

Programming Assignment 3

Shivam Choudhary (sc3973)

1. Problem 1

Answer:

- (a)
- prog1: C/C++, the files contains libstdc++ headers along with libc.so linkages which hint towards an executable of C/C++ type
 - prog2: C/C++, it has linkages libstdc(++) and libgcc.so.6
 - prog3: C/C++, because it includes libstdc++.so.6 and has function calls like fseek, malloc
 - prog4: C, Its C because it has linkages like pthread hinting towards pthreads and references like GNU GCC
 - prog5: Java, because it has references like prog6.java
 - prog6: Python, Because it has linkages to modules like scapy,dport ,sport which means that it is python module with scapy as a library

- (b) Running the analysis for n=1 and s=1 we see that the top 19 hex values match for prog1 and prog 2 as an be seen in Figure 1. Furthermore prog 3 and prog2 have also top 16 values matching which hints that they are of the same type. The rest of the programs don't actually match and as can be corroborated with ssdeep as well. When running the output with different slides and n we can see that the number of ngrams for both the prog1 and prog2 become equal.

As can be seen from the figure several matches between prog3 and prog2 is also found hinting at strong level of correlation between the prog2 and prog1 both.

1 Static ngram analysis results for prog1	1 Static ngram analysis results for prog2	1 Static ngram analysis results for prog3	1 Static ngram analysis results for prog4	1 Static ngram analysis results for prog5	1 Static ngram analysis results for prog6
2 =====	2 =====	2 =====	2 =====	2 =====	2 =====
3 length_ngrams(n) = 1	3 length_ngrams(n) = 1	3 length_ngrams(n) = 1	3 length_ngrams(n) = 1	3 length_ngrams(n) = 1	3 length_ngrams(n) = 1
4 length_slide(s) = 1	4 length_slide(s) = 1	4 length_slide(s) = 1	4 length_slide(s) = 1	4 length_slide(s) = 1	4 length_slide(s) = 1
5 Only Top 20 are shown	5 Only Top 20 are shown	5 Only Top 20 are shown	5 Only Top 20 are shown	5 Only Top 20 are shown	5 Only Top 20 are shown
6 =====	6 =====	6 =====	6 =====	6 =====	6 =====
7 HEX_VALUE:COUNT	7 HEX_VALUE:COUNT	7 HEX_VALUE:COUNT	7 HEX_VALUE:COUNT	7 HEX_VALUE:COUNT	7 HEX_VALUE:COUNT
8 00:16631	8 00:16916	8 00:19211	8 00:515	8 00:202	8 00:1004
9 4b:3359	9 4b:2776	9 4b:2902	9 17:293	9 61:171	9 74:147
10 ff:3146	10 ff:2472	10 ff:2304	10 07:230	10 01:48	10 64:121
11 5f:1720	11 5f:1726	11 5f:1857	11 ff:194	11 2f:41	11 73:102
12 8b:1710	12 8b:1481	12 8b:1612	12 5f:186	12 6e:34	12 05:98
13 45:1572	13 45:1407	13 45:1439	13 01:181	13 74:34	13 01:89
14 53:1026	14 53:1031	14 74:1012	14 18:175	14 69:29	14 69:81
15 74:1019	15 74:1022	15 0b:841	15 20:174	15 72:29	15 02:63
16 0b:888	16 61:841	16 e8:826	16 08:169	16 6c:28	16 70:58
17 e8:868	17 e8:745	17 23:197	17 f1:163	17 01:26	17 61:57
18 61:850	18 8b:789	18 65:743	18 40:162	18 65:24	18 2f:56
19 49:702	19 49:709	19 61:684	19 03:159	19 6a:24	19 03:54
20 72:631	20 72:677	20 49:667	20 0b:154	20 76:22	20 72:52
21 c7:659	21 40:577	21 40:634	21 02:150	21 07:19	21 79:45
22 40:638	22 65:943	22 c7:615	22 0c:144	22 4c:16	22 28:44
23 60:610	23 60:535	23 69:082	23 ff:144	23 0e:14	23 6e:43
24 65:550	24 c7:528	24 72:589	24 2f:139	24 53:13	24 6c:42
25 69:537	25 31:400	25 6f:523	25 06:138	25 0e:13	25 52:41
26 31:476	26 63:405	26 63:489	26 04:137	26 6f:13	26 03:40
27 63:473	27 73:451	27 6e:455	27 40:137	27 28:12	27 04:34

