# Wide-area Routing Dynamics of Malicious Networks

Maria Konte, Nick Feamster
Georgia Tech
{mkonte, feamster}@cc.gatech.edu

## ABSTRACT

This paper studies the routing dynamics of malicious networks. We characterize the routing behavior of malicious networks on both short and long timescales. We find that malicious networks more consistently advertise prefixes with short durations and long interarrival times; over longer timescales, we find that malicious ASes connect with more upstream providers than legitimate ASes, and they also change upstream providers more frequently.

**Categories and Subject Descriptors:** C.2.0 [Computer Communication Networks]: Security and protection.

**General Terms:** Measurement, Security, Management.

**Keywords:** Hostexploit, malicious networks, BGP, routing dynamics, security.

## 1. INTRODUCTION

Attackers use advanced techniques to launch spam campaigns or mount attacks; in particular, Ramchandran *et al.* [**?**] observed that some some attackers send spam from hijacked prefixes. Previous studies have studied different malicious activities in isolation such as worms, botnets, spam, and Internet scam infrastructure [**?**, **?**] and proposed detection solutions for each of them. Unfortunately, observing just the nature of the attack itself is not sufficient, because any given network or host may mount different attacks on different days. Rather than attempting to detect any individual attack, we characterize the *routing behavior* of malicious networks that are primarily responsible for cybercriminal activities and identify features of this behavior that may be more stable across time. Although our initial goal is simply to characterize this behavior, our belief is that ultimately certain aspects of routing behavior may serve as invariants for detecting malicious infrastructure, even as the attacks themselves evolve. Specifically, we believe that routing behavior may ultimately be useful for identifying and monitoring the activity of networks that mount attacks, even as the attacks themselves change.

We perform the first systematic study of the wide area routing behavior of malicious networks. To understand the nature of malicious routing behavior with respect to legitimate domains, we compare the routing behavior from autonomous systems (ASes) listed in Hostexploit to the routing behavior of the ASes that host the top 500 Alexa domains. Hostexploit [**?**], an organization that correlates data from multiple sources such as spam, malware, spam bots, botnet C&C servers, and phishing servers from industry partners, has been publicizing lists of ASes that show the highest levels of cybercriminal activity for the last three years. Hostexploit rates each AS with an index based on the activity of the AS weighted by the size of its allocated address space. There are examples of these networks that were detected and eventually disconnected including ATRIVO/Intercage on September 2008 and VolgaHost on January 2011.

We perform our analysis across two timescales. On short timescales, we study how malicious networks advertise their prefixes. For every prefix, we group the announcements into distinct events and study the duration and interarrival time of each event over the period of one month. We perform our analysis for two months, January and June, for four consecutive years (2008–2011). We show only the results for January 2011, but our results are qualitatively consistent for all time periods. We find that malicious ASes advertise a large portion of their prefixes (45%) over events that have both shorter duration *and* longer interarrival time than legitimate ASes. Over long timescales, we study the wiring trends of malicious ASes and find that malicious ASes connect with more providers than legitimate ASes; they also change their upstream providers more frequently.

Copyright is held by the author/owner(s).
*SIGCOMM'11,* August 15–19, 2011, Toronto, Ontario, Canada.
ACM 978-1-4503-0797-0/11/08.

## 2. DATA

We use the list of networks that are reported as top in Internet criminal activities according to Hostexploit from 2009–2011 to identify a set of malicious ASes. We augment this dataset with the following sources: (1) To study the behavior of malicious networks on *short timescales*, we collect the complete set of BGP updates from malicious networks, from all RouteViews monitors. We cluster the BGP updates we observe for each prefix into *prefix events*. A prefix event begins when a RouteViews monitor receives a new announcement for a prefix from some origin AS. A prefix event ends at the time when the same monitor receives the *first* withdrawal for the corresponding prefix. (2) To understand the behavior of malicious networks over *long timescales*, we obtain a publicly available dataset of customer-provider links formed among ASes over a ten-year period [**?**]. To understand how malicious behavior of malicious ASes differs from legitimate ASes, we perform the same analysis for the ASes that host the top 500 Alexa domains.

## 3. PRELIMINARY RESULTS

**Malicious networks advertise their prefixes over events with shorter duration and longer interarrival time than legitimate networks.** To study the behavior of malicious networks over short timescales, we proceed as follows: (1) We cluster all the announcements for all the prefixes that originate from malicious networks, as observed from all RouteViews monitors for one month into events as described in Section **??**. (2) We *jointly* examine the duration and

