PROFILING INTERNET USERS

Information Security and Privacy

VAMSHI SAGAR GADDE

U62820763

Objective: The main objective of this project is to check if the internet usage of a user is statistically indistinguishable when user's usage is compared with the internet usage of his and also with other users in a time period. Also, to display how profiling affects the problem.

Source Data : The source of data for this project is Cisco NetFlow version 5 i.e. the ip traffic of 54 users over a time period which includes data like octets, real first packet, and duration.

Approach:

Language Used is Python Language.

I started with a for loop which gives all the possible combinations between all files. In the for loop I have passed each file to the functions **userweek1avg(), userweek2avg()**.

These functions will remove the unwanted data i.e. the rows with duration values equal to zero, the rows which are not in between 8 am – 5 pm(converting epoch time to local time). Within this loop I will call another function **findaverage()** which will take the useful data and calculate the average values between the given window(i. e. 10 or 227 or 5 minutes).

userweek1avg() and userweek2avg():

These functions will take **sheet** of the excel file as argument and returns all the averages between given time window.

This function first converts the epoch time into local time using the datetime module i.e datetime.datetime.fromlocaltimestamp(epoch time). From this I will extract the date, hours, seconds etc. and store into variables. Now, I will extract the duration values by using the function sheet.cellvalue(row, column) and if it is zero then I will skip the row, else, I will now check if the time from epoch is between 8 am and 5pm else, I will skip this row also. If the time is between 8 and 5. I will check the date and will store the doctets/duration values into one list and times into another list. Now, I will pass the date list and doctets/duration list to another function **findaverage()** to calculate the averages for weeks. The same I will do for the second week also.

findaverage():

In this function first I will check whether the length of the passed list is zero. If it is zero then I will put zeroes for all the averages in a day in the given time window. If the length is not zero then, I

will check how many packets from the given list are between the given time window intervals and I will calculate the averages and store it into a list. If there is no single packet between the time interval then I will put average value to zero for that interval. Now, I will return the list with averages.

With the averages for the week1 and week2 for both the users, I will now calculate the spearman correlation coefficients by calling the function scipy.stats.spearmanr(list1, list2)

The correlation values are r1a2a(correlation value between average values of week1 and 2 of user a), r1a2b (correlation value between average values of week1 of user a and week 2 of user b) and r2a2b (correlation value between average values of week2 of user a and week2 of user b).

Now, the calculated values are used in the function to get the "z" values.

Now, using the "z" values calculated above, we can calculate the "p" values using the below formula.

Code Execution: There are three files which are p_10.py, p_227.py, p_300.py which are used to generate the 10 secs window, 227 secs window, 300 secs window respectively. In the files the directory path should be changed according to the os path convention. I have used pycharm to execute the files. We can just install the python latest version and can run the program on the interactive shell using the command "exec(open("./filename").read()".

Analysis:

The figures below are the snapshots of the "p values" for the users in the windows 10secs, 227secs, 300secs.

Snapshot of 10 secs window

Weeks 1&	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
User 1	0.5	0.999779	0.999984	0.998341	0	0.815479	0.731971	0.999986	0.191524	0.996526
User 2	0.625796	0.5	0.599227	0.630904	0	0.620626	0.812004	0.458599	0.535882	0.503671
User 3	0.989288	0.999985	0.5	0.999991	0	0.999991	0.000512	1	0.999991	0.999991
User 4	0.684701	0.515984	0.715007	0.5	0	0.48061	0.843897	0.606231	0.565573	0.52054
User 5	1	1	1	1	0.5	1	1	1	1	1
User 6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 7	1	1	0.999991	0.999991	0	1	0.5	1	1	1
User 8	0.999991	0.999991	1	0.999991	0	0.999991	1	0.5	1	0.999991
User 9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 10	0.220838	0.06937	0.196855	0.636852	0	0.459525	0.956355	0.414254	0.786161	0.5
User 11	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 12	0.999982	0.999991	0.999991	0.999987	0	0.999984	0.999667	0.999991	0.99999	0.999991
User 13	0.999991	0.999989	0.933396	0.999874	0	0.999991	0.605623	0.999991	0.999991	0.99999
User 14	0.708465	1.59E-05	0.896831	0.24811	0	0.691798	0.457987	0.109421	0.430583	0.332913
User 15	0.999926	0.999986	0.999991	0.999884	0	0.832404	0.99949	0.999986	0.944194	0.999987
User 16	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 17	0.99999	0.99999	0.942357	0.99991	0	0.999991	0.215182	0.031893	0.999991	0.99996
User 18	-8.3E-06	0.058169	0.560583	0.220915	0	0.353692	0.072058	0.641967	0.082382	0.015235
User 19	0.999989	0.999991	0.999991	0.999991	0	0.899708	0.668521	0.999898	0.999991	0.999991
User 20	1	0.999991	0.999991	0.999991	0	0.998686	0.547191	0.999991	0.999991	0.999991
User 21	0.758583	0.366787	0.025554	0.538832	0	0.672668	0.992202	0.861107	0.148076	0.383139
User 22	0.000365	0.771763	0.47082	0.532635	0	0.038309	0.003085	0.687379	-1.1E-06	0.382193
User 23	0.84483	0.984275	0.992674	0.99417	0	7.97E-05	0.849214	0.707075	-9E-06	0.985228
User 24	0.504814	0.498167	0.027699	0.404421	0	0.754011	0.017249	0.831353	0.596394	0.517758
User 25	0.999991	0.999991	1	0.999991	0	0.000941	0.999991	0.426563	-9E-06	0.999991
User 26	1	1	1	1	0	0.999991	1	1	1	1
User 27	0.476395	0.453955	0.155007	0.661271	0	0.688566	0.068864	0.000166	0.899497	0.438623
User 28	0.999991	0.999991	0.999991	0.999991	0	0.999991	0.992203	0.999991	0.999991	0.999991
User 29	1	0.999991	0.999991	0.999991	0	1	1	1	0.999991	1
User 30	1	1	1	1	0	1	1	1	1	1

