DATA SETS COMPOSITION AND UTILIZATION DESCRIPTION

ADFA-LD, ADFA-WD and ADFA-WDSAA are labelled data that contains following three different folders:

(i) Training data (contains only normal traces)
(ii) Validation data (contains only normal traces)
(iii) Attack data (contains only attack traces)

How to use?

In the case of anomaly intrusion detection system (IDS) design you can use just training normal data in training phase with normal label if you want to do supervised ML and at testing phase you can use all attack data and validation data to measure Detection rate (DR), False positive rate (FPR), False negative rate (FNR) and False alarm rate (FAR).

In the case of signature IDS design you can use all normal training data with label normal and some attacked data (from test attack data) with label attack while training phase and using supervised machine learning (ML). During testing you can use rest of the attack data that is not use in training and all validation normal data, to measure DR, FPR, FNR and FAR.
In the case of unsupervised ML specific to anomaly IDS design, the training phase would require normal training data without labels and for testing phase, all the attack data and validation data with labels can be used to measure DR, FPR, FNR and FAR.

Further you can use following research articles for study and reference as they utilized the above-mentioned datasets.

(1). Haider, Waqas, Jiankun Hu, and Miao Xie. "Towards reliable data feature retrieval and decision engine in host-based anomaly detection systems." *Industrial Electronics and Applications (ICIEA), 2015 IEEE 10th Conference on*. IEEE, 2015.

(2). Haider, Waqas, et al. "Integer Data Zero-Watermark Assisted System Calls Abstraction and Normalization for Host Based Anomaly Detection Systems." *Cyber Security and Cloud Computing (CSCloud), 2015 IEEE 2nd International Conference on*. IEEE, 2015.

(3). Haider, Waqas, et al. "Windows Based Data Sets for Evaluation of Robustness of Host Based Intrusion Detection Systems (IDS) to Zero-Day and Stealth Attacks." *Future Internet* 8.3 (2016): 29.

Note: Any further query that we believe is available in the above articles or is the part of your own study might not be entertained.

Provided By: Prof. Jinakun Hu.
Email : J.Hu@adfa.edu.au