Reaction to "Impression Fraud in On-line Advertising via Pay-Per-View Networks "

1) Technical details (approach/technique) that you found novel/ Something specific you learned that you didn't know before

I have to say I found nothing novel idea in this paper that I don't know before. Basically this paper is an analysis of data collected from fraudulent traffic providers and trying to identify the characteristic of such traffic. It doesn't involve much technical details that interest me.

2)Could I have done this work if I had the idea why or why not?

I think there's no difficulties for me to repeat the entire work if I have the intention to do so. The authors purchase the traffic from traffic merchants, which is open to everyone. The authors build their honeypots website using Amazon EC2, which is also open to public. It is just a matter of time to repeat everything they have done in the paper.

3)Is there anything I could do to repeat or validate?

A large dataset is always an interesting to look into. One thing I want to repeat is to collect the data set mentioned in this paper. As one of the most important thing in dealing with fraudulent traffic is to identify them, I am interested in looking for other characteristics that I can figure out from the data set.

4) What is my best idea for follow on work that I could personally do?

I noticed that they are using mustats.com to evaluate the popularity of websites. Mustats.com seems be able to do evaluation on any of the website without any data collection work in advance. I am very interesting in how this can be done and I want to see more into this.

There are several countermeasures mentioned in the paper, but not in depth. I think that it would be an interesting topic to think about building systems that works for both advertisers and intermediaries to detect and block such kind of invalid traffic.

5)What is my best idea for follow on work that I'd like to see the authors do?

As I mentioned in the previous question, I think it would be a good idea talking about building a fraudulent traffic recognizer for Pay-per-view advertisements to prevent forged impressions. I am personally interested in this topic. I also want to see more work done by the author.

6)Any logistical experimental lessons I learned?

One lesson I learned in this paper is some methods to build a honeypot on internet and prevent natural traffic from accessing it. As mentioned in the paper, the authors use techniques like editing robot.txt and adding "noindex, nofollow" in meta tags of the webpage. I think more can be done to build a more reliable honeypot.

7)How does this compare to the other papers we read? Most similar? How different? Other comparisons?

This paper also focuses on underground market. This topic is similar to the one that talks about underground market of twitter account. They both talk about underground transactions, do analysis to their economic scope and purchase data from providers to do further analysis on the characteristics of the data.

However, in the twitter paper, the authors have cooperation with twitter company, which gives them an advantage to apply their classification algorithm to twitter online environment and verify its efficiency. On the contrast, the persons who writes the paper to analyze purchased traffic only gives some general analysis of countermeasures, without any concrete method of defense.

8)What is your biggest criticism of the paper?

One thing I want to mention about this paper is that it is only a superficial analysis to the existence of fraudulent traffic without any in depth analysis of prevention method. Also, it seems to only look at part of the problematic traffic, which provides no strong similarity to the traffic it analyze.