## NAME

traffic_learner - Samba tool to assist with traffic generation.

## SYNOPSIS

traffic_learner {-o OUTPUT_FILE ...} [-h] [--dns-mode {inline|count}] [SUMMARY_FILE] [SUMMARY_FILE ...]

## DESCRIPTION

This tool is part of the **samba**(7) suite.

This tool assists with generation of Samba traffic. It takes a traffic-summary file (produced by traffic_summary.pl) as input and produces a traffic-model file that can be used by traffic_replay for traffic generation.

The model file summarizes the types of traffic ('conversations' between a host and a Samba DC) that occur on a network. The model file describes the traffic in a way that allows it to be scaled so that either more (or fewer) packets get sent, and the packets can be sent at a faster (or slower) rate than that seen in the network.

## OPTIONS

-h|--help
: Print a summary of command line options.

SUMMARY_FILE
: File containing a network traffic-summary. The traffic-summary file should be generated by traffic_summary.pl from a packet capture of actual network traffic. More than one file can be specified, in which case the traffic will be combined into a single traffic-model. If no SUMMARY_FILE is specified, this tool will read the traffic-summary from STDIN, i.e. you can pipe the output from traffic_summary.pl directly to this tool.

-o|--out OUTPUT_FILE
: The traffic-model that is produced will be written to this file. The OUTPUT_FILE can then be passed to traffic_replay to generate (and manipulate) Samba network traffic.

--dns-mode [inline|count]
: How DNS traffic should be handled by the model.

## EXAMPLES

To take a traffic-summary file and produce a traffic-model file, use:

traffic_learner traffic-summary.txt -o traffic-model.txt

To generate a traffic-model from a packet capture, you can pipe the traffic summary to STDIN using:

tshark -r capture.pcapng -T pdml | traffic_summary.pl | traffic_learner -o traffic-model.txt

## OUTPUT FILE FORMAT

The output model file describes a Markov model estimating the probability of a packet occurring given the last two packets.

The count of each continuation after a pair of successive packets is stored, and the ratios of these counts is used to calculate probabilities for the next packet.

The model is stored in JSON format, and also contains information about the packet rate and DNS traffic rate.

### Example ngram listing

The following listing shows a contrived example of a single ngram entry.

```
"ngrams": {
  "ldap:0\tdcerpc:11": {
     "lsarpc:77": 1,
     "ldap:2": 370,
     "ldap:3": 62,
     "wait:3": 2,
     "-": 1
  }, [...]
}
```

This counts the observed continuations after an ldap packet with opcode 0 (a bind) followed by a dcerpc packet with opcode 11 (also a bind). The most common next packet is "ldap:2" which is an unbind, so this is the most likely packet type to be selected in replay. At the other extreme, lsarpc opcode 77 (lookup names) has been seen only once, and it is unlikely but possible that this will be selected in replay.

There are two special packet types here. "wait:3" refers to a temporary pause in the conversation, where the "3" pseudo-opcode indicates the length of the wait on an exponential scale. That is, a "wait:4" pause would be about 2.7 times longer that a "wait:3", which in turn would be similarly longer than a "wait:2".

The other special packet is "-", which represents the limit of the conversation. In the example, this indicates that one observed conversation ended after this particular ngram. This special opcode is also used at the beginning of conversations, which are indicated by the ngram "-\t-".

## VERSION

This man page is complete for version 4.13.17 of the Samba suite.

## SEE ALSO

**traffic_replay**(7).

## AUTHOR

The original Samba software and related utilities were created by Andrew Tridgell. Samba is now developed by the Samba Team as an Open Source project similar to the way the Linux kernel is developed.

The traffic_learner tool was developed by the Samba team at Catalyst IT Ltd.

The traffic_learner manpage was written by Tim Beale.