Appendix for the paper:
Alice Marascu, Suleiman Ali Khan, Themis Palpanas. ***Scalable Similarity Matching in Streaming Time Series***. Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Kuala Lumpur, Malaysia, 2012

1 Appendix-A: Data sets Description

All data sets obtained from the UCR Time Series Data Archive
http://www.cs.ucr.edu/~eamonn/time_series_data/

All Data sets were z-normalized (μ=0, σ=1) for consistency. Patterns of length 50,100,250 and 500 were extracted from random locations[*1] in each stream. Each stream was matched against 10 random patterns, and the results averaged.

Dataset (each dataset contains more than one time series)	**# of processed Streams in data set**	**# of Data points per Stream**
2D Time Series	9 [*2]	900-64732 [*4]
Green Screen	8 [*2]	16902-22502[*4]
Power demand Italy AEM	1 [*3]	29931
Gene Expression	2 [*2]	17820-23760[*4]
ECG_znorm205	1 [*3]	11536
EEG_heart_rate	1	7200
Fluid_dynamics	1	10000
Light_curve	1	27291
Physiological_data_B1	1 [*3]	51000
Physiological_data_B2	1 [*3]	51000
Realitycheck	1 [*3]	14000
ballbeam	1 [*3]	2000
balloon	1 [*3]	4002
burstin	1	50000
chaotic	1	1800

cstr	1 [*3]	22500
earthquake	1	4096
eeg	1 [*3]	10752
evaporator	1 [*3]	37830
foetal_ecg	1 [*3]	20000
glassfurnace	1 [*3]	11223
greatlakes	1 [*3]	4920
infrasound_beamd	1	8192
memory	1	6875
muscle_activation	1	29900
network	1	18000
ocean	1	4096
ocean_shear	1	4096
packet	1	360000
phdata	1 [*3]	6003
power_data	1	35040
powerplant	1	2400
shuttle	1 [*3]	6000
spot_exrates	1 [*3]	30792
standardandpoor500	1	17610
steamgen	1 [*3]	38400
synthetic_control	1 [*3]	36000
tide	1	8746
wind	1 [*3]	78888
winding	1 [*3]	22500

[*1] rand()*StreamLength was used to determine the starting location of each pattern from the stream data. ‘X’ number of data points following rand()*StreamLength were extracted as Pattern, whereas ‘X’ was set to any one of 50, 100, 250 and 500.

[*2] Multiple time series of different types in the data set containing were concatenated to form one stream of each type. For example the Sign Language data set contained 10 different types of time series. Each type contained 20 time series of 30 data points with 4 dimensions, concatenating the dimensions and multiple time series resulted in 20x30x4=2400 data points in each of the 10 streams.

[*3] Multiple time series of the same type in the data set were concatenated to make a single stream. As an example “Physiological_data_B1” contained 3 time series of 17000 data points each, which were concatenated to make a single stream of size 51000 data points.

[*4] The number of data points varies for different streams in the data set.