Understanding the Twitter user networks of Viruses and Ransomware attacks

Michelangelo Puliga[1,2,*], Guido Caldarelli[1,2,3,†], Alessandro Chessa[1,2,‡], and Rocco De Nicola[1,§]

[1] Scuola IMT Alti Studi Lucca, Piazza San Francesco 19 55100, Italy michelangelo.puliga@imtlucca.it
[2] Laboratorio Linkalab, Cagliari, Italy
[3] London Institute for Mathematical Sciences, 35a South St. Mayfair London UK

Abstract

We study the networks of Twitter users posting information about Ransomware and Virus and other malware since 2010. We collected more than 200k tweets about 25 attacks measuring the impact of these outbreaks on the social network. We used the mention network as paradigm of network analysis showing that the networks have a similar behavior in terms of topology and tweet/retweet volumes. A detailed analysis on the data allowed us to better understand the role of the major technical web sites in diffusing the news of each new epidemic, while a study of the social media response reveal how this one is strictly correlated with the media hype but it is not directly proportional to the virus/ransomware diffusion. In fact ransomware is perceived as a problem hundred times more relevant than worms or botnets. We investigated the hypothesis of *Early Warning signals* in Twitter of malware attacks showing that, despite the popularity of the platform and its large user base, the chances of identifying early warning signals are pretty low. Finally we study the most active users, their distribution and their tendency of discussing more attack and how in time the users switch from a topic to another. Investigating the quality of the information on Twitter about malware we saw a great quality and the possibility to use this information as automatic classification of new attacks.

Introduction

Since the early days of personal computers the diffusion of malicious code has triggered concern and attention. During the first years the majority of the malware was mostly related to the ”worm” or ”virus” categories and their diffusion used mostly physical devices (for instance the floppy disk) or basic networking services. Nowadays the attacks are mostly related to social engineering manipulation of the victims using psychological tricks to convince people to download and install malicious email attachments or online offensive code. The development of Ransomware tools with encryption strategies able to block the system and ask for a ransom is a relatively recent threat. The development of crypto-currencies such as the BitCoin gave to hackers and organized crime a new source of financing using the ransomware:[3] discuss the pros and cons of paying or not paying the ransom. Other authors[6] made an economic model taking into account the bargaining phase of the attack when a user can ask for a discount in the ransom he eventually have to pay. This will create a market that involves a limited possibility of negotiation for users and cyber criminals and some feedback loops that further develop the market. More recently new generations of Ransomware attacks are specializing in targeting companies, that are more vulnerable to the disruption caused by the encryption and lock strategy. Finally with the rise of the IoT devices connected to the network, minimally managed and with basic hardware, the threat is becoming alarming as several ransomware are able to infect also industrial plants

*Corresponding author. The author collected the data, conceived the experiment(s), analysed the results.
†The author conceived the experiment(s) and the reviewed the manuscript.
‡The author conceived the experiment(s) and the reviewed the manuscript.
§The author conceived the experiment(s) and the reviewed the manuscript.

relying on IoT devices. There are concerns (https://www.eset.com/us/business/resources/tech-briefs/rot-ransomware-of-things/) that the IoT devices can be transformed in RoT "Ransomware of Things" a new menace to industrial business, civil infrastructures potentially very lucrative for the criminal groups.

In this paper we study the networks of Twitter users discussing about 25 relevant and worldwide diffused Viruses (Trojan, Worms, Spywares) and Ransomware attacks, as well as a Botnet. Twitter microblogging platform is used for posting news and discussing trending topics. Press agencies, bloggers, professionals and users interact with short messages discussing popular subjects (mostly related to politics, sport and gossip) but also more specialized ones like coding and ICT security. In the space of 140 characters users post general info, create hashtags, mention other users and attach short urls pointing to news. Following and counting the news from links it is possible to estimate the prevalence of a source in the debate and the overall diffusion of a topic.

We have tracked the evolution of posts about malwares from 2010 till September 2017. For each malware, we created the mention networks of the users discussing about the attacks and have computed the main features of those networks from a topological perspective; we then estimated the social media impact and the driving force of the news sources (i.e. the relative importance of the specialized press in spreading the news about the outbreak).