Snapshot of 227 secs window

Weeks 1&	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
User 1	0.5	0.999707	0.99997	0.948514	0	0.992566	0.948648	0.999991	0.632905	0.999943
User 2	0.712809	0.5	0.831499	0.853465	0	0.450075	0.863906	0.490482	0.516732	0.52842
User 3	0.996627	0.999517	0.5	0.999503	0	0.999433	0.904163	0.999947	0.999937	0.996302
User 4	0.437102	0.27621	0.328026	0.5	0	0.181379	0.718482	0.348881	0.297965	0.330476
User 5	1	1	1	1	0.5	1	1	1	1	1
User 6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 7	0.999991	0.999991	0.999991	0.999991	0	0.999991	0.5	0.999991	0.999991	0.999991
User 8	0.999842	0.978709	0.999991	0.999742	0	0.997522	0.999857	0.5	0.999963	0.997774
User 9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 10	0.016125	0.572694	0.72902	0.431309	0	0.807862	0.427334	0.155128	0.579473	0.5
User 11	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 12	0.999798	0.999991	0.999954	0.99999	0	0.99999	0.999991	0.999956	0.999802	0.999988
User 13	0.999908	0.996382	0.576554	0.978609	0	0.998276	0.547147	0.999935	0.99543	0.995269
User 14	0.847986	0.00771	0.992526	0.259181	0	0.427804	0.113182	0.271028	0.652505	0.199644
User 15	0.999968	0.999832	0.999991	0.999807	0	0.994256	0.95009	0.999991	0.979919	0.999869
User 16	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 17	0.997657	0.993209	0.967633	0.850577	0	0.818357	0.910115	0.498572	0.992084	0.955057
User 18	-9E-06	0.18909	0.72848	0.171428	0	0.173885	0.037951	0.150145	0.049881	0.06254
User 19	0.999968	0.999989	0.999991	0.999976	0	0.998402	0.994483	0.999977	0.999989	0.999952
User 20	0.999991	0.999991	0.999991	0.999991	0	0.999983	0.999852	0.999981	0.999984	0.999988
User 21	0.498756	0.350419	0.091655	0.207544	0	0.117302	0.754113	0.263462	0.011154	0.413108
User 22	0.005248	0.937876	0.987956	0.789394	0	0.782403	0.004306	0.637056	0.770551	0.669974
User 23	0.176659	0.697768	0.849397	0.372309	0	0.036045	0.329764	0.170514	7.02E-06	0.47357
User 24	0.432136	0.467055	0.205457	0.100129	0	0.420055	0.146024	0.014862	0.312211	0.404061
User 25	0.582091	0.164196	0.476408	0.084188	0	0.00063	0.110709	0.007797	-7.6E-06	0.45925
User 26	0.999919	0.999985	0.999961	0.98895	0	0.933805	0.999773	0.971725	0.992836	0.999985
User 27	0.443472	0.433144	0.392827	0.744162	0	0.450365	0.007232	0.131667	0.768565	0.431831
User 28	0.991193	0.997328	0.945914	0.996115	0	0.997815	0.528896	0.997978	0.99663	0.973305
User 29	0.999914	0.994243	0.982241	0.991099	0	0.998665	0.883948	0.996151	0.997269	0.998409
User 30	0.999842	0.999961	0.999088	0.999896	0	0.999756	0.998207	0.999989	0.99998	0.999986

