribution in our experiments, we randomly selected 1,000 email conversations from the ENRON Dataset, with an average of 40 messages per exchange.

4. AUTOMATIC DETECTION OF ADVANCE FEE EMAIL COMMUNICATIONS

In this section we discuss our methodology to automatically distinguish between fraudulent advance fee emails and regular professional and personal email conversations. Our approach to this task is based on text categorisation and involves the creation of document representations based on a selected set of linguistic features, feature selection using statistical techniques, and classification using machine learning algorithms. We describe each of these steps in the following subsections.

4.1 Data Pre-processing

Postings on scam-baiting websites can contain annotations or comments from the scam-baiters which are inserted to explain events or to comment upon the strategies of the scammers. Because this text was not part of the original email conversations, the first step in pre-processing consisted of removing it from the transcripts. This was accomplished through a semi-automated process, whereby comments inserted according to a conventional format were extracted automatically and further unconventional comments were manually removed. Likewise, any direct references to Enron were removed from the ENRON Dataset.

Next, we tokenised all messages. Because of the non-standard language use that is often present in the data, the tokenization process consisted of splitting up each message in a list of words and punctuation marks, but concatenated forms (e.g. "somepeopleâĂİ) and incorrectly spelled words (e.g. "whelther") were left unchanged, because such "linguistic noise" could be informative to distinguish between scammer and non-scammer emails during the experiments. We stripped punctuation, standardised all tokens to lower-case, replaced numeric sequences with a number token, and replaced sequences of 'X'-marks with a token indicating omitted information. Finally, an implementation of the Porter stemming algorithm was used to reduce words to canonical forms.

4.2 Feature Selection

In text categorisation, a document representation (in our case a series of consecutive emails written by a scammer, a scam-baiter or an Enron employee) is composed of different types of linguistic features, the selection of which can significantly affect the performance of the machine learning algorithm. In this case, four different feature sets were explored:

- **BOW features**: the set of all words occurring in the data-set (Bag-Of-Words features).
- **Frequent Term features**: the set of terms occurring more than a cut-off frequency (40 times) in the document corpus, excluding common English stop-words and any obvious personal or corporate names. Term frequencies were recorded for each message.
- **Semantic features**: manually-created categories from the list of frequent terms, such as terms relating to finance (e.g. "money", "cash", "amount") or legalities ("lawyer", "form", "contract"). These categories were partially based on prior work by [4]. Additionally, after manually analyzing part of the scam-baiting conversations, we decided to add a number of extra categories that seemed typical of online scammers' persuasive strategies. Category frequencies were recorded for each message. All categories, together with a few examples for each category are displayed in Appendix B.
- **LIWC features**: previously-existing dictionary terms mapped to categories, from the LIWC 2015 dictionary [7]. Category frequencies were recorded for each message.

All features were drawn from the subject line and main text of each message.

4.3 Experiments and Results

Using our labelled datasets of scammer/scam-baiter/regular exchanges, we first trained and evaluated a variety of classifiers for the task of distinguishing between regular email conversations (NON-SCAM messages drawn from the ENRON dataset) and scam-baiting exchanges (SCAM_EITHER from the ADVANCE FEE SCAM-BAITING dataset). Additionally,

[4] http://www.419eater.com
[5] http://www.whatsthebloodypoint.com/
[6] https://www.cs.cmu.edu/~./enron/ (last accessed on 26/01/17).

Classifier	Scam_Either				Non-scam		
	Acc.	Prec.	Rec.	F-sc.	Prec.	Rec.	F-sc.
NB_BOW	98.3	85.5	100.0	92.2	100.0	98.1	99.0
NB_LIWC	93.3	73.4	51.8	60.7	94.8	97.9	96.3
NB_Semantic	92.5	82.2	33.0	47.1	93.0	99.2	96.0
NB_FreqTerms	94.0	63.5	94.6	76.0	99.4	93.9	96.6
LR_BOW	99.4	100.0	93.5	96.7	99.3	100.0	99.7
LR_LIWC	99.4	100.0	93.5	96.7	99.3	100.0	99.7
LR_Semantic	98.8	99.0	88.9	93.7	98.8	99.9	99.4
LR_FreqTerms	99.4	100.0	93.5	96.7	99.3	100.0	99.7
SVM_BOW	99.4	100.0	93.5	96.7	99.3	100.0	99.7
SVM_LIWC	99.4	100.0	93.5	96.7	99.3	100.0	99.7
SVM_Semantic	99.3	100.0	92.6	96.2	99.2	100.0	99.6
SVM_FreqTerms	99.5	100.0	94.4	**97.1**	99.4	100.0	99.7