We focused on questions related to the characteristics of the users tweeting about malware, trying to answer to questions like: who are the early reporters? are they mostly from specialized press or rather individuals acting independently? We compare their hashtags and the precision they use in describing the malware threat. Finally, we used temporal networks analysis to study the effectiveness of the networks in diffusing the information among other users. In particular we investigated whether the users discussing of an attack were also active commenting other malwares. This overlap is of fundamental importance to understand if there is a core of users that constantly report news about malware; these users will be likely a source of information about new unseen attacks and thus of early warnings.

In our paper we decided to leverage on Twitter to gather informations about each attack. Our goals is to estimate the impact in terms of overall discussion (number of tweets, users involved, hashtags created and so on), temporal patterns (attention peaks, user overlap among different attacks) and media diffusion (analysis of cited blogs and unique news in tweets). Users tweeting about malware do improve our knowledge of the infection, and help us in estimating the perceived importance of a topic and the level of awareness. In specific domains such as political debate and online voting the Twitter platform proved also to be a good forecast tool (see the case of American Idol in[2]) or simply a tool to understand the development of the events like the Chilean student movement's protest described in[5]. In a more theoretical way authors like[8] are studying the diffusion of rumors on Twitter using epidemiological models. The attention of the users on the social media platform remains low: in[7] the authors show how the majority of the interactions of each user are focusing on a small subsample of his/her social network. Indeed not all friends and followers are equally important for a user. Moreover the number of topics discussed daily in the platform is not large: the general tendency is commenting the daily news. In our perspective this limited attention of the users can partly explain why the topic of the cybersecurity in Twitter is far less popular than other threats such as terrorism: media generally ignore the topic and prefers other menaces. Monitoring Twitter to detect emergent information (topic trends) is an important research topic. We aknowledge the contribution to the general discussion to several early works such as *TwitterMonitor*[9], *enBlogue* from[1] and frameworks for real time tweet analysis[4]. We reused the concept of dynamic sliding time window from this last work in order to explore the volumes of tweets for Ransomware epidemics across time. The weekly interval is enough to cover the required time frame for volumes and hashtags. The resulting mention networks, however, are studied globally.

Methods

The selection of the attacks was done with the main criteria of ensuring a uniform temporal coverage: we wanted to study the evolution of the interest on malware since 2010 till present (mid Sep 2017). We do distinguish the attacks as worms, viruses and ransomware, as well as a botnet: Mirai. The complete list of the malware is reported in 1.

The data from Twitter were collected using the Search interface directly from the main website using scraping techniques in order to overcome the limitation of the Search API that limits the available results to the last 7 days of each search query. The scraping action allowed to recover 214463 tweets that after a filtering process become 206040 tweets related to the 25 attacks. The filter on queries and tweets we introduced is based on the BoW (Bag of Words) methodology i.e. searching and saving only the tweets that contain one or more words in a selected set. In the search query we specify terms closely related to the malware and variants of the malware name. For instance to study the WannaCry infection the word "WannaCry" is in the BoW along "WannaCrypt" (a variant of the original name) as well as the word "EternalBlue" that is the code name for the exploit used to infect the machines with the ransomware. To acquire this knowledge on the attacks we revised several news articles on specialized press and the referring security bulletins. As a general remark there are limited possibilities of collision when the name of an attack can be confused with a username: this is the case of "VBmania" an old virus that collides with a username calling himself "vbmania". In this case in the filtering procedure we removed the tweets having as author "VBmania".

Several user networks can be created using the Twitter data: a) the direct network of friendship and *followership* b) a user-to-user mention network c) a hashtag network d) a topic network.

The first approach based on a direct network of friendship/followership is possible when the dataset of tweets refers to a recent period of time and the number of users is not too large. In this case a query with the Twitter API is used to extract the connections of each user in the database. This approach can only be used to extract the network as it is today and it is not suitable for the attacks that took place before 2017: the social network around each user might have changed since then in ways that we cannot reconstruct.

The second approach of building a mention network is the one we followed for our analysis. It is the most robust method as mentions are likely to imply also a friendship or a follower relation. Notice however that the mention network does not imply existence of a reciprocal link, i.e. the mentioning act has a direction. In fact, there are users that cite other users and are never cited back. In our analysis this approach has the main advantage of enablin us to reconstruct the network as it was at the time of the attack without the need of using the actual social relationships that can be different from those at the time of the attacks.

