Decision Theoretic Approach

Overview

In this project we study and investigate network anomaly detection algorithms for Internet Paths. We also develop a *Decision Theoretic Approach* (DTA) based on our observations regarding the characteristics of the performance-measurement statistics obtained from the IEPM-BW project.

To study and compare the algorithms we use the data sets collected by IEPM-BW spanning approximately 3 years (i.e. 2005 - 2008). The Internet paths observed were the links between Stanford Linear Accelerator Center (SLAC) and the following sites:

1. University of Toronto, Canada.
2. Deutsches Elektronen-Synchrotron, Germany.
3. Forschungszentrum Karlsruhe, Germany.
4. European Organization for Nuclear Research, Geneva, Switzerland.
5. San Diego Supercomputing Center, USA.
6. Switch, Switzerland.
7. University of Florida, USA.
8. National Laboratory for Particle and Nuclear Physics, Canada.
9. Oak Ridge National Laboratory, USA.
10. Budker Institute of Nuclear Physics, Russia.
11. Daresbury Laboratory, United Kingdom.
12. California Institute of Technology - CACR, USA.
13. Istituto Nazionale di Fisica Nucleare, Italy.
14. Czech NREN Operator, Czech Republic.
15. Brookhaven National Laboratory, USA.
16. Argonne National Laboratory, USA.
17. California Institute of Technology - Ultralight, USA.

The topology of the monitoring framework is shown in figure 1.

Fig. 1: Topology of IEPM as of 07/2008

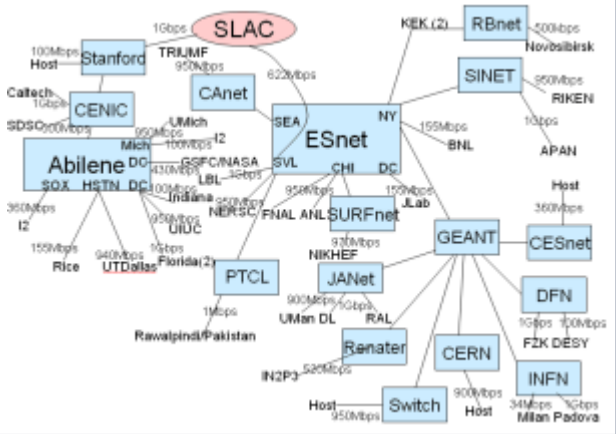
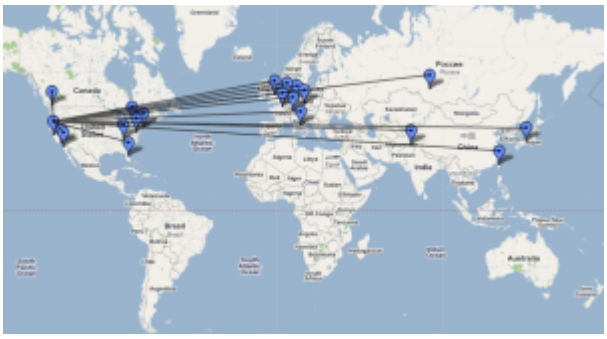


Fig. 2: Deployment of Selected Sites



The number of measurements made to the following sites from SLAC:

Site	pathchirp	iperf	thrulay
cern.ch	48647	24586	39510
desy.de	32247	4522	28689
fzk.de	65536	4874	42708
nslabs.ufl.edu	41206	1549	28613
switch.edu	19668	4638	28744
sdsc.edu	21176	4416	22456
triumf.ca	26425	4669	27021
utoronto.ca	40614	5003	21646
ornl.gov	35339	5182	18375
anl.gov	17968	1	27559
bnl.org	23580	20708	16000
cacr.caltech.edu	61871	25525	37293
dl.ac.uk	27806	6096	28058
nsk.su	20117	1	26845

cesnet.cz	23618	3062	28426
infn.it	30372	4343	28573
ultralight.caltech	3739	88	1534
SubTotal	**539929**	**119263**	**452050**

Data Sets

The data sets used in the study may be downloaded from the links listed below. These data sets were collected by the IEPM-BW project

Table 1: Performance measurement statistics compiled by IEPM, as seen from SLAC.

