Internet research 2017/18
Internet Research final report
Application IR170017 From Dr Yun Sing Koh
Form Submitted 8 Aug 2019, 7:01pm NZST

Project Report

* indicates a required field

Research Title
Investigating Changes in Phishing Models for Social Networks
This question is read only.

Please provide a short summary of the work that was completed as part of this project / research *
The main aim of this research is to develop a robust spam phishing detection system to investigate how the New Zealand community is affected by spam phishing in social networks. The project is divided into four phases.

We have completed two of the four phases. The main findings are we developed a new unsupervised machine learning algorithm that can detect potential phishing attacks in online social networks. This technique can be used for early detection of potential phishing attacks and does not require pre-existing assumptions about the type of data or understanding of the characteristics of the potential attack. The current accuracy is 87% for the system. We are currently investigating and designing the technique to automatically detect unusual behaviour or changes in online social networks.
Describe the 'who, what, where, when and why' of your initiative

Timing

Is your project / research complete? *
◉ Yes ○ No
If your initiative is still in progress, pick 'no'

Start Date	**Finish Date**
01/04/2018	30/06/2019
Must be a date.	Must be a date.

Milestones

What have been the major steps / stages (i.e. milestones) involved in delivering your initiative to date?

Milestone	**Description**

Phase 1: Initial dataset collection and processing from the Twitter API completed.	We have built a Real-time Twitter Engineering Framework that leverages AWS to collect tweets and label the tweets at scale. Over 9 months we have collected 595,843 tweets, 160,634 are spam and 462 are phishing. Also, we have presented the features that we will use for our phishing detection technique and the cost of running in a real-time environment, such as not being able to capture potentially important features, such as re-tweets and favourite counts.
Phase 2: Development of unsupervised learning technique for phishing detection completed	Initially, we had the goal of developing an unsupervised technique so the technique could adapt to changes in phishing attacks, however, this was not achievable given the constraints. Instead, we achieved the same outcome with a semi-supervised technique. This technique was developed to be able to adapt to changes in phishing attacks, real-time processing presented on a real-world data set. We have presented a new real-time semi-supervised phishing detection algorithm applied to a real-world scenario. We have shown that Pelican performs better than the benchmark techniques, particularly when the evaluation is not in a sandbox environment where class imbalances exist. We showed that Pelican can capture more phishing tweets compared to benchmark techniques despite loss in non real-time features.
Phase 3: Development of the phishing drift detection technique to monitor the changes of the phishing models.	The technique uses a change detector that enables automatic retraining when there is an unusual behaviour detected. This enabled the technique to different phishing models as they change. We have shown that Pelican performs better with a drift detection technique compared to without. Also, the methodology behind the phishing detection uses different techniques to look at the structured and unstructured data that allows different phishing attacks to be detected.

Phase 4: Compare and contrast the types of phishing attacks on community world-wide v ersus the phishing attacks on communities i n New Zealand.	We have applied our phishing detection technique to a real-world scenario and applied transfer learning techniques to enable us to detect phishing in small population countries such as NZ. We have also shown that inductive transfer learning performs better than the benchmark technique and the direct model transfer as we can capture a wider range of phishing tweets from the domain source that reduced the bias of the model. We have also shown interesting insights behind the difference in attributes between US phishing and US non-phishing as well as the difference between US phishing and NZ phishing. We have also presented the phishing landscape of individual communities and overall. We discover that the phishing tweets are similar and different at the same time, each community has its unique types of phishing attacks depending on the culture of the community. Not only do they have different types, but there is also a difference in the volume of tweets and the amount of proportionate phishing. NZ, in particular, has the lowest phishing proportionate to tweets in the region, as well as the lowest amount of phishing compared to the other communities.
e.g. planning; major activities; evaluation	

Outcomes

What outcomes were generated as a result of this project / research?

Outcomes are the changes that have occurred for the beneficiaries of your initiative. Generally outcomes can be framed as an increase or decrease in one or more of the following:

- Skills, knowledge, confidence, aspiration, motivation, (these are generally **immediate** or short-term outcomes)
- Actions, behaviour, change in policy (these are generally **intermediate** or medium-term outcomes)
- Social, financial, environmental, physical conditions (these are generally **long-term** outcomes)

Immediate outcomes occur directly following an activity (e.g. within 1 month); intermediate outcomes are those that fall between the immediate and long-term (e.g. between 1 month and 2 years); and long-term outcomes are those we expect to see years later (e.g. 2, 5, 10 or 50 years after the activity).

We also want to learn more about how you tracked the outcomes of your initiative - what you measured and how.

Internet research 2017/18
Internet Research final report
Application IR170017 From Dr Yun Sing Koh
Form Submitted 8 Aug 2019, 7:01pm NZST

If you need more help understanding what outcomes are, read the help sheets at www.ourcommunity.com.au/evaluation

List your initiative's outcomes and attached information in the following table. Leave blank any fields that do not apply to your project.

Outcome	Were these outcomes anticipated?	Timeframe	Indicator	Verification Method
A better understanding of the problem of spam phishing attacks on social networks.	Anticipated	Immediate	Phishing Landscape Survey on NZ and the types of phishing that is common.	Displayed a timeline showing which phishing attacks occur by region, we presented that phishing attacks are tailored to the community's culture.
We are changing the landscape of how current research into detecting spam attacks on social networks is carried out. Techniques need to be more proactive and detection mechanisms should be near real-time	Anticipated	Intermediate	The algorithm uses modern phishing detection techniques that work in real-time.	Ran the algorithm on 9 months worth of collected data from US, SG, AU and NZ.
We will be sharing the research including open-source code created for research purposes.	Anticipated	Immediate	Open source code via github	Able to view code online
We will be looking at the number of cases where New Zealanders are affected by phishing spams compared to other countries.	Anticipated	Immediate	Phishing Landscape Survey.	Displayed a timeline showing which phishing attacks occur by region, we have found that NZ has the lowest level of phishing compared to US, SG and AU over 9 months.