The third approach, that requires building a hashtag network, uses the hashtag overlap in the tweet of two users to infer an undirected connection. If two users use the same set of hashtags when twitting about the same attack it is likely that they are connected or interested in the same topics. This relation is of semantic nature and can be explored only using large databases of tweets for several attacks that, currently, we do not have. Another disadvantage of this approach is that the search query procedure is focusing on a constant set of words (the bag of words) and it does not consider all possible hashtags related to the given attack. Indeed to build a network of hashtags we need a dataset of tweets containing an exhaustive set of hashtags about each attack and not only the original BoW. The feasibility of this approach is limited when the volumes of tweets are scarce.

The last approach is a generalization of the hashtag network model, it uses the broad concept of topic to establish a connection among two users that will be connected if, for a given malware, will have most of the topics in common. This methodology is even less accurate of the one of the hashtags as it relies on the ability of completing a list of the topics related to an attack and correctly classifying the

tweets per each topic.

Creating a network using the mentions is for Twitter the most straightfoward approach, in our case the selection of tweets is also taking into account the majority of the well known hashtags related to each infection. In fact to search and filter an attak we need a BoW considering all terms, also the hashtags. As a final result despite che choice of the mention network to analyze the malware, the presence of the hashtags ensures a better level of completeness in the data acquisition phase.

To build the mention networks we count the number of direct citations between two users as weights of the links. To guarantee that the weights remain in the interval 0-1, they are rescaled according to the simple Min-Max rescaler:

$$w_{ij} = \frac{w_{ij} - min(\mathbf{w})}{max(\mathbf{w}) - min(\mathbf{w})}$$

where w_{ij} is the number of citations between node i and node j (user i author of the tweets $1,2,...,m$ is citing user j in the tweets).

The analysis on the mention networks of the attacks aims at the simplest topological characteristics: edge density, average degree, and size of the network. Instead to better understand the existence of temporal patterns we computed the pairwise Jaccard index for each network looking at the nodes of each graph:

$$J_{ij} = \frac{|N_i \cap N_j|}{|N_i| + |N_j| - |N_i \cap N_j|} \tag{1}$$

where $|N_i|$ indicated the size of the network i, and $|N_i \cap N_j|$ is the number of common nodes of the pair i, j. The idea behind this index is computing the fraction of nodes that are common to two different networks. We want to understand if there is a common set of specialized users that constantly talk about the attacks and if this set change in time.

Results

Table 1. reports the basic statistics of the 25 Twitter user networks of mentions about viruses and ransomwares that we collected from the social media platform. From the data, we see that the largest attacks for tweets volume are the WannaCry and Petya ransomware. The latter is more interesting as it is related to a case of international espionage and political controversy as it was used to attack the Ukrainian power plants during the Ukrainian revolution in mid 2016 rumors that the infection reached the Chernobyl power plant were in the press in Jun 2017 [1]. At the bottom, in terms of scarce popularity there is the Alureon Virus that was present in Twitter with just 4 tweets during the first year of the infection [2]. The less popular attacks have less than 10 tweets in total, the most popular ones thousands in a single day during the peak. The size of the attacks is linked to the ransomware epidemics and in particular to the vastly popularized WannaCry malware. In this case the mainstream media reported the news and the hashtag "wannacry" became worldwide *trending topic* the 12th of May 2017 (see `https://trendogate.com/search/?trend=wannacry`). While the Petya ransomware became trending topic the 12th of April 2016 (see `https://trendogate.com/search/?trend=petya`) no other malware attack ever became trending topic. This simple indication confirms the perception of ransowares as the most disruptive attacks. We notice that the amount of the tweets exceed largely the real impact of those malware; despite

[1] `https://www.theverge.com/2017/7/2/15910826/nato-response-petya-attack-state-actor-russia-ukraine` and `https://www.theverge.com/2017/6/28/15888632/petya-goldeneye-ransomware-cyberattack-ukraine-russia`

[2] The number of collected tweets refer to a single year after the discovery of the attack: we discarded late references to the malware

the great attention on the media, the estimated number of computers infected by WannaCry is estimated in 400k (https://blog.barkly.com/wannacry-ransomware-statistics-2017), a number The most interesting and spectacular change we see in the data (Fig. 5) is linked to the popularity of the ransomware attacks. Petya and Wannacry became trending topics and their total impact on twitter is two order of magnitude larger of each previous case. Clearly the volumes of reconstructed discussion on Twitter can be influenced by precision of our analysis and by our ability of using the correct set of words for the Bag of Word model. Further investigations will be carried out to check whether the 2016 drop in Fig. 5 is real or due our choice of search keywords. smaller than the far less popular CryptoLocker (http://www.hongkiat.com/blog/famous-malicious-computer-viruses/) whose victims were around 500k. Finally, the attention to a botnet such as Mirai (https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet) is far behind the concern for the ransomware epidemics, in spite of the immense power of such tools. With Mirai, hundreds of thousands of infected zombie computers and enslaved IoT devices (that are particularly difficult to secure and patch) are able to shutdown large portions of the national Internet infrastructure with massive DDos attacks. Notice that all the techniques, botnets, viruses in form of network worms, and encryption payloads can be used together in sophisticate attacks. For instance a botnet can spread malicious emails containing a payload that infects a computer with a worm, able to further explore the network looking for potential new victims running non secured computers. The worm can eventually activate a crypto attack and ask for a ransomware, or can instead change its nature enslaving the computer to the original botnet. The network analysis in Tab. 2 focuses on basic network measures: we see that, despite the different sizes of the networks, if a graph has at least 100 nodes the average degree remains quite constant. Figure 1 confirms that the networks with at least 100 nodes have a strict proportionality of number of nodes and number of edges. This is a simple yet powerful indication of the network formation around those topics on Twitter: networks of citations in Twitter have by constraint an upper boundary in terms of degree; with only 140 characters the number of users that can be cited is not large. However it is still not obvious that a user is citing in his tweets no more than two other users for all the outbreaks. The results is a network that is almost fully connected even if the average degree is not so large.

attack	users	unique htags	relative tweets	total tweets
WannaCry	62660	28162	147009	147065
Petya	19745	9140	42680	43963
CryptoLocker	3335	1462	6498	10226
Mirai	2559	1498	5708	8178
Linux.Darlloz	385	86	937	1295
SpyEye	428	397	887	1208
TorrentLocker	213	111	295	380
Gameover_ZeuS	195	101	307	350
VBMania	209	72	305	305
Shamoon	172	109	219	230
Duqu	134	105	214	214
sKyWIper	91	36	170	170
Reveton	87	77	117	153
Tiny_Banker_Trojan	74	97	127	127
ZeroAccess	66	64	123	123
Stuxnet	50	44	116	118
BASHLITE	39	112	74	95
CryptoLocker.F	10	3	66	69
Xafecopy	46	44	57	57
Kenzero	19	28	49	54
WaleDac	37	26	54	54
NGRBot	7	5	20	20
Fusob	4	3	4	5
Alureon	2	3	4	4

Table 1: The table represents the main data about the Twitter dataset about of 25 viruses and ransomware attack popularized in Twitter since 2010

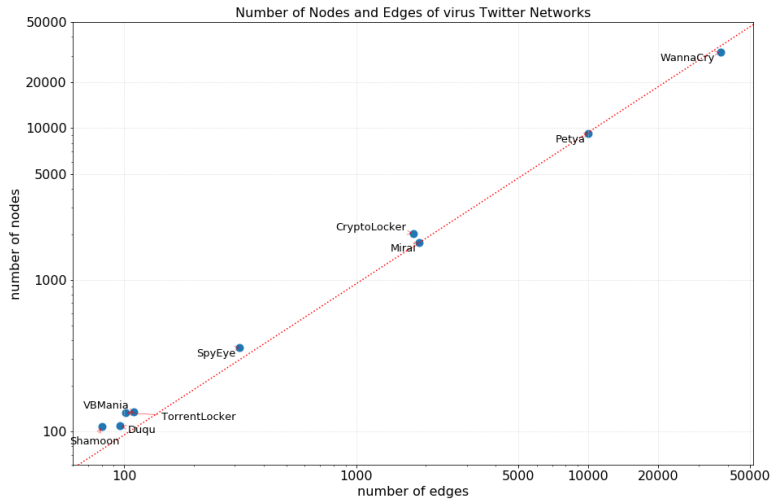

Figure 1: Nodes and Edges proportionality for the largest Viruses and Ransomware user mention networks as obtained from Twitter in a timespan of 7 years from 2010.

The temporal evolution of the attacks is shown in Fig. 2 where, for graphical purposes, the activity is normalized to the unity: for instance the WannaCry crisis was thousands of times more intense that the WaledDac infection. In general the activity is represented by a low count of tweets followed by an outbreak in the form of a peak (see Fig. 3). The users are commenting with low volumes till the point a hashtag becomes popular or the mainstream media start reporting about the infection. The more an attack was popularized the more the infection has sudden peaks of large size. For instance while Petya and WannaCry have large peaks, the Linux.Darlloz worm with less than one thousand tweets have a more flat and low rate of citations during the entire year of recording. Fig. 4 shows how the different attacks were covered on the media (as extracted from the tweets resolving the media url inside) and on Twitter in a similar way: there is a strict proportionality of the number of media news and number of tweets. This is true for all sizes: from the 20 unique tweets of the NGRBot to the 147k tweets of the WannaCry outbreak. The relatively constant average degree and the strict proportionality of number of tweets and media is something that we believe be embedded in the Twitter platform and its 140 character message. It is also interesting to notice that this behaviour remained almost unchanged for the last 7 years. The most interesting and spectacular change we see in the data (Fig. 5) is linked to the popularity of the ransomware attacks. Petya and Wannacry became trending topics and their total impact on twitter is two order of magnitude larger of each previous case. Clearly the volumes of reconstructed discussion on Twitter can be influenced by precision of our analysis and by our ability of using the correct set of words for the Bag of Word model. Further investigations will be carried out to check whether the 2016 drop in Fig. 5 is real or due our choice of search keywords.

attack	nodes	edges	avg_degree	edge_density	start	end
WannaCry	31854	37352	2.35	0.0001	2017-02-01	2017-09-01
Petya	9242	10036	2.17	0.0002	2017-02-01	2017-09-21
CryptoLocker	2019	1768	1.75	0.0009	2013-08-01	2014-08-01
Mirai	1761	1865	2.12	0.0012	2016-08-01	2017-08-01
SpyEye	358	314	1.75	0.0049	2011-01-01	2012-01-01
VBMania	134	110	1.64	0.0123	2010-08-01	2011-08-01
TorrentLocker	132	101	1.53	0.0117	2014-01-01	2015-01-01
Duqu	108	96	1.78	0.0166	2011-07-01	2012-07-01
Shamoon	107	80	1.50	0.0141	2012-07-01	2013-07-01
Gameover ZeuS	87	65	1.49	0.0174	2013-11-01	2014-11-01
Linux.Darlloz	83	57	1.37	0.0167	2013-11-01	2014-11-01
sKyWIper	74	72	1.95	0.0267	2012-04-01	2013-04-01
Reveton	60	44	1.47	0.0249	2012-01-01	2013-01-01
ZeroAccess	45	31	1.38	0.0313	2011-06-01	2012-06-01
Tiny Banker Trojan	42	31	1.48	0.0360	2016-01-01	2017-01-01
Stuxnet	41	34	1.66	0.0415	2010-05-01	2011-05-01
WaleDac	24	15	1.25	0.0543	2010-01-01	2011-01-01
Kenzero	19	21	2.21	0.1228	2010-08-01	2011-08-01
BASHLITE	17	10	1.18	0.0735	2015-01-01	2016-01-01
NGRBot	13	18	2.77	0.2308	2012-07-01	2013-07-01
Xafecopy	13	8	1.23	0.1026	2017-08-01	2017-09-21
CryptoLocker.F	10	0	0.00	0.0000	2013-08-01	2014-08-01
Fusob	7	5	1.43	0.2381	2015-03-01	2016-03-01
Alureon	2	0	0.00	0.0000	2010-01-01	2011-01-01
Regin Trojan	0	0	nan	nan	2014-10-01	2015-10-01

Table 2: The table the network measures of the Twitter user mention networks. The dates start/end represents the time period of data collection for the tweet dataset.

