Enriched data sets provided by Tor Metrics

Part X

data-set-human-readable-name

Draft

tbd
The Tor Project
e-mail
2017

Contents

1 Introduction 1
2 General description of data-set 1
3 Statistics, estimates, and predictions 2
4 Summary 2

Abstract
This document is concerned with the statistics and enriched data sets provided by Tor Metrics. ...

1 Introduction

...

2 General description of data-set

In general, all data sets are provided as comma separated value files and have a row based record structure. The rows in one data set contain different types of information, and a row usually contains several fields that identify properties of the data record, e.g., a date or server type, and other fields that contain counts of items, e.g., clients or servers, confirming to the properties. Empty value fields
in a row signify either that the items in these particular records don’t have the respective properties or that the record refers to all items without separating by the value of the property.

2.1 Details

Name the files the data set consists of and their general use.

The following lists describe the contents of the data set(s).

2.1.1 Fields of data-set

Data set data-set-file-name consists of records of X values.

date: Contains the UTC date of the beginning of the measurement interval, for which this row provides estimates. The date is derived from extra-info descriptor field dirreq-stats-end.

...: ...

...: one for each column

...: ...

3 Statistics, estimates, and predictions

general description of method used incl. assumptions etc.

3.1 Calculation

Details incl. formulae for computing the statistics/aggregated data sets.

3.2 Implementation

Implementation details that influence the resulting data set(s).

3.3 Reliability estimates

estimates for the quality of the provided data set(s) based on the two previous sections. Incl. error ranges and influences of outliers etc.

3.4 Improvements

Maybe add possible areas of improvement? But a little out of scope here.

4 Summary