Clickbot Detection via Behavioral Analysis

Chris Neasbitt
Computer Science Department
The University of Georgia
cjneasbi@uga.edu

Introduction

We are investigating how to detect malware that perpetrates click fraud (“clickbots”) without code inspection. Click fraud is a problem that has the potential to undermine the primary revenue source for no end-user cost internet services. Modern malware employs measures to thwart reverse engineering attempts. In order to defraud an ad-network, a clickbot must communicate using the protocol specified by the network, which is most commonly HTTP. We expect our work will speed the detection of clickbots.

Background /Related Work

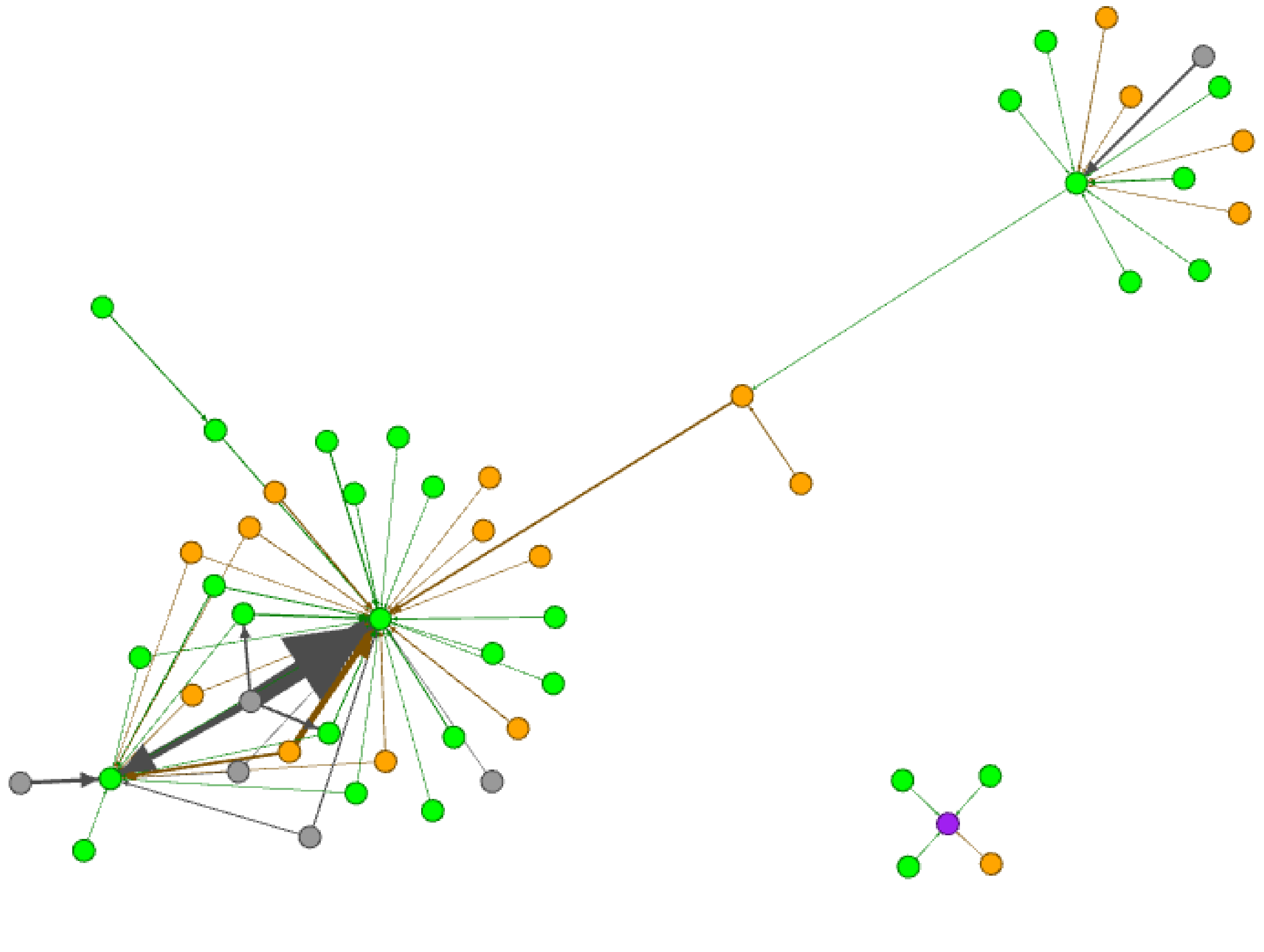
Previous work in reverse engineering clickbots has revealed several identifiable aspects of their operation, [1] [2]. These include forged referrers and requests to sites listed on well-known blacklists.

Approach

Our approach is twofold. First, we construct and analyze graphs of HTTP traffic generated by malware looking for patterns that resemble click fraud. Click identification is critical in recognizing these patterns. Disconnected subgraphs can indicate a possible manual interaction with the browser (“click”). Second, we plan to refine our click detection accuracy by instrumenting a browser to replay traffic from a trace.

Below: Referrer Graph of a HTTP Session

Each node represents a host to which an HTTP request was sent. Edges represent referrer relationships between requests. Each edge is directed from referee to referrer. Edge thickness represents the number of referrer relationships between nodes. Green nodes represent white-listed hosts. Orange nodes represent known ad network hosts. Purple nodes represent referrers not seen in the traffic trace. Gray nodes are of an undetermined type.



Acknowledgments

We would like to thank Dr. Roberto Perdisci, Phani Vadrevu, and the NSF for their support of and contributions to this work.

Discussion

We are currently experimenting two sets of data: one set of 2,038 malware traces and one set of 1,978 malware traces.. We have developed a two stage filtering process in order to focus our analysis on only those traces that may be produced by a clickbot. Our coarse-grain filter selects only those traces with at least one request sent to a well-known ad network. After applying the coarse-grain filter to the datasets, we derived two sets of candidate traces totaling 907 and 378. We are currently developing our fine-grain filter to generate and analyze the referrer graphs from the candidate traces.

Contributions

- Identification of a malware as a clickbot.
- Detection of clicks from HTTP traffic traces.

References

1. Daswani, N. & Stoppelman, M. (2007). The anatomy of Clickbot.A. In , *Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets* (pp. pp. 11-11). : USENIX Association.

2. Miller, B., Pearce, P., Grier, C., Kreibich, C. & Paxson, V. (2011). What’s Clicking What? Techniques and Innovations of Today’s Clickbots. In, *Lecture Notes in Computer Science*. , 6739, 164-183.