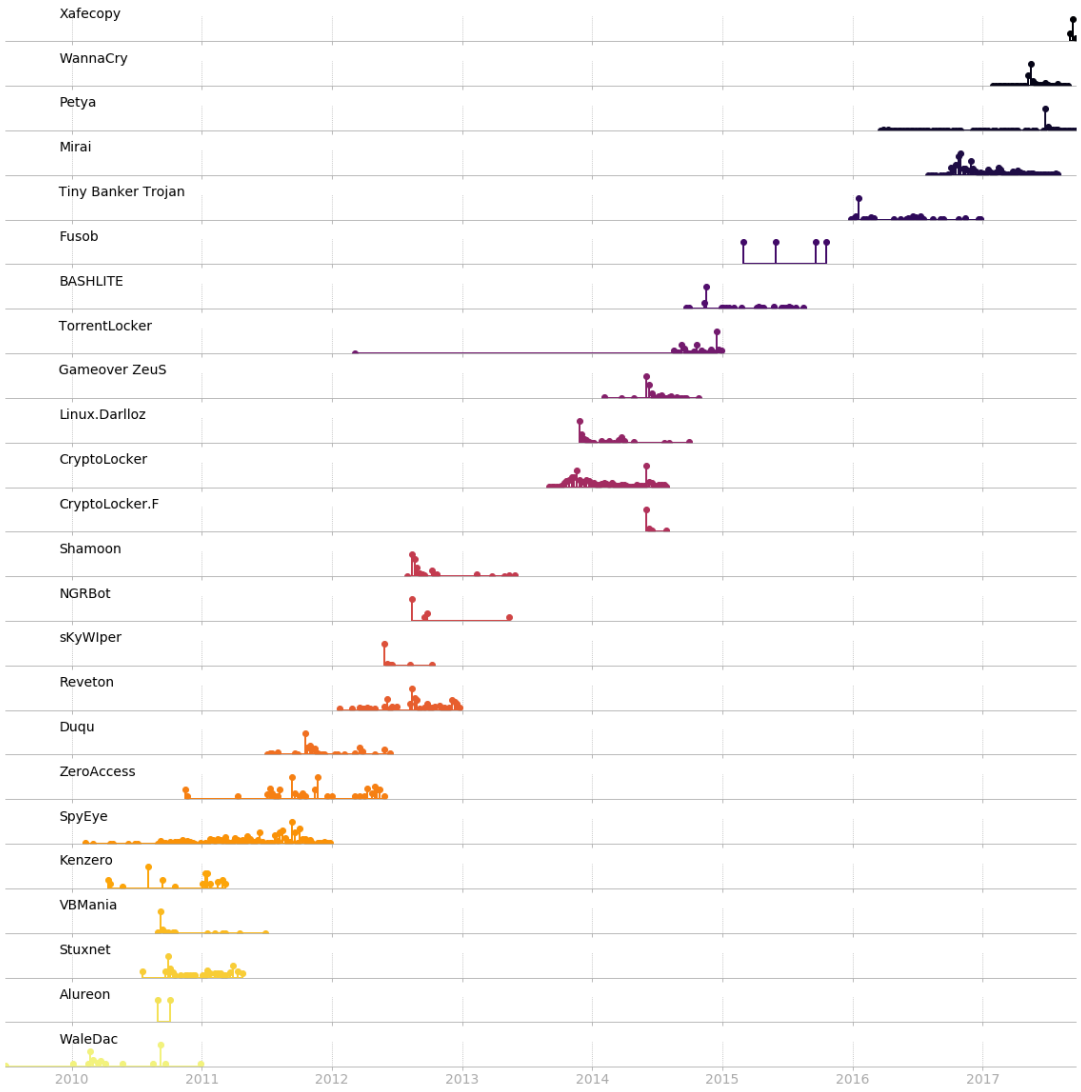

Figure 2: Timeline of Ransomware/Virus attacks from Twitter. The plot is normalized to the unity for each case. The scale of the WannaCry attack is thousands of time larger than the Alureon one as reported in Tab. 1. For the TorrentLocker timeline it is evident the early ”special” tweet of early 2012 (see main text for details).

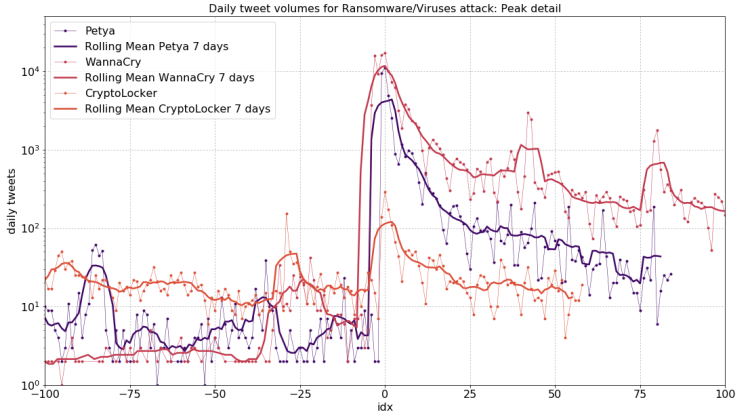

Figure 3: The daily tweet volumes of the three largest (by tweet number) malwares. In this case we artificially aligned the peaks to show how the burst and the following decay were similar for all three large and popular attacks.

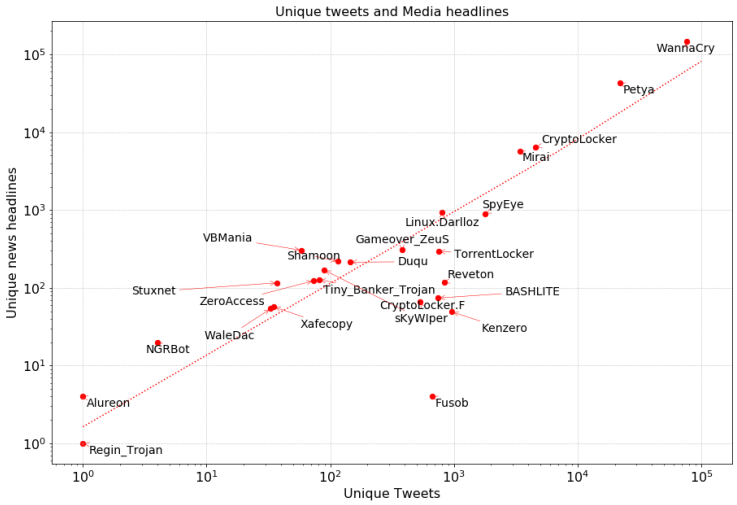

Figure 4: Media coverage and unique tweets of Virus/Ransomware attacks. The media addresses in the tweets are strictly proportional to the number of tweets collected for each attack: in face of the different sizes a constant fraction of users is citing media references for the incident. The proportionality is respected for all attacks having at least 10 tweets in the database.

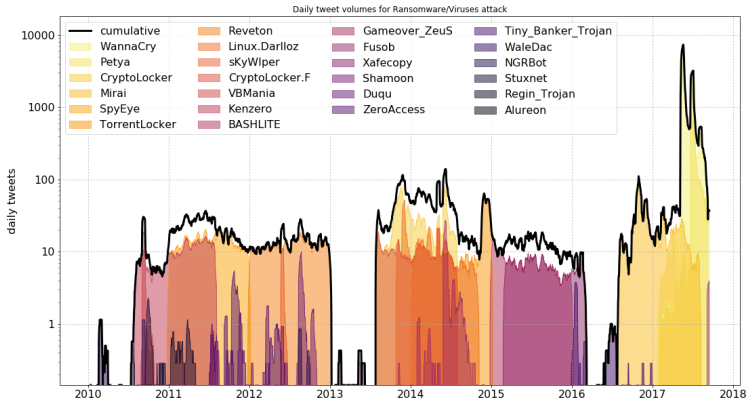

Figure 5: The daily tweet volumes of all malware attacks. The black line represent the sum of all the activities. Scale is semilog.

To better understand how an attack develops on Twitter, we can compare the peaks of three main incidents that we selected because of their large volumes. In Fig. 3 the large peak and following decay of WannaCry is shown with the tendency line (using a 7 days moving average to remove the weekly periodicity) in comparison with Petya, and CryptoLocker. In spite of the different size of the peaks, the plots show an identical behavior of burst and decay. The decay part is similar also in the time interval: for more than 100 days the topic is still very active on Twitter, while the rise of the peak is almost instantaneous.

Temporal analysis and media coverage

Applying the equation 1 to each pair of networks we verify the existence of a small cluster from Cryptolocker to TorrentLocker and Fusob: during the year 2014 and 2015 (see Fig. 6). This overlap of the user base (nodes of the networks) is mostly due to the contemporaneity of the attacks, but the same temporal patter of overlap is not always present in the data. For instance Petya and WannaCry are developing in the same period but their Jaccard index is much lower, i.e. their user bases tend to overlap less. Long range overlaps are less likely to appear: the fraction of users in common between the old SpyEye and WannaCry is not large compared to the almost contemporaneous infection Kenzero.