	Data Sets with Events	**Data Sets with no Events**
IEPM	[rar] 3.4 MB, [zip] 3.6 MB	[rar] 3.3 MB, [zip] 3.5 MB

All files with name "filename_raw_dataset.pathchirp" contain the raw data i.e the available bandwidth measurements along with the timestamps which are used in all algorithms.
All files with name "filename_event_file.txt" contain the list of events identified.

Technical Report - Labeling and Comparative Analysis

The technical report titled "A performance evaluation of anomaly detection algorithms for Internet Paths" will be available soon.

Input/Tuning parameters

Plateau Algorithm (PL)

History Buffer Length (H)	**Trigger Buffer Length (T)**	**Threshold (th)**	**Sensitivity (s)**
240	6 - 45	0.10 - 0.70	1.0 - 2.8

Kalman Filters Method (KF)

Sensitivity (K)	**Time Window (h)**
0.001 - 11.0	6 - 20

Holt Winter's Method (HW)

? - alpha	**? - beta**	**? - gamma**	**? - sigma**
0.1	0.1 - 0.3	0.1 - 0.5	2.0

Adaptive Fault Detector (AFD)

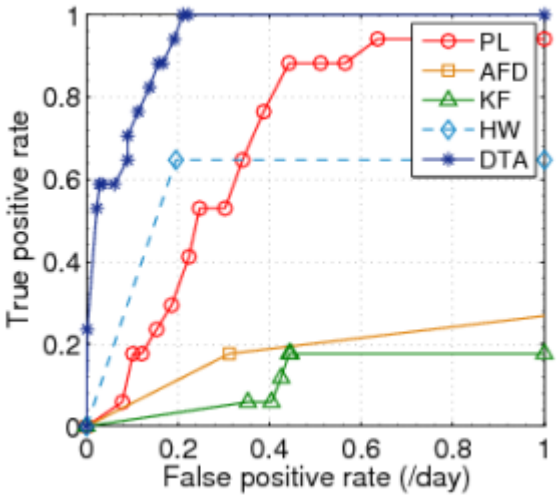
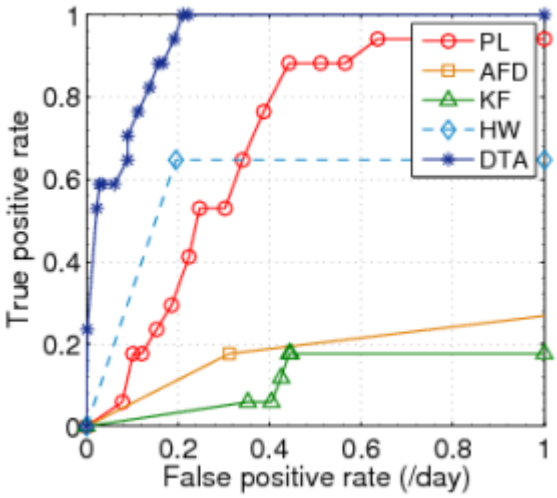
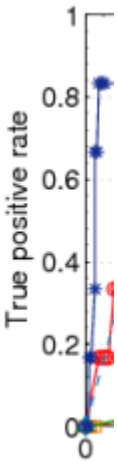
Window Size (N)	**? - alpha**	**? - beta**	**No. of Training Data (Hn)**
20	0.95	0.0015 - 0.1	100

