

Active Countermeasures for Email Fraud

Wentao Chen, Fuzhou Wang, *Matthew Edwards*

2023-07-04

Bristol Cyber Security Group

Great Investment Opportunity

My name is Rolf Michielson. I am a financial consultant/attorney. My client has a substantial amount of money \$10.5M to invest in any good business that will bring good return on investment either in your country or anywhere in the world. My client is looking for a reliable, trustworthy and competent person to partner with. If you are interested, I will like you to email me the following:

- 1) BUSINESS PLAN
- 2) ANNUAL RETURN ON INVESTMENT OF THE PROPOSAL
- 3) BUSINESS DURATION
- 4) YOUR DIRECT TELEPHONE NUMBER

My client is willing to negotiate the terms and conditions to suit your best interest. [...]

Email Fraud

Crime Type ¹	Victims	Losses (\$)
Advanced Fee	11,264	104,325,444
Lottery/Inheritance	5,650	83,602,376
Extortion	39,416	54,335,128
Investment	30,529	3,311,742,206
Confidence/Romance	19,021	735,882,192
	105,880	4,289,887,346

¹IC3 2022 figures, see <https://www.ic3.gov/>

Canham & Tuthill (2022) ²argue:

The current state of information security can therefore be simplified as follows; high numbers of cybercrimes are committed, extremely low arrest and conviction rates present little to no deterrence to threat actors, and humans continue to be the attack vector of choice.

[...]

employing social engineering active defense (SEAD) will be more effective to countering malicious actors than maintaining the traditional passive defensive strategy

²Canham, M. and Tuthill, J. *Planting a poison SEAD: Using social engineering active defense (SEAD) to counter cybercriminals*. In International Conference on Human-Computer Interaction, pages 48–57. Springer, 2022

Scam-Baiting

- A manual example of active countermeasures.
- Various manifestations
 - reverse-scamming/cash-baiting
 - ‘tropy-hunting’
 - ‘straight’ scam-baiting.
- Increases the false positive rate for scammers³.
- Frustrates scammers⁴

³Herley, C. *Why do Nigerian scammers say they are from Nigeria?*. In Proceedings of the Workshop on the Economics of Information Security, 2012

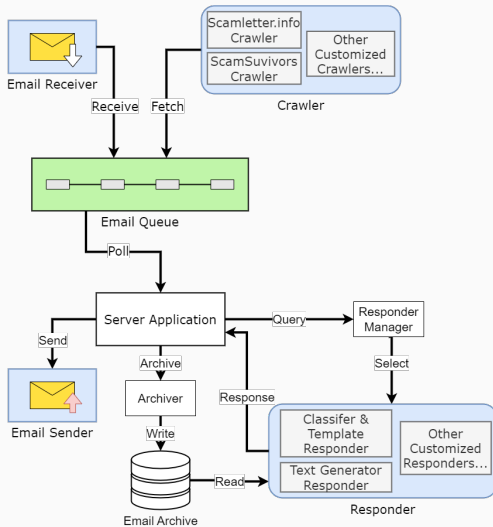
⁴Edwards, M., Peersman, C., and Rashid, A. *Scamming the scammers: Towards automatic detection of persuasion in advance fee frauds*. In WWW Companion. pages 1291–1299, 2017.

Also:

- Volunteer/hobbyist activity.
- Takes up scam-baiter time & energy.
- Doesn't scale.
- Not a good evidence-base.

What we want is an *automatic* scam-baiter.

Server Structure

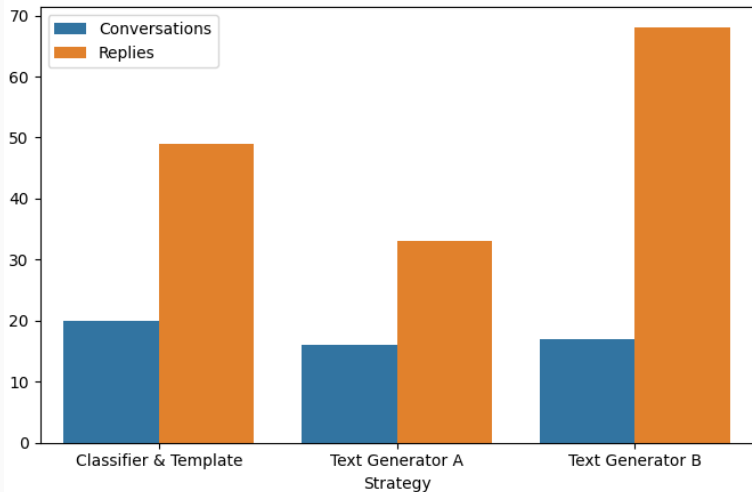


Classifier & Template Identify fraud format; respond with randomly-selected prewritten template option.