A phishing technique that can handle a real-world situation where phishing will be scarce.	Unanticipated	Intermediate	Detects phishing on a real-world dataset better than benchmark techniques that are designed for sandbox environments.	Showed that our algorithm performs better.
Outcomes are the changes that you believe were generated or influenced by your initiative. See information above.	Choose from the list	Choose from the list (see description above)	What you used to measure this outcome - e.g. 'change in teenage pregnancy rates from x to y'	e.g. survey; interviews; focus groups

What (if anything) did you change in your approach and practices as your project? research proceeded, and why? *
Wernsen Wong, the master student started in July 2018 instead of May 2018. This affected the plan slightly but we are currently on-track for deliverables on the 30/5/2019.

Have not attended netHui 2018, but expect to do so in 2019 in Wellington, when we are closer to the end of the project.

We may use this information to help inform others undertaking similar work

What did you learn as a result of undertaking this project/program? *
The computing resources needed for cloud computing was lower than expected, we were able to collect, label and evaluate Twitter data.

The time taken to build a scalable framework to ingest the number of tweets took longer than expected. We originally expected it to take 1 month however it took 2 months to figure out the Streaming API as well. It would have been better to start the collection of data well before the start of the project.

The number of phishing tweets was getting lower and lower. Twitter is improving the accounts challenged from year to year, the amount of phishing is lower than reported in other related work. We suspect that it was due to Twitter's policy on phishing attacks becoming more strict.

The features lost in a real-time environment may heavily affect the accuracy of the algorithm. With the loss in features such as re-tweets count and favourites count, we had to adapt the algorithm to improve the accuracy of the algorithm.

We are particularly interested in lessons that may help others undertaking similar work. Think about what you learned about your inputs (money, skills, personnel, time - too much; too little; about right?); your assumptions (were they 100% right, only partly right, or were the results a complete surprise?); and the context of the project/program (timing; targeted beneficiaries; geographic settings - were they right; wrong; about right?)

How will you share your learnings from this project/research? *
We have developed a website that will be available for public use.

We will be attending NetHui to discuss our research at the conference.

We have created an open source public Github repo with the code so others can use and extend it.

Internet research 2017/18
Internet Research final report
Application IR170017 From Dr Yun Sing Koh
Form Submitted 8 Aug 2019, 7:01pm NZST

We have submitted the research to an international peer reviewed conference: CIKM (http://www.cikm2019.net/).

What mediums were used to share the learnings? Have you reached the audience you expected?

We'd love to see some visual and audio representations of your work. Please share below.

Upload files:

Filename: CIKM2019.pdf
File size: 626.5 kB

Filename: InternetNZ_report.pdf
File size: 564.2 kB

and/or

Provide web link:

http://pelican-apdt.s3-website-ap-southeast-2.amazonaws.com
Must be a URL

and/or

Provide additional details:

Please include captions, if relevant

Can we use your media content in our own communications?

◉ Yes ○ No ○ Please contact us first
e.g. in our annual report

Financial Report

* indicates a required field

Project Income & Expenditure

Please provide details of any project income (funds received) and project expenditure (funds spent) to date.

Use the 'Notes' column to provide any additional information you think we should be aware of.

Income Description	Income Type	Confirmed Funding?	Income Amount ($)	Notes
InternetNZ	Other Income *	Confirmed *	$10,500.00	InternetNz Funds

Expenditure Description	Expenditure Type	Expenditure Amount ($)	Notes
Consumables	Administrative and Infrastructure *	$516.63	Disk storage, computation cost, printing.
NetHui Travel - for Masters Student	Other Expenditure	$544.58	Travel, Accommodation, Registration Fees
Master Students Fees	Salaries and Wages	$7,713.91	

Income and Expenditure Totals

Total Income Amount	Total Expenditure Amount	Income - Expenditure
$10,500.00 This number/amount is calculated.	$8,775.12 This number/amount is calculated.	$1,724.88 This number/amount is calculated.

Have you experienced any issues with your intended project budget to date? If so, please explain reasons for any major variances or for providing incomplete information:
AWS processing for students/educational purposes were cheaper than originally intended, due to accounts for educational purposes, and we received free credits for the computational processing.

Printing and binding is not completed as yet. As the Master's student has to finally print and bind his thesis after the examination process is completed. the Masters student has already submitted the thesis for examination in July 2019 and the examination normally takes about 3-6 months. Printing for thesis/binding normally cost around $283. (https://www.library.auckland.ac.nz/sites/public/files/documents/thesis-binding-form-06-2019.pdf)

Certification and Feedback

Feedback

You are now nearing the end of this form. Before you review your application and click the **SUBMIT** button please take a few moments to provide some feedback. (If you would rather provide anonymous feedback, please go to **{{ Grantmakers: provide a link to an anonymous survey or delete this sentence }}**

Please indicate how you found the acquittal process:
○ Very easy ◉ Easy ○ Neutral ○ Difficult ○ Very Difficult

How many minutes in total did it take you to complete this form?
120
Estimate in minutes (i.e. 1 hour = 60 minutes)

Please provide us with your suggestions about any improvements and/or additions to this form that you think we need to consider: