Blackmarket Forum Analysis

Bruno, Daniel, Lucca

April 2019

1 Introduction

Motivation To understand the life-cycle of vulnerabilities it is instrumental to figure what users are discussing in the dark web (forums) about CVEs and so on. Since we have access to CrimeBB dataset, could be founded some main agent between different forums in order to found how are spread cybercrime tools ?

Gap in prior art There is no prior work studying the reputation of users across forums

Goals we want to answer the following questions

1. what are the externalities with respect to reputation? Knowning the reputation in one forum can we learn the reputation in another forum?
2. who are the bridges? Who are the users that send information across forums?
3. are there relationships of cooperation and punishment with respect to users that post spam or that sell things that have no value?
4. are there correlations at all across forums?

Challenges how to detect if two users are the same? One possibility is to see if they posted the same material, or to see if they have the same sentiment profile

2 Describing Data of Reputation

In CrimeBB each forum is a different database, and each user have a column in users database called "Reputation". This is a pear to pear score built when user interact and evaluate another user interaction. Each forum has its own way to do it, and some of them not measure it. Because of it each forum data have a very different distribution of reputation. in the following image the description of the data:

Forum	antichat	garage4hackers	greysec	hackforums	kernelmode	mpgh	offensivecommunity	raidforums	safeskyhacks	stresserforums
count	77,806.00	872.00	438.00	588,411.00	1,441.00	476,073.00	10,714.00	43,834.00	7,350.00	764.00
mean	1.35	0.00	0.86	2.41	4.35	11.17	0.03	3.21	0.00	0.58
std	23.97	0.00	3.05	29.85	24.60	36.31	1.57	58.02	0.00	3.77
min	-97.00	0.00	-3.00	-589.00	-20.00	-632.00	-1.00	-963.00	0.00	-11.00
25%	0.00	0.00	0.00	0.00	0.00	10.00	0.00	0.00	0.00	0.00
50%	0.00	0.00	0.00	0.00	0.00	10.00	0.00	0.00	0.00	0.00
75%	0.00	0.00	0.00	0.00	0.00	10.00	0.00	0.00	0.00	0.00
max	1,555.00	0.00	45.00	2,527.00	571.00	5,453.00	158.00	2,681.00	0.00	60.00

3 Methodology

3.1 Excluding zeros

Locking at the previous description we can see that Garage4Hackes and RaidForums do not measure reputation, Looking at 25%, 25% and 75% we can see zeros in most of the forums, this symbolize users that had not made interactions that had not been evaluated are the major part, and the same happen with the value 10 for MPGH. Excluding these user that have not change this scores yet, the data looks like:

Forum	antichat	garage4hackers	greysec	hackforums	kernelmode	mpgh	offensivecommunity	raidforums	safeskyhacks	stresserforums
count	4,783.00	0.00	108.00	119,363.00	293.00	11,411.00	106.00	2,275.00	0.00	84.00
mean	21.96	nan	3.47	11.89	21.39	58.82	3.48	61.94	nan	5.26
std	94.30	nan	5.37	65.41	51.17	229.51	15.42	247.50	nan	10.29
min	-97.00	nan	-3.00	-589.00	-20.00	-632.00	-1.00	-963.00	nan	-11.00
25%	-1.00	nan	1.00	-3.00	2.00	11.00	1.00	-3.00	nan	0.50
50%	2.00	nan	2.00	2.00	5.00	17.00	1.00	3.00	nan	3.00
75%	9.00	nan	4.00	6.00	19.00	34.00	2.00	50.00	nan	6.00
max	1,555.00	nan	45.00	2,527.00	571.00	5,453.00	158.00	2,681.00	nan	60.00

3.2 Standard Score

Each forum data is not comparable with another cause they have different scales. In order to make it comparable, we will use standard score when we intent to compare reputations in different forums. Data describe after Standard Score:

Forum	antichat	garage4hackers	greysec	hackforums	kernelmode	mpgh	offensivecommunity	raidforums	safeskyhacks	stresserforums
count	4,783.00	0.00	108.00	119,363.00	293.00	11,411.00	106.00	2,275.00	0.00	84.00
mean	0.00	nan	-0.00	0.00	0.00	0.00	-0.00	0.00	nan	-0.00
std	1.00	nan	1.00	1.00	1.00	1.00	1.00	1.00	nan	1.00
min	-1.26	nan	-1.21	-9.19	-0.81	-3.01	-0.29	-4.14	nan	-1.58
25%	-0.24	nan	-0.46	-0.23	-0.38	-0.21	-0.16	-0.26	nan	-0.46
50%	-0.21	nan	-0.27	-0.15	-0.32	-0.18	-0.16	-0.24	nan	-0.22
75%	-0.14	nan	0.10	-0.09	-0.05	-0.11	-0.10	-0.05	nan	0.07
max	16.26	nan	7.74	38.45	10.74	23.50	10.02	10.58	nan	5.32

3.3 Criterion to find same users in different forums

For this research we needs to assume that a user is the same in two or more forums when he has the same nickname in these forums. That is not a good way to find a bridge user, although we are doing it at this stage of the research.

3.4 Users in common between forums

Using the previous Criterion now we can classify uses in common in these forums. The following table is displaying how many users are in common in each couple of forums

	antichat	garage4hackers	greysec	hackforums	kernelmode	mpgh	offensivecommunity	raidforums	safeskyhacks	stresserforums
antichat	4783	0	1	88	2	16	0	10	0	0
garage4hackers	0	0	0	0	0	0	0	0	0	0
greysec	1	0	108	16	0	4	0	3	0	0
hackforums	88	0	16	119363	9	548	8	147	0	11
kernelmode	2	0	0	9	293	3	0	1	0	0
mpgh	16	0	4	548	3	11411	1	22	0	0
offensivecommunity	0	0	0	8	0	1	106	1	0	0
raidforums	10	0	3	147	1	22	1	2275	0	3
safeskyhacks	0	0	0	0	0	0	0	0	0	0
stresserforums	0	0	0	11	0	0	0	3	0	84

3.5 Pearson Correlation

Using Pearson Correlation, we can not see any important correlation in any couple of forums that have a substantial set of in common users

Forum	antichat	garage4hackers	greysec	hackforums	kernelmode	mpgh	offensivecommunity	raidforums	safeskyhacks	stresserforums
Forum										
antichat	1.00	nan	nan	-0.03	1.00	0.23	nan	-0.21	nan	nan
garage4hackers	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
greysec	nan	nan	1.00	-0.12	nan	-0.51	nan	0.42	nan	nan
hackforums	-0.03	nan	-0.12	1.00	-0.18	0.04	-0.55	-0.07	nan	0.16
kernelmode	1.00	nan	nan	-0.18	1.00	-0.90	nan	nan	nan	nan
mpgh	0.23	nan	-0.51	0.04	-0.90	1.00	nan	0.10	nan	nan
offensivecommunity	nan	nan	nan	-0.55	nan	nan	1.00	nan	nan	nan
raidforums	-0.21	nan	0.42	-0.07	nan	0.10	nan	1.00	nan	-0.97
safeskyhacks	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
stresserforums	nan	nan	nan	0.16	nan	nan	nan	-0.97	nan	1.00

3.6 Restricting the Scope

3.7 Extracting User with a good reputation between Hackforums and MPGH

4 Preliminary results

present linear results

References