Text Generator A GPT-Neo trained on human scam-baiting exchanges in `<email,response>` pairs.

Text Generator B GPT-Neo trained on mix of scam-baiting & normal email conversations.

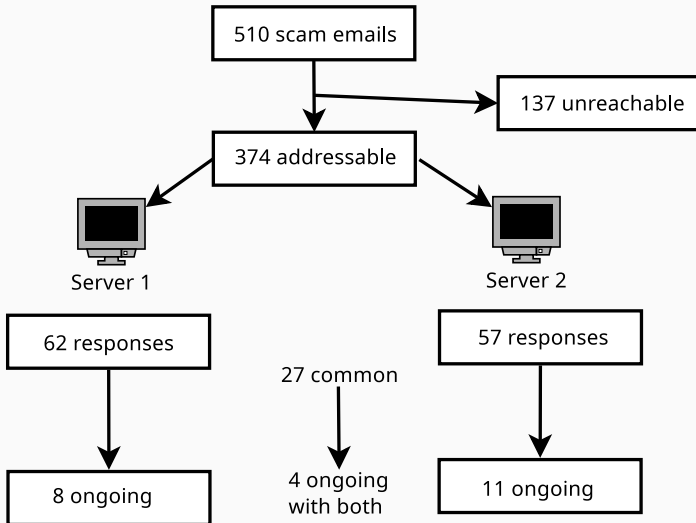
Results



Results

	Hook	Replies		Max. Time
		Avg.	Max.	
Classifier & Template	7%	2.45	5	17.2 days
Text Generator A	5%	2.06	5	17.3 days
Text Generator B	6%	4	23	21.5 days

Concurrent Engagement



Limitations

Scammer

The money is \$2,500

Scammer

The shipment is on now

Text Generator B

Dear David, doing what I'm supposed to do, going to bring the US to Israel, maybe take some ham and gruff out of the plane if maybe it's some hardwood muslin. But sure you will give it lot of thought. Here are my For short talk, but in the end you and are really going to have ham and wubbbut. chat.ze link. Best, Michael

Limitations

Text Generator B

Welcome,uffs, goodFoo, yo boxes Best, Michael

Scammer

Pay the money for the shipment

Scammer

??

??

??

How does ChatGPT do?

Very well^a. 28-day experiment testing 2 different ChatGPT prompting approaches (+ 1 control), found:

- Much improved (32-34%) scammer response rate.
- ChatGPT established longer conversations, some lasting the entire study period.
- Over 1,000 replies from 250+ scammers.
- New conversation-stoppers.

^a(Bajaj & Edwards, in preparation).

Sample

Thank you for your email regarding the release of my funds approved by the IMF. I appreciate your prompt response and would like to know more about the processing and transfer cost of US Nineteen Thousand Five Hundred and Fifty United States Dollars Only. Can you please provide me with a breakdown of the cost and the reason for such a high amount? Also, I would like to know more about the official Bitcoin payment method that you mentioned.[...]

Automatic Scam-Baiting as an Experimental Toolkit

1-month experiment⁵ assigning scammers to conversations with one of 4 'personalities' (+control). Total engagement: 1.5k responses, 294 scammers.

Alex is an optimistic, informal young man who sends bubbly messages with smiley faces and poor spelling.

Dan is a rude, impulsive middle-aged man who makes greedy and selfish demands.

Sam is a professional businessman who displays his credentials and always remains polite.

Doris is a kind, trusting old woman, who recently lost her husband and apologises for being slow.

*Which personality sustained **significantly** more engagement than the control?*

⁵(Robinson & Edwards, in preparation)

`https://github.com/scambaitermailbox/`

- A corpus of historic human scam-baiter interactions (658 conversations, \approx 37k messages).
- Code for running scam-baiting experiments.
- A growing corpus of automatic scam-baiting conversations. (723 conversations, \approx 5k messages).

Future Directions

1. Application to media other than email.
2. Longer-term deployments.
3. Testing further hypotheses about scammer interests.
4. Designing automatic scam-baiters to gather information.
5. Methods robust to detection techniques.

Thank you

Questions?