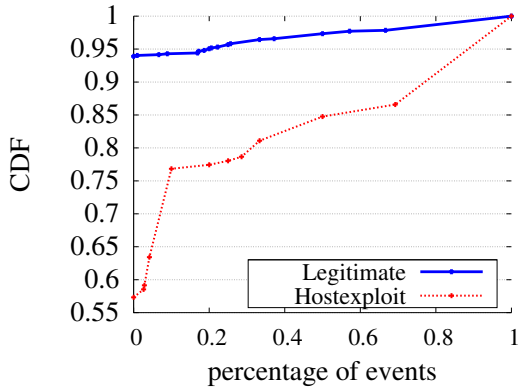
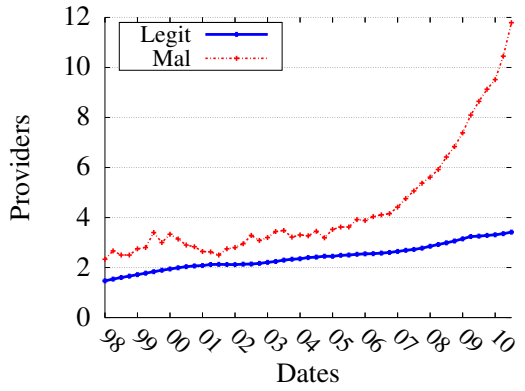
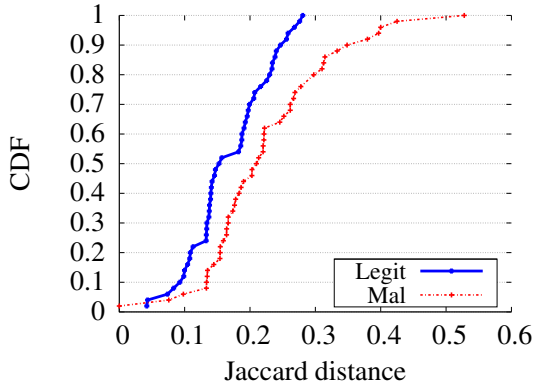



**Figure 1: Malicious networks advertise their prefixes over events with shorter duration and longer interarrival time than legitimate networks.**




(a) Malicious ASes connect to more providers over time.




(b) Malicious ASes are more aggressive in connecting to new upstream providers.

**Figure 2: Malicious ECs link with more providers through their lifetime and change their upstream connectivity more frequently than legitimate EC.**

the interarrival time of these events. More specifically, for each prefix, we categorize its events into four types: a) long duration - short interarrival time, b) long duration - long interarrival time, c) short duration - short interarrival time and d) short duration - long interarrival time. We consider the *duration* of a prefix event to be short if is less than five minutes, which is the median advertisement duration of about 40% of the prefixes. Similarly, we consider the *interarrival time* of a prefix event to be short if it is less than 100 seconds, which is the median interarrival value for about 40% of the prefixes for all of their events. (3) For every prefix, we compute the fraction of the events we observe for each event type. Figure **??** shows the distribution of events per prefix for each event type. The most striking finding is that about 80% of the prefixes advertised by malicious ASes are advertised mostly with events of short duration and long interarrival time for about 70% of prefix events, whereas roughly 85% of the legitimate prefixes have no events of this type at all.

**Malicious networks link with more providers through their lifetime than legitimate networks.** We examine the total number of providers that ASes connect to over their lifetime. Here, we focus on the ASes that are Enterprise Customers (EC) according to their AS classification [**?**]. Figure **??** shows the cumulative number of average providers that malicious and legitimate ASes connect to over time. We note that both the number of ASes and the cumulative number of providers increase as time progresses, but the rate at which the two increase is not always the same. For some snapshots we observe a small decrease in the average cumulative number of providers. We observe that malicious ASes link on average with a total of twelve providers over the course of twelve years, whereas the legitimate ASes connect to an average of four providers. AS23456 linked with a total of 183 providers from April 2007 through January 2010.

**Malicious networks change their upstream connectivity more frequently than legitimate networks.** To quantify the aggressiveness of an AS in changing its upstream connectivity, we consider the distance between the set of the AS's providers for two consecutive snapshots. We use the Jaccard distance as a metric of distance between the set of providers between two consecutive snapshots. For example, a Jaccard distance of 0.8 indicates that 80% of the links seen in the two snapshots are observed in one of the two snapshots but not in both. We calculated the Jaccard distance between two consecutive snapshots for each AS throughout its lifetime. Figure **??** shows the distribution of the Jaccard distance values. We observe that malicious ASes more frequently change their upstream connectivity than legitimate ASes.

## Acknowledgements

This work is funded by the National Science Foundation under NSF CAREER Award CNS-0643974 and CNS-0831300, and the Department of Homeland Security under contract number FA8750-08-2-0141.

## REFERENCES

[1] A. Dhamdhere and C. Dovrolis. Ten Years in the Evolution of the Internet Ecosystem. In *Proceedings of ACM SIGCOMM/USENIX Internet Measurement Conference (IMC).*, 2008.

[2] Hostexploit. `http://www.hostexploit.com`.

[3] H. A. Kim and B. Karp. Autograph: Toward automated, distributed worm signature detection. In *the 13th conference on USENIX Security Symposium*, 2004.

[4] C. Kreibich and J. Crowcroft. Honeycomb: Creating intrusion detection signatures using honeypots. In *2nd Workshop on Hot Topics in Networks (HotNets-II)*, 2003.

[5] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. In *Proceedings of Sigcomm*, 2006.