Figure 1: n=1,slide=1

As we can see from the output of ssdeep that prog1 matches prog2 which gives us a way to further inspect the other values of n and s.

```
shivam@shivam-VirtualBox: ~/hmkw3
shivam@shivam-VirtualBox:~/hmkw3$ ssdeep -x prog1ss prog2ss prog3ss prog4ss
prog1ss:/home/shivam/hmkw3/prog1 matches prog2ss:/home/shivam/hmkw3/prog2 (69)
prog2ss:/home/shivam/hmkw3/prog2 matches prog1ss:/home/shivam/hmkw3/prog1 (69)
shivam@shivam-VirtualBox:~/hmkw3$
```

Figure 2: ssdeep output

Static ngram analysis results for prog1	Static ngram analysis results for prog2	Static ngram analysis results for prog3	Static ngram analysis results for prog4	Static ngram analysis results for prog5	Static ngram analysis results for prog6
1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:12375 9 fffff:1670 10 4889:1597 11 4889:619 12 ff48:759 13 c768:581 14 89c7:581 15 488d:532 16 8045:477 17 4080:439 18 5374:486 19 0048:401 20 4953:322 21 005f:310 22 8085:201 23 7249:288 24 5f58:278 25 4545:245 26 0080:236 27 0080:229	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:12384 9 4889:1386 10 fffff:1313 11 4889:670 12 ff48:550 13 c768:472 14 89c7:472 15 5374:486 16 4080:397 17 8045:377 18 488d:340 19 0048:325 20 4953:322 21 005f:318 22 7249:288 23 5f5a:271 24 4545:245 25 0080:241 26 0080:224 27 0080:218	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:14318 9 4889:1489 10 fffff:1231 11 4889:751 12 89c7:547 13 c768:546 14 8045:477 15 4880:475 16 0048:399 17 ff48:338 18 005f:313 19 5f72:265 20 8080:243 21 4545:244 22 0080:237 23 5374:224 24 4889:227 25 488d:220 26 0048:10 27 fffff:119	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:214 9 3c97:25 10 fffff:21 11 4d63:18 12 0280:17 13 346d:16 14 4080:15 15 d334:14 16 698a:14 17 ff17:13 18 973c:13 19 0002:13 20 fffff:12 21 a609:12 22 830c:12 23 2083:12 24 1414:11 25 4889:10 26 0048:10 27 fffff:9	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0100:45 9 0000:35 10 7661:22 11 6a67:22 12 6a61:21 13 6176:21 14 617f:20 15 0780:15 16 616e:15 17 672f:14 18 696a:14 19 6a61:14 20 2f6c:14 21 0c80:12 22 4c6a:11 23 7269:11 24 5374:10 25 7472:10 26 0a80:10 27 0081:9	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:535 9 0064:86 10 0280:52 11 0380:52 12 0074:49 13 0100:46 14 0073:41 15 0480:31 16 0052:27 17 7079:27 18 0083:25 19 0580:23 20 007c:21 21 0080:20 22 2080:20 23 0060:20 24 0780:19 25 0073:18 26 0065:18 27 6a74:17