Table 1: Overall accuracy, precision and recall for distinguishing between the Enron and the Advance Fee Scam-baiting dataset.

Classifier	Scammer				Scam-baiter		
	Acc.	Prec.	Rec.	F-sc.	Prec.	Rec.	F-sc.
NB_BOW	95.5	94.7	96.4	95.6	96.4	94.6	95.5
NB_LIWC	68.8	66.7	75.0	70.6	71.4	62.5	66.7
NB_Semantic	70.5	73.5	64.3	68.7	68.2	76.8	72.3
NB_FreqTerms	92.7	90.0	96.4	93.1	96.2	89.3	92.6
LR_BOW	95.4	95.0	96.6	95.8	95.8	93.9	94.9
LR_LIWC	81.5	85.5	79.7	82.5	77.4	83.7	80.4
LR_Semantic	82.4	83.3	84.5	84.0	81.3	79.6	80.4
LR_FreqTerms	93.5	91.9	96.6	94.2	95.7	89.8	92.6
SVM_BOW	**96.3**	95.1	98.3	96.7	97.9	93.9	95.8
SVM_LIWC	85.2	87.7	84.8	86.2	82.4	85.7	84.0
SVM_Semantic	85.2	86.4	86.4	86.4	83.7	83.7	83.7
SVM_FreqTerms	94.4	92.1	98.3	95.1	97.8	89.8	93.6

Table 2: Overall accuracy, precision and recall for distinguishing between the scammers and scam-baiters in the Advance Fee Scam-baiting dataset.

we set up experiments in which we attempted to separate messages authored by scammers (Scammer) from those authored by scam-baiters (Scam-baiter) within the Advance Fee Scam-baiting dataset.

Our system was developed using ten-fold cross validation (cf. [10]). In this experimental regime, the available data is randomised and divided into ten equally sized folds or partitions. Subsequently, each partition is used nine times in training and once in test. During the splitting, the email messages were clustered by conversation, to ensure that no message used in validation was part of a conversation with other messages visible in training, which prevented overfitting of features which may be particular to a given scam or pair of conversation partners.

The feature sets we described in Section 4.2 were trialled in three different classifiers, in evaluations carried out over our two datasets. The classifiers used were support vector machines with a linear kernel (SVM), naive Bayesian classifiers (NB) and binomial logistic regression (LR). Parameters were experimentally determined on a development set of each training partition during cross validation.

Table 1 and 2 present the overall accuracy of each classifier and show the precision, recall and F-score for each category in both experimental set-ups. During the first series of experiments (Non-scam vs. Scam_Either) the best result was achieved by the SVM classifier using Frequent Term features. The scam-baiting email conversations were successfully identified with an F-score of 97.1%. Of the individual feature sets, the frequent terms were generally the best performing, and the Semantic the poorest performing. When distinguishing between scammers and scam-baiters, the BOW features yielded the best results. When detecting the scammers of our Advance Fee Scam-baiting dataset the SVM yielded a best overall accuracy score of 96.3%. Combining different feature types did not produce significantly better results.

5. PERSUASIVE MESSAGING IN ADVANCE FEE SCAMS

Although scammers are constantly changing the content of their emails in order to prevent detection by scam filter software, the use of persuasion remains a critical marker for fraudsters attempting to trick or manipulate people into giving up money. In this section, we discuss our analysis of semantic markers within the scammer and scam-baiter conversations, scrutinising such persuasive content as it occurs in an extended discussion, rather than only the initial solicitation email.