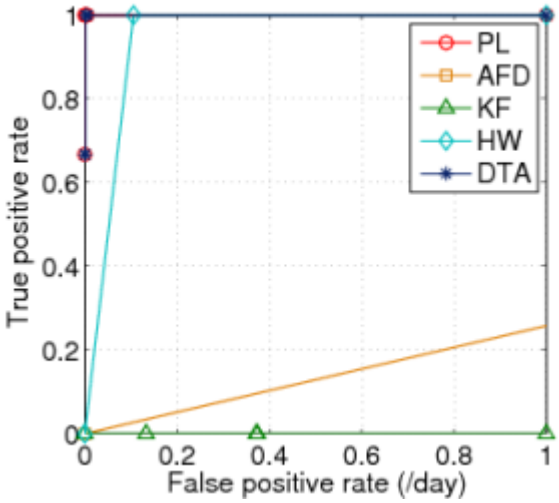
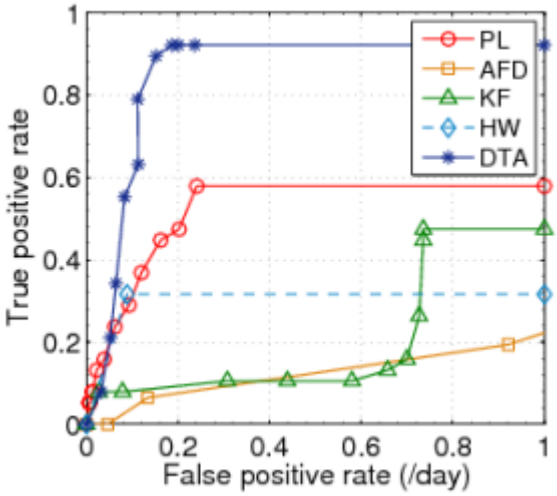
Decision Theoretic Approach (DTA)

History Buffer Length (N)	**? - alpha**	**? - beta**	**Median filter length (n)**
30 - 90	0.01 - 0.34	0.99	100

ROC Results

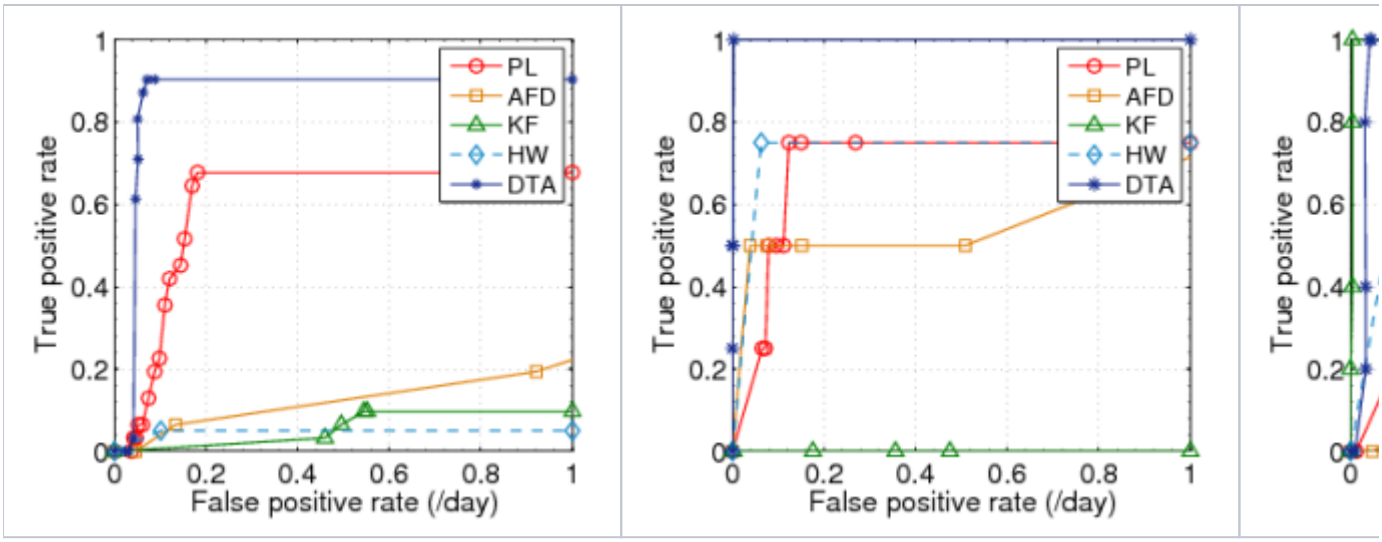
Datasets with Gaussian Distributions

CERN	FZK	SDSC
		

TRIMUF	UTORONTO
	

Datasets with Weibull Distributions

DESY	NSLABS	SWITCH


True positive rate
False positive rate (/day)
PL
AFD
KF
HW
DTA
0
0.2
0.4
0.6
0.8
1