

KDD Cup 1999 Data

Abstract

This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 The Fifth International Conference on Knowledge Discovery and Data Mining. The competition task was to build a network intrusion detector, a predictive model capable of distinguishing between ``bad" connections, called intrusions or attacks, and ``good" normal connections. This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment.

Information files:

- [task description](#). This is the original task description given to competition participants.

Data files:

- [kddcup.names](#) A list of features.
- [kddcup.data.gz](#) The full data set (18M; 743M Uncompressed)
- [kddcup.data_10_percent.gz](#) A 10% subset. (2.1M; 75M Uncompressed)
- [kddcup.newtestdata_10_percent_unlabeled.gz](#) (1.4M; 45M Uncompressed)
- [kddcup.testdata.unlabeled.gz](#) (11.2M; 430M Uncompressed)
- [kddcup.testdata.unlabeled_10_percent.gz](#) (1.4M; 45M Uncompressed)
- [corrected.gz](#) Test data with corrected labels.
- [training_attack_types](#) A list of intrusion types.
- [typo-correction.txt](#) A brief note on a typo in the data set that has been corrected (6/26/07)

[The UCI KDD Archive](#)
[Information and Computer Science](#)
[University of California, Irvine](#)
Irvine, CA 92697-3425
Last modified: October 28, 1999