5.1 Conversation Level Analysis

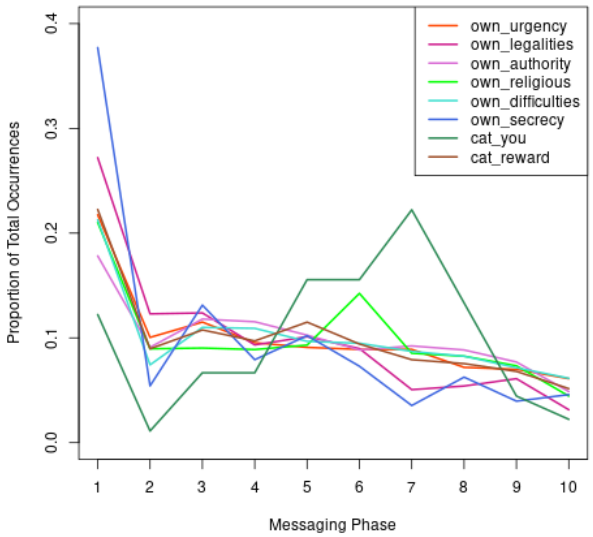
Conversations were normalised by binning messages according to which decile of the overall transcript length they fell into. Feature weights were then calculated for each bin according to the proportion of that feature's overall presentation which appears in each location. For interpretation purposes, a uniform distribution (the null model) would present as a proportion of 0.1 in each decile.

As Figures 1a and 1b show, a scammer's initial solicitation email and its immediate follow-up are packed with details of the scam set-up, selling why and how the victim should engage with the con. Traces of some typical persuasive strategies are displayed, such as the assumption of authority and language usage stressing the urgency of a transaction. Also visible is a concentration of attention on secrecy and security of communication, all within the first few exchanges.

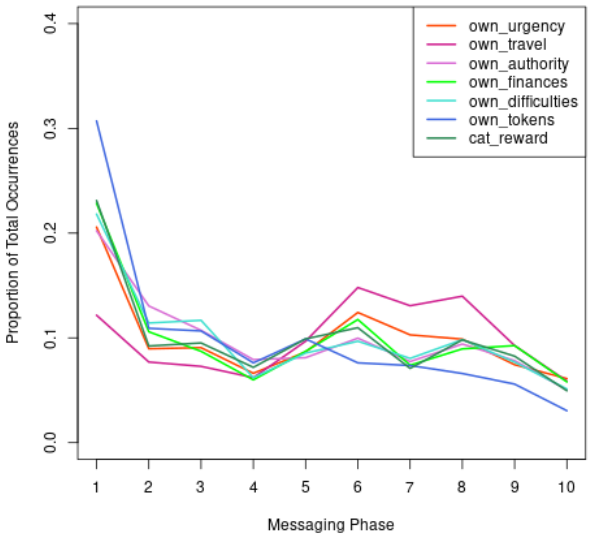
As the conversation continues, however, these elements become secondary to the focus on the second-person. The transition suggests that the scammer is moving away from strategies based on combinations of principles of authority and distraction, and towards strategies based on liking, similarity and reciprocity [4].

For the scam-baiters, a very similar picture appears, with their initial messages more densely packed with counterploys that they hope to hook the scammer's attention with, including the presentation of difficulties complying with instruction, and demands for tokens of good faith. These tactics are either enabled or abandoned relatively soon after. Notably low in the initial exchange are references to travel, a gambit which appears to be deployed only later on in the proceedings.

Looking at traces of emotional terms in Figures 2a and 2b, an interesting pattern emerges. The scammers' use of swearwords seems to peak around the fourth decile of the exchange, possibly reflecting an early loss of patience, or an

