RELAZIONE PROGETTO GESTIONE DI RETI

Claudio Burrafato mat.520333
Email: c.burrafato@studenti.unipi.it

Il progetto consiste nel aggiungere ulteriori statistiche stampate dal tool di analisi ndpiReader contenuto in nDPI, dove vengono raccolti i nomi al dominio dei flussi:
* flow->ssh_tls.server_info
* flow->host_server_name
e vengono inseriti in una UT_hash_table, ordinandoli per numero di occorrenze, dove in questo caso le occorrenze indicano quante volte viene contattato tale dominio nell'analisi di un flusso, in modo che alla fine del programma possiamo stampare, in ordine decrescente, i nomi al dominio per numero di tali occorrenze (i top X nomi al dominio contattati). In questo modo appare più chiaro quali domini vengono contattati più spesso durante una comunicazione(i flussi che compongono tale comunicazione), analizzando i pacchetti .pcap catturati in precedenza da uno sniffer(wireshark ad esempio).
Riporto un esempio "tagliato" di output della versione modificata di ndpiReader:

```
./example/ndpiReader -t -i ./tests/pcap/1kxun.pcap -v4
[...]
        NOTE: as one flow can have multiple risks set, the sum of the
              last column can exceed the number of flows with risks.



pic.1kxun.com                           15
239.255.255.250:1900                    12
mangaweb.1kxun.mobi                     11
charming-pc                             7
hkbn.content.1kxun.com                  6
???????????                             6
joanna-pc                               5
kankan.1kxun.com                        5
isatap                                  5
kevin-pc                                5
jp.kankan.1kxun.mobi                    4
hybird.rayjump.com                      4
caesar-thinkpad                         4
192.168.115.75                          4
setting.rayjump.com                     4
wangs-ltw                               4
ro_x1c                                  4
ws.1kxun.mobi                           3
jason-pc                                3
dl-obs.official.line.naver.jp           3
net.rayjump.com                         3
wpad                                    2
cdn.liftoff.io                          2
tw.api.vpon.com                         2
de01.rayjump.com                        2
sonusav                                 2
notebook                                2
sanji-lifebook-                         2
analytics.rayjump.com                   2
usher-pc                                2
vv.video.qq.com                         2
218.244.135.170                         1
play.google.com                         1
```

```
gfile                                   1
qzonestyle.gtimg.cn                     1
_googlecast._tcp.local                  1
```

Per la consegna del codice ho effettuato una pull request https://github.com/ntop/nDPI/pull/1587 direttamente nel github del progetto open source https://github.com/ntop/nDPI

Per far questo, quindi, ho innanzitutto creato una struttura dati che ho chiamato hash_stats; dove vengono usati tre campi: un char* per il nome del dominio usato come chiave della tabella, un int per l'occorenza usato come valore della tabella e l'handle per la UT_hash_table(il campo hh chiamato così per utilizzare le macro della libreria uthash).
Dopodiché, nella funzione help, ho aggiunto alla riga 475 tra le opzioni suggerite per l'argomento v, il 4. Questo perché ho scelto di usare un opzione a parte per vedere, in maniera diretta, cosa veniva stampato dal mio codice.
Fatto questo, nella porzione di codice prima di printFlowsStats(), ho aggiunto i codici di due funzioni(hash_stats_to_order e hash_stats_to_print) da utilizzare nelle macro della UT_hash,HASH_SORT, per ordinare la tabella in ordine crescente, nel caso della hash_stats_to_order, in modo che le cancellazioni da fare per l'harvesting fossero fatte per i nomi con occorrenza minore; e per ordinare la tabella in ordine decrescente per stampare i nomi con più occorrenze, fino ai nomi con meno occorrenze .
In fine ho modificato la funzione printFlowsStats() aggiungendo il caso in cui verbose fosse==4. In questo blocco viene costruita la hash table in cui inserire le statistiche da stampare.
La struct hostsHashT è il puntatore principale della hash table, host_iter viene usato per i vari cicli della macro HASH_ITER, dove vengono memorizzate le struct a mano mano visitate dal ciclo; inoltre nei vari HASH_ITER viene usato il campo della struct hh come viene suggerito dalla guida sulle UT_HASH. Infine ad ogni ITER vengono usati delle struct temporane, chiamate in questo caso tmp e seguite dai vari numeri.
Per riempire la table, viene prima controllato se il dominio è già una chiave presente, e nel caso contrario viene inserita usando la macro HASH_ADD_KEYPTR (viene usata questa in quanto il campo è un char*). Se invece la chiave è già presente, viene incrementato il rispettivo valore che rappresenta il numero di occorenze nel flusso.
Nel caso in cui viene inserita una chiave, si controlla se la lunghezza della table non superi un certo valore(la variabile len_table_max); nel caso in cui questo valore venga superato si cancellano un tot di elementi(la variabile toDelete) per liberare spazio nella table(harvesting).
Dopo gli inserimenti, con HASH_SORT viene ordinata la table utilizzando la funzione definita prima.
Nelle righe successive viene stampato il contenuto della tabella. Nelle righe 2774-2776 ho scritto questo codice in modo da stampare la colonna delle occorrenze in maniera allineata, in quanto mi sono accorto che usando \t della printf, alcuni numeri non venivano stampati allineati.
Nelle ultime righe viene, infine, cancellata la tabella e vengono deallocate le struct che la compongono.
Nel terminale, per compilare ed eseguire ndpiReader, ho usato questa istruzione: ./nDPI/example/ndpiReader -i ../tests/pcap/XXXXX.pcap -v 4. I file .pcap usati per i test sono quelli della cartella tests di nDPI.