This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001119C0102. The opinions, findings, and conclusions stated herein are those of the authors and do not necessarily reflect those of DARPA.

Preprocessing for the Greenwich Meanie

Introduction

The “Greenwich Meanie” (also known simply as the “Meanie”) is an interesting IBR phenomenon involving a large number of IBR sources behaving in an unusual manner that appears to be coordinated, but does not appear to be directly related to any other IBR behavior.

The Meanie was brought to our attention by Michael Collins at ISI, who had identified an unexpected phenomenon in the IBR gathered by his small telescope in the USC/ISI network. Using our larger telescope, we were able to confirm his findings, and refine our understanding of the Meanie. Since his initial observation, we have found pervasive evidence of the Meanie in the data of every telescope we have been able examine, and learned that other researchers noticed the Meanie behavior in their own data starting in early 2020.

Characteristics of the Meanie

Meanie packets are always UDP, unfragmented, and contain a payload. The payloads have a uniformly distributed size in a fixed range, and are either encrypted or random – we have detected no structure to the payload, or any hints in the distribution of byte values in the payload.

The most interesting thing about the Meanie packets is that the packets for a given day all have the same destination port (in the range that appears to be from 48K to 64K-1), and the destination port changes at midnight UTC every day. It seems reasonable to think that the destination port is chosen as a function of the date, but the choice appears random and we have not been able to predict future ports.

The Meanie is unusual for several additional reasons:

- *There are a large number of packet sources.*

  Even after almost four years of observing Meanie sources, we still see large numbers of new sources *every day.* For a small telescope, most of the Meanie sources observed during a given day are sources that *have never been observed before.*

  This is different from most IBR sources, which are relatively stable.

- *Each sources only sends a small number of packets.*

  In four years of data collection, our telescope has only received one or two packets from most Meanie sources. Most IBR sources engage in periodic episodes of scanning, sending thousands of packets into our telescope when they are active. In contrast, Meanie sources typically send packets into our telescope every few years.

- *The packet sources are dispersed widely around IPv4 space.*

  Most scanners that have more than one source have sources that are related in some manner, such as all belonging to a collection of /24 subnets or part of a block allocated to a hosting provider. For example, we have seen scanners from Brazil that use hundreds or thousands of different sources to scan our telescope (with each source only sending a handful of probes), but all of the sources are part of a single subnet (i.e. a /18).

  The Meanie packets come from many sources, but also from many subnets. In fact, the addresses of the sources is spread out that the number of unique source /24 subnets is almost as high as the number of sources – instead of being clumped together in subnets owned by a single organization, Meanie sources are usually relatively distant from each other.

- *The distribution of destination addresses is not uniform.*

Although we have received many millions of Meanie packets in our telescope (which has only contained, at its largest, a few thousand destination addresses), many destination addresses in our telescope have *almost never* been the destination of a Meanie packet.

Almost all scanners (or groups of scanners working together) cover the entire telescope, or at least entire subnets within the telescope, but the Meanie omits some addresses entirely, and addresses some of the other addresses very rarely (rarely enough that the apparent Meanie packets might be caused by a different scanner that happened to collide with the Meanie port by random).

For example, for one of our /24 subnets, from 2/2020 until 12/2023, we observed 2.35 million Meanie packets. Of the 256 possible destinations, 30 addresses were never seen and another 54 were only seen once or twice. An additional 19 were seen less than 2000 times, and then there are 96 addresses approximately 9700 times, 37 seen approximately 19800 times , 17 seen approximately 29400 times, three that were seen about 39000 times, and one that was 49325 times. There is no obvious relationship between the value of an address and the number of Meanie packets addressed to it.

- *The size of the payload is uniformly distributed between 65 and 199 bytes.*

  The distribution is nearly uniform (with a slight skew towards the larger values). For the same subnet mentioned above, the sizes vary between 17088 and 19282, with a median of 17340.

- *The payload appears to be random, or encrypted in a manner that makes it appear random.*

Processing steps

The `find-meanies.sh` script finds Meanie packets in new input files.

Like most of the other scripts, `find-meanies.sh` takes its parameters from a parameter file, environment variables, or values on its commandline. See `../../params/example.params` for a description of the parameters. The required parameters are `DATANAME`, `PCAPDIR`, `FCSVDIR`, and `MEANIEDIR`.

The first step for each new day (starting at midnight UTC) is to find the destination port that the Meanie is using that day, using heuristics based on the unusual characteristics of the Meanie (described above). The best candidate for each day is stored in in a file in the `ports` subdirectory of the Meanie output.

After the best candidate has been found, the packets for the corresponding port are extracted from the complete pcap files, and stored in the `pcap` subdirectory of the Meanie output.

Next, the resulting pcap files are converted to CSV (in the format described in `../../C/meanie2csv.c`).

Finally, any “missing” hourly files (for hours during which there was no data) are created, so that every hour has a file, even if it is empty.

Some example post-processing tasks are also included, such as computing the number of Meanie packets observed each hour, either for the entire telescope or a subset.