Snapshot of 300 secs window

Weeks 1&	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
User 1	0.5	0.995239	0.999939	0.931654	0	0.990702	0.956544	0.999991	0.743413	0.999904
User 2	0.719224	0.5	0.743774	0.889962	0	0.477511	0.858825	0.381133	0.513344	0.191546
User 3	0.999625	0.999886	0.5	0.999852	0	0.999673	0.985332	0.999982	0.999984	0.999021
User 4	0.46019	0.350453	0.224034	0.5	0	0.130786	0.705678	0.391802	0.368867	0.413037
User 5	1	1	1	1	0.5	1	1	1	1	1
User 6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 7	0.999991	0.999991	0.999991	0.999989	0	0.999991	0.5	0.999991	0.999991	0.999991
User 8	0.994101	0.96735	0.999986	0.995351	0	0.956867	0.999451	0.5	0.998481	0.99165
User 9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 10	0.019912	0.332284	0.693684	0.387717	0	0.569748	0.285988	0.07762	0.44261	0.5
User 11	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 12	0.999788	0.999988	0.999383	0.999926	0	0.999955	0.999991	0.999668	0.998343	0.999945
User 13	0.998336	0.984179	0.375543	0.846529	0	0.992723	0.531296	0.999522	0.977807	0.941461
User 14	0.858026	0.094585	0.993586	0.271541	0	0.444014	0.125986	0.364162	0.681977	0.24268
User 15	0.999799	0.99988	0.999975	0.998916	0	0.97967	0.886774	0.999991	0.968173	0.999519
User 16	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
User 17	0.997249	0.995195	0.958645	0.737854	0	0.778191	0.955313	0.769844	0.984191	0.969536
User 18	-7.4E-06	0.298605	0.803345	0.212287	0	0.163205	0.036373	0.16451	0.085731	0.124131
User 19	0.999988	0.999988	0.999991	0.999984	0	0.999007	0.993173	0.999986	0.999984	0.999934
User 20	0.999982	0.999991	0.999978	0.99999	0	0.999922	0.996926	0.999784	0.999962	0.999963
User 21	0.725422	0.598526	0.135663	0.433425	0	0.322241	0.896198	0.473287	0.061468	0.657163
User 22	0.004276	0.927575	0.975247	0.637262	0	0.597296	0.001799	0.687245	0.65544	0.572873
User 23	0.259494	0.473001	0.544151	0.115896	0	0.131798	0.461909	0.115981	0.000445	0.451629
User 24	0.807706	0.6604	0.382863	0.178989	0	0.608296	0.283448	0.016032	0.594208	0.528125
User 25	0.496595	0.293923	0.424939	0.144028	0	0.017395	0.127534	0.034284	0.000149	0.439774
User 26	0.999906	0.999983	0.999901	0.968247	0	0.89803	0.999544	0.975122	0.98673	0.999979
User 27	0.508	0.67065	0.693369	0.726585	0	0.455979	0.075603	0.259644	0.918365	0.678829
User 28	0.934734	0.992483	0.895086	0.984932	0	0.985265	0.520089	0.941814	0.984725	0.909776
User 29	0.999861	0.967505	0.944912	0.969135	0	0.994594	0.886073	0.993639	0.996547	0.996636
User 30	0.999149	0.999681	0.993035	0.998423	0	0.999447	0.998242	0.999894	0.999918	0.999914

Result of Analysis:

In the **10** seconds window there are **2512**(i. e p value > 0.05) number of user combinations that are indistinguishable from each other while **404**(i. e. p value <= 0.05) number of combinations are distinguishable.

In the **227** seconds window there are **2577**(i. e p value > 0.05) number of user combinations that are indistinguishable from each other while **339**(i. e. p value <= 0.05) number of combinations are distinguishable.

In the **300** seconds window there are **2608**(i. e p value > 0.05) number of user combinations that are indistinguishable from each other while **308**(i. e. p value <= 0.05) number of combinations are distinguishable.

Conclusion:

We can see that the p values of 227 seconds is better than that of p values of 300 seconds window. But, 10 seconds window has better distinguishability than that of 227 seconds window. From the above analysis we can say that 10 seconds window has better distinguishability when compared to other windows of the users.