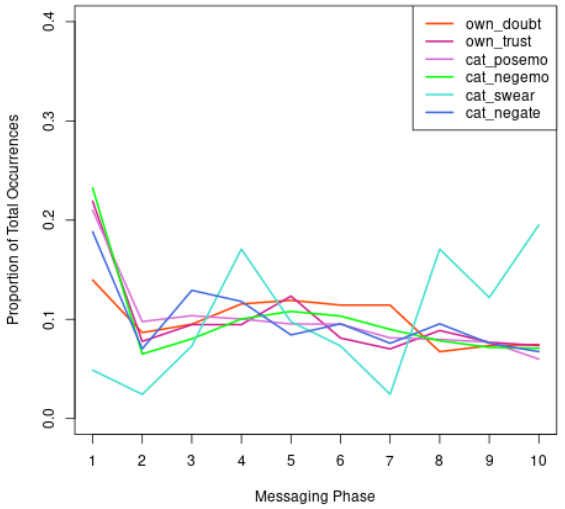


(a) Scammers' use of tactical terms

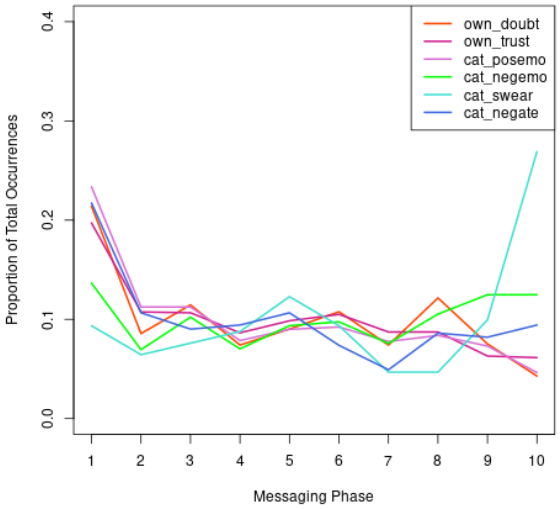


(b) Baiters' use of tactical terms

Figure 1: Term categories related to scammer and scam-baiter strategies, over email transcripts



(a) Scammers' use of emotionally-loaded terms



(b) Baiters' use of emotionally-loaded terms

Figure 2: Emotion and sincerity over email transcripts

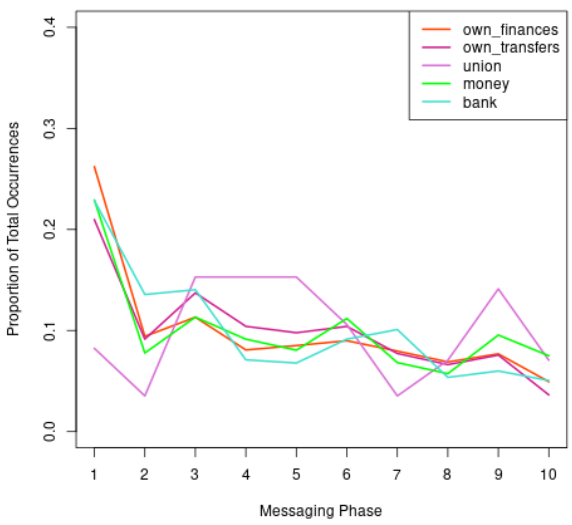


Figure 3: Scammers' use of financially-connected terms

attempt at intimidation. Around the same time, language usage reflecting termination also increases, indicating the scammer is threatening to call off the arrangement. This is followed by a spike in words related to trustworthiness and sincerity and a drop in swearword usage, which then returns dramatically towards the end of the conversation. This corresponds roughly with an increase in references to urgency and travel on the part of the scam-baiter. In combination, these tracks suggest a common pattern whereby the scammer loses patience midway through the conversation – perhaps because the scam-baiter has yet to deliver on their end of what is required for the con – and is pacified by the scam-baiter with a recognition of the urgency of the transaction, and a suggestion of a new means of delivery (which is in fact just another gambit to waste the scammer's time and resources).

Finally, with regard to financial references in particular, which include the scammers' ultimate goal of persuading their victims to transfer money, Figure 3 shows that scammers make less mention of money transfer services (e.g., Western Union) in the first part of their conversations. Instead, they raise the matter further on into the process. This is typically followed by the peak in verbal aggression we mention above. The distribution of financial terms declines slightly as the conversation moves into the liking and similarity phase and then rises again towards the end of the conversation, in a final effort to obtain the money they are seeking.

