Mathematical Prediction of Botnet Propagation Dynamics

Julian Rrushi, Ali A. Ghorbani

Faculty of Computer Science, University of New Brunswick

Motivation

The tactical design of botnets has been evolving towards higher degrees of sophistication.

Mathematical models that capture propagation dynamics of botnets and malware in general have drawn considerable research attention in the last decade. Nevertheless, the majority of that research focuses on the mathematical models per se and not on possible ways of exploiting those models for botnet mitigation.

We present an approach that explores mathematical modeling of botnet propagation dynamics to detect early stage botnet infections in an enterprise network. The ultimate objective behind this research is to enable botnet containment in an effective and timely fashion.

Botnet containment is conducted by logically grouping and controlling infected hosts via automatic establishment and configuration of a Virtual Local Area Network (VLAN).

A VLAN allows for separating infected hosts from uninfected hosts regardless of their physical location in the enterprise network. Clearly the research challenge takes the form of detecting the first few infections in order to guarantee containment.

Method

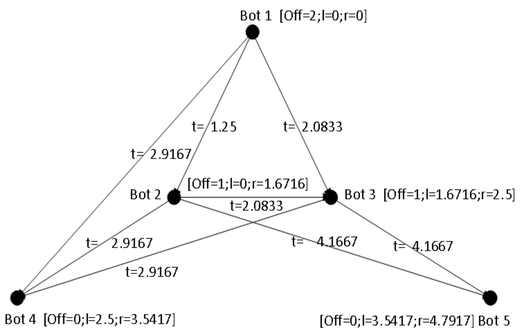
Hosts are modelled as vertices, while end-to-end network communications between hosts are modeled as directed edges. Hosts are labelled by IP addresses. Edges are assigned attribute vectors to characterize the network connections. Vertices are assigned attribute vectors that indicate the corresponding number of offspring. We refer to this graph as the data graph.

During practical experimentation we observed that early stage botnet infections form weakly connected subgraphs within the data graph. Thus in this research, formation of a weakly connected subgraph is indicative of a suspected botnet outbreak.

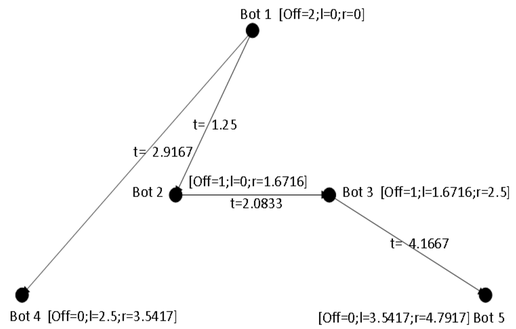
We have devised a novel application of statistics and number theory that allows for inferring the botnet propagation dynamics from network traffic analysis. That inference is materialized into a model graph.

In the model graph, vertices represent bots and are labelled sequentially according to their time of birth in the enterprise network. Directed edges indicate parent-child relationships between bots. Each edge is assigned an attribute vector that indicates the mean time at which the corresponding infection is predicted to haven take place.

We have devised a subgraph search algorithm that seeks to find an isomorphism between the model graph and a weakly connected subgraph formed within the data graph. If a match between the two is found, we conclude that the weakly connected subgraph in question indeed represents a botnet outbreak.



Example of a model graph as generated by our approach



An instance of that model graph with mutually exclusive concurrent edges removed

The labels of the vertices in the weakly connected subgraph give us the IP addresses of the hosts that have been infected by bots.

The subgraph search algorithm is error-tolerant, and thus is able to tolerate distortions. It makes use of an error model incorporated within the model graph.

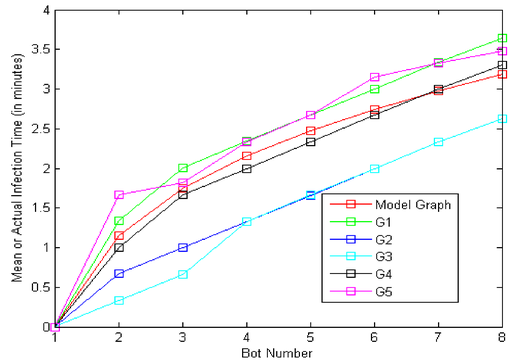
The error model specifies variability around infection times, mutually exclusive concurrent edges, and variability around the number of offspring of each bot. Those data are conveyed as attribute vectors assigned to the vertices of the model graph.

Evaluation

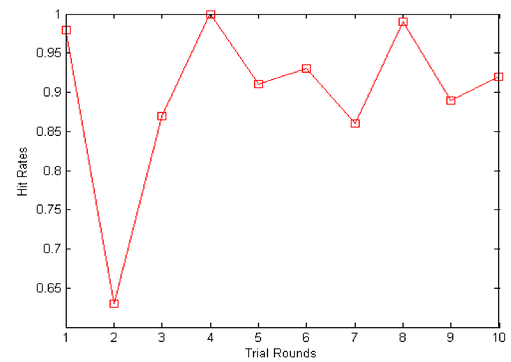
We have implemented this research as a tool that we dubbed *Variant*. The part of this research that regards mathematical modeling and statistical inference of botnet propagation dynamics was implemented in Matlab. The subgraph isomorphism search algorithm was implemented in Perl.

We wrote test software that emulates a botnet in a controlled fashion. Extensive practical experiments with this research in relation to the botnet emulating software were conducted in the Emulab network testbed.

Empirical observations reveal that our approach can predict botnet propagation dynamics with low error rates. Our approach yields variably high hit rates with infrequent disturbances due to the highly stochastic nature of the botnet infection process.



Host infection times in the model graph and in some of the weakly connected subgraphs targeted by our approach in the Emulab network testbed.



Hit rates observed empirically for various rounds of trials with a test botnet in the Emulab network testbed

Conclusions

This research demonstrated the viability of mathematical models of botnet propagation dynamics as effective means of enabling botnet containment in an enterprise network.

The experiments that we conducted along with experimental data that we obtained are indicative of the effectiveness of our approach, and also suggest a clear potential of mathematical models of botnet propagation dynamics for practical botnet mitigation.

The *Variant* tool is a workable prototype implementation of this research. It receives in input pcap files generated by *Tcpdump* or any other derivative network sniffing tool. In the case of a positive botnet interception, the output produced by the *Variant* tool consists of a textual description of the final form of the weakly connected subgraph, which was formed by the botnet infection process within the data graph.

Acknowledgements

This research was supported by a NSERC grant awarded to Ali A. Ghorbani. The views expressed in this poster are those of the authors only.