Figure 3: n=2,slide=1

Static ngram analysis results for prog1	Static ngram analysis results for prog2	Static ngram analysis results for prog3	Static ngram analysis results for prog4	Static ngram analysis results for prog5	Static ngram analysis results for prog6
1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 2 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:6586 9 fffff:1807 10 4889:819 11 488d:459 12 4080:376 13 ff48:371 14 c768:307 15 89c7:274 16 488d:274 17 5374:216 18 0048:210 19 0b45:203 20 0048:105 21 2200:191 22 0080:185 23 0c10:166 24 0041:165 25 0002:164 26 158d:164 27 1c00:160	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 2 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:6717 9 fffff:1837 10 4889:734 11 488d:459 12 4080:362 13 ff48:262 14 c768:253 15 89c7:219 16 0080:215 17 5374:201 18 2200:191 19 0080:181 20 0c10:175 21 0e10:167 22 0041:166 23 0002:165 24 488d:165 25 138e:165 26 1c00:162 27 4953:162	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 2 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:7738 9 fffff:1801 10 4889:735 11 488d:458 12 4080:482 13 c768:276 14 89c7:271 15 0000:217 16 0048:214 17 0c45:207 18 2200:196 19 0080:185 20 0c10:176 21 0041:176 22 0042:174 23 438d:174 24 ff48:167 25 1c00:166 26 005f:162 27 5548:151	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 2 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:117 9 3c97:24 10 0280:12 11 0176:16 12 0100:18 13 fffff:9 14 698a:8 15 2083:8 16 ff17:7 17 4080:7 18 0572:7 19 0787:7 20 d334:7 21 0100:7 22 0110:6 23 0017:6 24 830c:6 25 346d:6 26 c12a:5 27 4883:5	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 2 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0100:27 9 0080:17 10 6176:16 11 617f:15 12 6a67:13 13 0780:12 14 6c61:11 15 0960:10 16 4c6a:8 17 2123:6 18 7472:6 19 0a80:5 20 0015:5 21 7269:5 22 0001:5 23 0002:5 24 0000:5 25 6a61:5 26 7073:5 27 5374:4	1 ===== 2 3 length_ngrams(n) = 2 4 length_slide(s) = 2 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 0000:271 9 0064:44 10 0280:28 11 0380:24 12 0100:22 13 0074:22 14 0083:18 15 0480:17 16 0052:16 17 0073:15 18 007c:12 19 0580:13 20 007f:12 21 7079:12 22 007c:11 23 2080:11 24 0080:11 25 0082:10 26 0080:10 27 0780:9

Figure 4: n=2,slide=2

Static ngram analysis results for prog1	Static ngram analysis results for prog2	Static ngram analysis results for prog3	Static ngram analysis results for prog4	Static ngram analysis results for prog5	Static ngram analysis results for prog6
1 ===== 2 3 length_ngrams(n) = 3 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 000000:9485 9 fffff:1734 10 89c768:581 11 4889c7:577 12 488d65:470 13 ff4889:333 14 408080:328 15 488085:201 16 000048:277 17 005f5a:268 18 fffff:240 19 ff488d:234 20 008000:226 21 fffff:221 22 000000:208 23 004889:206 24 5f54a6:204 25 000022:188 26 002280:187 27 220000:183	1 ===== 2 3 length_ngrams(n) = 3 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 000000:9967 9 fffff:1840 10 89c768:472 11 4889c7:468 12 488d65:368 13 408000:329 14 005f5a:269 15 000048:242 16 ff4889:240 17 008000:222 18 000000:210 19 5f54a6:204 20 000022:195 21 000005:193 22 000022:188 23 002280:187 24 220000:183 25 537434:176 26 746f72:173 27 607249:171	1 ===== 2 3 length_ngrams(n) = 3 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 000000:11807 9 89c768:546 10 4889c7:546 11 488d65:456 12 408000:344 13 fffff:338 14 005f5a:262 15 000048:223 16 008000:235 17 008000:217 18 746f72:210 19 000022:193 20 004889:194 21 fffff:183 22 002280:192 23 220000:187 24 1c0000:187 25 5f54a6:187 26 4889c7:178 27 8c7083:177	1 ===== 2 3 length_ngrams(n) = 3 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 000000:136 9 3c973c:13 10 000000:20 11 000200:8 12 d334d8:8 13 000002:7 14 555858:7 15 004080:7 16 4d0334:7 17 a6999a:7 18 000000:6 19 344d03:7 20 020800:6 21 000000:6 22 010000:5 23 830c32:5 24 20830c:5 25 141414:5 26 699acc:4 27 699acc:4	1 ===== 2 3 length_ngrams(n) = 3 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 6a6176:21 9 000074:19 10 000000:20 11 76612f:20 12 6c616a:14 13 6a672f:14 14 612f6c:14 15 616e67:14 16 2f6c61:14 17 4c6a61:11 18 72696a:11 19 537472:10 20 742698:8 21 672458:8 22 000100:8 23 696e67:8 24 2f5747:7 25 3b0100:5 26 284c6a:5 27 000000:5	1 ===== 2 3 length_ngrams(n) = 3 4 length_slide(s) = 1 5 Only Top 20 are shown 6 ===== 7 HEX_VALUE:COUNT 8 000000:272 9 000073:41 10 000074:39 11 038000:28 12 000052:25 13 208000:19 14 000028:18 15 048000:17 16 000200:16 17 740380:14 18 060000:14 19 640200:13 20 740400:13 21 657733:13 22 000100:12 23 000402:12 24 020000:12 25 010000:12 26 707974:11 27 070000:11