5.2 Persuasion Stage Model

Based on the results of our analysis, an outline of the persuasive strategies that are employed by advance fee scammers is given in Figure 4, as described below:

1. SOLICITATION. In the solicitation email, the scammer's most visible communication, a widely-distributed email typically outlines the plot of the scam. This is a hard sell, and scammers tend to use an entire range

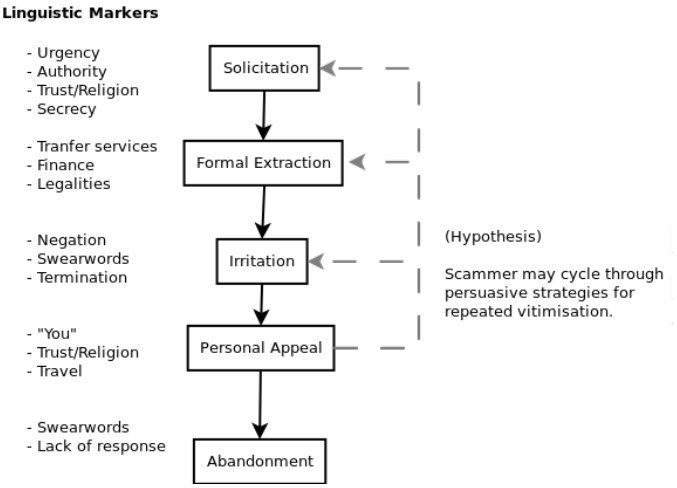


Figure 4: Scammer persuasive process

of persuasive aids to bolster their case. To start with, they create a false persona. With this strategy, scammers attempt to persuade their victims either by pretending to be a person of **authority** – people tend to respond to requests from a figure of authority (e.g. a bank director) – or by creating a likeable persona to whom the victim can relate to or sympathise with (e.g. a poor orphan). In the latter plot, scammers tend to present themselves as being **trustworthy**. As was mentioned by [4], people feel more confident in their decisions if they like the person they are following or if the person seems familiar or similar to themselves. In both cases, the fraudster attempts to distract the victim with promises of a **reward** (a large amount of money or a dream job) and the **urgency** of the victim's reply. Additionally, stress is made of the **secrecy** required. If not, the reward will somehow be lost.

2. FORMAL EXTRACTION. The recipient having responded, the scammers reply in line with their ploy, providing information for means of further communication and the method by which an advance fee should be paid or account information transferred. The recipient is typically asked to transfer the money through a **money transfer service** in which the recipient can remain anonymous, the transaction cannot be reversed, the money cannot be traced afterwards and the recipient can collect it at one of many locations (e.g. Western Union or MoneyGram). The focus here is funnelling the recipient into a procedure – references to **legalities** and **financial words** are numerous.

3. IRRITATION. If the recipient proves reluctant or presents difficulties, the scammer responds first of all with **negation** and reminders of **secrecy** and **urgency**, and then with a spike of verbal aggression (**swearwords**) and a threat to end their association (**termination**). This increases the pressure on the recipient to submit to demands.

4. PERSONAL APPEAL. If appeased, the scammer shifts from their previous strategy of guiding the recipient through a procedure, and shifts to more of a personal

focus on the recipient (**you**), with a renewed effort to present themselves as likeable and trustworthy (**religious**). They may agree for or even encourage the recipient to **travel** in person to deliver funds.

5. ABANDONMENT. If the conversation drags on too long, or becomes too tiresome, the scammer will grow irritated, verbal aggression will return, and they often stop replying and move on to other targets.

6. CONCLUSIONS AND DISCUSSION

In this paper, we showed that it is feasible to design a system that can automatically separate advance fee email conversations from regular professional and personal email exchanges while maintaining the complex conditions of a real life scenario – a large, highly skewed dataset. Additionally, we demonstrated that the system can identify both the fraudster and the potential victim on the conversational level with high accuracy, despite the similarities we attested in the ADVANCE FEE SCAM-BAITING dataset.

Additionally, the analysis we presented in the second part of this study revealed a number of hints about how scammer-victim conversations may progress. Most importantly, there is a strong indication that the scammer solicitation emails studied by previous literature are not representative of the rest of their interaction with victims, a result which stresses the importance of studying full exchanges. In our analysis, we identified a shift in language use on the part of scammers as transcripts progress, and postulate corresponding stages in scammer strategy. Further work will explore whether these stages can be manually labelled with consistency, and if so, then whether these stages can be identified by machine learning processes.

Our dataset of scam-baiting transcripts by no means exhausts the available data, which exists in a variety of formats. Additional data collection may lend itself to new insights about scam-baiting conversations and scammer strategy.

However, it must be acknowledged that the conclusions drawn from scam-baiting exchanges must be viewed cautiously for application to scammer-victim interactions. Scam-baiters are not victims, and in many ways their contributions to exchanges may more closely resemble those of scammers. The classifier we have developed in this paper demonstrates that this issue is not insurmountable, but the influence that they exert on the conversation remains to be examined.

For example, it is widely known that the fees required by advance fee fraudsters are rarely singular in nature – a single payment extracted will be followed by demands for yet more payments. As such, it is likely that stages 2, 3 and 4 in our process described above are cyclical in nature, with scammers entering the process via the solicitation and rotating through the tactics until the victim ceases to present a good prospect, and then abandoning them. As scam-baiters never surrender money, it is not possible to properly observe these cycles in scam-baiting transcripts. As a result, our future research will focus on establishing whether this is evident in actual scammer-victim transcripts, and generally examine the comparability of scam-baiter and scammer-victim exchanges.

7. ACKNOWLEDGEMENTS

The research reported in this paper is supported by award `EP/N028112/1` "DAPM: Detecting and Preventing Mass-Marketing Fraud (MMF)", from the UK Engineering and Physical Sciences Research Council.

8. REFERENCES

[1] J. J. Chang. An analysis of advance fee fraud on the internet. *Journal of Financial Crime*, 15(1):71–81, 2008.

[2] R. B. Cialdini and N. Garde. *Influence: the psychology of persuasion*, volume 3. A. Michel, 1987.

[3] A. Ferreira, L. Coventry, and G. Lenzini. Principles of persuasion in social engineering and their use in phishing. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, pages 36–47. Springer, 2015.

[4] M. Jakobsson. *Understanding social engineering based scams*. Springer, 2016.

[5] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamalytics: An empirical analysis of spam marketing conversion. In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 3–14. ACM, 2008.

[6] B. Klimt and Y. Yang. Introducing the enron corpus. In *CEAS*, 2004.

[7] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn. The development and psychometric properties of liwc2015. Technical report, 2015.

[8] D. Schaffer. The language of scam spams: linguistic features of "nigerian fraud" e-mails. *et Cetera*, 69(2):157, 2012.

[9] C. Seife. *Virtual Unreality: Just Because the Internet Told You, how Do You Know It's True?* Penguin, 2014.

[10] S. M. Weiss, N. Indurkhya, T. Zhang, and F. Damerau. *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media, 2010.

[11] A. Zingerle. Towards a categorization of scambaiting strategies against online advance fee fraud. *International Journal of Art, Culture and Design Technologies (IJACDT)*, 4(2):39–50, 2014.

APPENDIX

A. TEXT SAMPLE OF THE ADVANCE FEE SCAM-BAITING DATASET

Scam-baiter	Mssr Kouame, I was directed to contact you by Barrister Andy Coulibally. He had quoted fees that were to be charged for supplying documents, as below: Change of Ownership;- $950.00. Affidavit of Oath;- $1,025.00. Letter of Authorization; $900.00. Total. $2,875.00. IMO these fees are outrageous. I would like you to explain why they are so high. Regards, Eliza Dane
Scammer	Attention Eliza Dane, this is prior to your mail to this court regarding the fee to obtain some legal documents as requested by barrister Andy Coulibally,I want to tell you that the fee as was issued the barrister is for both processing legalisation at the justice ministry here. But if you want the processing and legalisation fee to be deducted from the fee then the total fee to get the mentioned documents is $1,950 only,the processing and legalisation fee is $925. you are advice therefore to liase with your local representative here barrister Andy Coulibally and remmit the fee to him to enable him come and pay the fee for collection of the 3 documents respectively. But if you wish,you can also pay in the money here directly through the informations as was given to you by the barrister. Thanks for your understanding. Messr Richard Kouame
Scam-baiter	Mr Kouame, Please answer the question. Eliza Dane
Scammer	Ms Dane, What answer do you require again?.I have told you in my last mail that the cost as was highlighted to you by Barrister Andy Coulibally is for both the processing and legalisation of the documents at the ministry of justice.The normal cost to get the three documents is $1,950 while $925 is for processing and legalisation at the jusice ministry.The above price cost is the normal governmental price to purchase such documents and oaths in this country. We have no room for any curiousity solution here,For more details you can liase with your representative to get more facts. Messr Richard Kouame
Scam-baiter	Mr Kouame, To put it bluntly, I don't believe you. I did a bit of research and found that what you're charging is equal to something like two years' income for the average citizen of Cote d'Ivoire. How you can write "The normal cost to get the three documents is $1,950 while $925 is for processing and legalisation at the jusice ministry... cost is the normal governmental price to purchase such documents and oaths in this country" and keep a straight face is a mystery. I want the real price, not the one you quote when you try to gouge foreigners. Eliza Dane
Scammer	Ms Dane, We have no time for all this cross questioning,you have been told that this fee as highlighted is the minimum fee imposed by the government to get such ducuments in this country.If you need redress of this fee then it's you who will state how much you will be able to pay.If your bargain is acceptebale by the judicial commitee,then we will proceed but if it's not accepted,that means quit. Indicate what you will be able to pay and stop embibing curiuosity. Messr Richard Kouame
Scam-baiter	Mr Kouame, Fine – if you insist on charging this much, send me an invoice for the fee. Eliza Dane

B. SEMANTIC DICTIONARY

Category	Examples
numbers	“one”,“two”,“three”
locations	“nigeria”,“africa”,“ghana”
times	“monday”,“tuesday”,“wednesday”
urgency	“urgent”,“asap”,“immedi”
titles	“mrs”,“mr”,“dr”
address	“attn”,“dear”,“hello”
organisations	“ministri”,“foundat”,“board”
communicationmediums	“phone”,“telephon”,“email”
communicationforms	“promis”,“clear”,“say”
postal	“post”,“packag”,“mail”
insults	“stink”,“turd”,“dumb”
finances	“money”,“advanc”,“amount”
legalities	“form”,“receipt”,“offici”
travel	“travel”,“meet”,“flight”
religious	“bless”,“church”,“lord”
relationships	“bother”,“sister”,“father”
secondperson	“ye”,“you”,“your”
firstsingulars	“i”,“me”,“mine”
firstplurals	“we”,“us”,“togeth”
business	“secretari”,“job”,“presid”
assistance	“assist”,“help”,“develop”
emotion_bad	“concern”,“sorri”,“afraid”
emotion_good	“delight”,“like”,“interest”
desire	“hope”,“wish”,“interest”
termination	“end”,“death”,“close”
beginning	“start”,“make”,“pose”
alteration	“chang”,“els”,“therefor”
trust	“assur”,“ensur”,“believ”
doubt	“question”,“actual”,“wast”
authority	“mp”,“repres”,“director”
deal	“process”,“ventur”,“plan”
secrecy	“code”,“privat”,“password”
difficulties	“busi”,“far”,“lost”
exaggerated	“entir”,“extrem”,“real”
small	“small”,“little”
retention	“keep”,“remain”,“rememb”
reflection	“rememb”,“know”,“thought”
assent	“yes”,“okay”,“ok”
persons	“man”,“person”,“peopl”
searching	“found”,“request”,“find”
decision	“choos”,“pick”,“put”
tokens	“sampl”,“prove”,“someth”
objects	“meat”,“bread”,“ship”
foreign	“foreign”,“state”,“white”
life	“life”,“live”,“old”
common	“usual”,“real”,“alway”
uses	“put”,“use”,“fill”
transfers	“western”,“union”,“moneygram”