Identifying Internet Mediated Securities Fraud: Trends and Technology

Jake van der Laan, Brodie M. Shannon, and Christopher J.O. Baker

University of New Brunswick and New Brunswick Securities Commission

The Problem

The world wide web makes it much easier to commit securities fraud. Securities regulators do not become aware of these frauds until after the damage has been done. Identifying these operations early is difficult using traditional surveillance methods. Innovative technology driven solutions are required. Creating such tools requires both a multi-disciplinary understanding and approach to this problem.

What is Securities Fraud?

Securities fraud takes many different forms, but most involve either the selling of investments or the manipulation of their value. In many cases, the internet is used to advertise, promote, or effect the actual securities transaction.

Prevalence on the Web

Since 2001 internet mediated securities fraud has been one of the most common types of securities fraud and will likely become the dominant form in the coming years.

Types of Internet Securities Fraud

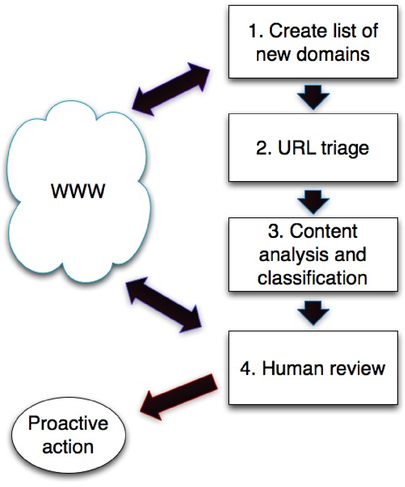
- Illegal distributions (sometimes referred to as boiler rooms. Securities offered by unregistered individuals or without proper disclosure.)
- Market manipulation (artificially drive a particular stock price up or down, in order to reap a profit from this change.)

Drivers of Internet Fraud

The internet makes identity easier to falsify and more difficult to authenticate, lowers the economic resources needed to set up, expands the "target market", and makes the proceeds of crime easier to divert. All of these factors weigh against effective law enforcement.

Scamalyzr

In 2009 we developed "Scamalyzr", a simple word based text classification tool which searches a corpus of continuously updated new website instances (retrieved from the web) for prevalence of a pre-determined set of relevant words, and then ranks them based on the presence and frequency of these words.
The tool is currently being used by the New Brunswick Securities Commission and identifies potential securities fraud websites on a daily basis.



Scamalyzr process flow

Results

Between March and September 2009, Scamalyzr processed over 13 million domains and classified these into 4 categories: suspect, not suspect, no-content, and parked. The no-content and parked categories were regularly re-scanned for content and re-classified accordingly. Fom the list of suspects, the highest ranking domains were reviewed by an analyst as a result of which boiler rooms were identified, and were placed on the New Brunswick Securities Commission website's Caution List or referred to law enforcement agencies and banks in the United States, the United Kingdom, Australia, and Canada.

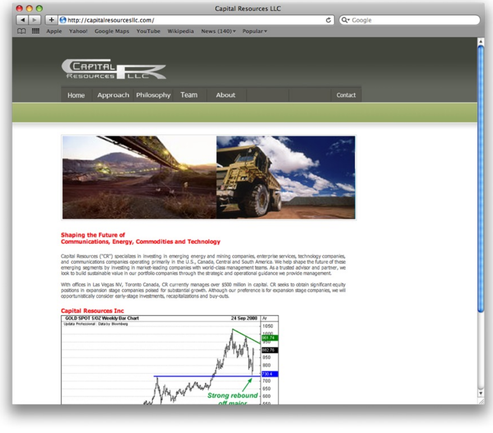
Opportunity

Even though the internet has enabled boiler rooms, it has also created a potential way of fighting them which was not previously available. This potential lies in the fact that the desired information (the promotional website and details of the touted investment) is accessible, as soon as the website is launched. The problem, of course, is finding it.

Challenges and Further Work

Despite providing valuable information, the system currently generates a high incidence of false positives ($> 80\%$). A more rigorous information extraction process will be implemented in the next phase of Scamalyzr development, using an ontology and machine learning approach (OBIE). We intend to extract information consistent with the concepts and relationships defined in the domain ontology, and then to maintain that information as instances of OWL ontologies, in turn enabling access to the output of the OBIE system by the Semantic Web.

The absence of useful metrics in assessing our current system is a problem and our future work will adapt established "precision" and "recall" metrics for information retrieval to our context.



Fraudulent investment website detected by Scamalyzr

The Importance of Web Science

It is truly surprising how people continue to surrender their hard earned money to someone they have never met. It is difficult to reconcile how people develop a high degree of trust in someone who calls them on the phone and convinces them, over a period of a few days by pointing them to a website and without actually meeting face to face, to invest in something which in most cases does not even exist.

Dealing with this is very much a multi-disciplinary problem. It requires not only an understanding of the technologies being used by scammers, but also, and perhaps more importantly, a better understanding of how people resolve trust and deception issues in a wired world, and how persuasion is effected online. These issues are not well understood at this time, in large measure due to the speed with which the internet, and particularly the web, has arrived on the scene, and started to fundamentally impact human interaction.

It is for that reason that Tim Berners-Lee and others' suggestion that a science of the web be created to seek to address the impact of technology and particularly the web is so timely and important. We need to engage in social analysis of the Web. We must, in particular, learn to understand how people respond to and deal with the negative aspects of web-based social interaction so that we can seek to ameliorate them.

Note

The primary author of this paper is a legal and securities industry professional interested in multi-disciplinary aspects of web science. The author invites comments or enquiries on the topic of this paper.

NEW BRUNSWICK SECURITIES COMMISSION

COMMISSION DES VALEURS MOBILIÈRES DU NOUVEAU-BRUNSWICK

UNB