Figure 5: n=3,slide=1

1 Static ngram analysis results for prog1	1 Static ngram analysis results for prog2	1 Static ngram analysis results for prog3	1 Static ngram analysis results for prog4	1 Static ngram analysis results for prog5	1 Static ngram analysis results for prog6
2 =====	2 =====	2 =====	2 =====	2 =====	2 =====
3 length_ngrams(n) = 3	3 length_ngrams(n) = 3	3 length_ngrams(n) = 3	3 length_ngrams(n) = 3	3 length_ngrams(n) = 3	3 length_ngrams(n) = 3
4 length_slide(s) = 3	4 length_slide(s) = 3	4 length_slide(s) = 3	4 length_slide(s) = 3	4 length_slide(s) = 3	4 length_slide(s) = 3
5 Only Top 20 are shown	5 Only Top 20 are shown	5 Only Top 20 are shown	5 Only Top 20 are shown	5 Only Top 20 are shown	5 Only Top 20 are shown
6 =====	6 =====	6 =====	6 =====	6 =====	6 =====
7 HEX VALUE:COUNT	7 HEX VALUE:COUNT	7 HEX VALUE:COUNT	7 HEX VALUE:COUNT	7 HEX VALUE:COUNT	7 HEX VALUE:COUNT
8 000000:3239	8 000000:3409	8 000000:3968	8 000000:43	8 6a6176:10	8 000000:96
9 ffff48:240	9 002200:187	9 89c7e8:212	9 3c973c:5	9 612f6c:8	9 000073:13
10 000000:207	10 ffff48:187	10 002200:191	10 020000:4	10 616e67:8	10 000074:12
11 4889c7:189	11 4889c7:169	11 4889c7:176	11 83eefc:4	11 76612f:7	11 000052:10
12 000022:187	12 89c7e8:149	12 488945:149	12 973c97:4	12 6c616e:5	12 030000:9
13 89c7e8:173	13 488945:123	13 ffff48:111	13 000000:4	13 2f5374:4	13 006402:8
14 488945:159	14 001200:112	14 000048:91	14 8b1e48:4	14 4c0a61:4	14 008302:8
15 ff4889:122	15 005f5a:83	15 005f5a:91	15 a6699a:4	15 3b0100:4	15 740300:7
16 488945:102	16 000048:82	16 000000:74	16 20839c:4	16 637651:4	16 002300:6
17 000048:91	17 5f5a4e:78	17 001200:69	17 6999a6:3	17 72699e:4	17 008400:6
18 005f5a:90	18 000000:77	18 004889:68	18 344d03:3	18 537472:4	18 280000:6
19 000000:77	19 ff4889:73	19 746f72:68	19 41ff43:3	19 000000:4	19 006403:5
20 5f5a4e:75	20 fdffff:63	20 ffffff:66	20 555058:3	20 010016:3	20 7c0200:5
21 004889:74	21 488945:62	21 024300:64	21 40d334:3	21 673b29:3	21 617963:5
22 feffff:72	22 746f72:61	22 488945:64	22 11d08a:3	22 743b29:2	22 000000:5
23 ffffff:71	23 800243:59	23 0e1086:64	23 000200:3	23 696e74:2	23 740700:5
24 ff4889:60	24 410e10:59	24 5f5a4e:64	24 6099a6:2	24 2f5072:2	24 640200:4
25 000012:65	25 488976:59	25 1c0000:59	25 ffffff:112	25 757469:2	25 000028:4
26 697249:61	26 ffffff:58	26 070800:59	26 322e01:2	26 4c6973:2	26 740800:4
27 347061:61	27 697249:58	27 000041:59	27 196490:2	27 646572:2	27 000000:4

Figure 6: n=3,slide=3