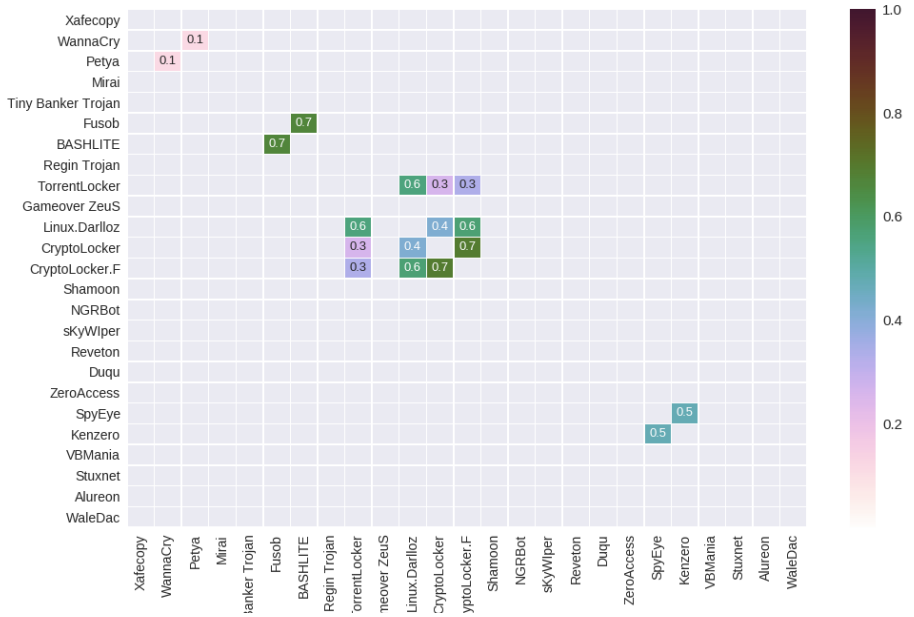

Figure 6: Pairwise Jaccard index of malware user bases (network nodes). The matrix plot shows which malware have an user base that overlaps significantly with the others. The more the color is intense the larger the overlap, the number inside the cell is the fraction of overlap (in the interval $(0,1)$) . The user base is equivalent to the set of nodes of each malware mentioning network. During 2014 the same users talking about CryptoLocker were discussing about the TorrentLocker, while the contemporaneity of WannaCry and Petya was not related to a strong level of overlap with only 0.1 (10%) of the users discussing both Petya and WannaCry. Other attacks have a negligible amount of overalapping.

To improve our understanding of each infection we studied two other social dimensions: a) the most popular media covering each attack b) the most popular hashtags in each dataset.

The media coverage analysis is a simple count of how many times a given source reported information about the incident. In particular we count the appearance of each baseline domain (such as nytimes.com or theguardian.com) in the entire sequence of tweets. Notice that in Twitter the majority of the links are shortened by services such as bit.ly and we have to use a software to *unshorten* each url. After many years several urls are broken and no more available from the original redirect service and also some shortening services have been shut down. Moreover a small part of urls were not correctly pasted in the original tweets. The sum of all those broken links is reported in Tab. 3 as "unresolved". We also performed a manual classification of the first 100 sites, distinguishing several categories: hacking blogs or tech magazines, general news, advertising sites, generic information or service websites, tools and software such as github. Other fine grain classifications can eventually be obtained by considering individual sites. We notice, for instance, that the majority of links are internals from twitter to twitter; i.e. they are the results of users citing other tweets (not a retweet) or adding links to facebook pages (there is also room to expand the analysis also in Facebook following those links). The large prevalence of hacking related magazines is a sign of the high technical skills of the users posting about malwares. Among others the hackernews.com magazine is the most relevant source of information on malwares on

Twitter. We noticed also several corporate web sites of companies working in security.

WannaCry	**Petya**	**CryptoLocker**	**Mirai**	**Linux.Darlloz**	**SpyEye**
#wannacry	#petya	#cryptolocker	#mirai	#linux	#spyeye
#ransomware	#ransomware	#ransomware	#botnet	#onlinevirus	#zeus
#cybersecurity	#wannacry	#malware	#iot	#infosec	#malware
#infosec	#notpetya	#security	#ddos	#security	#trojan
#petya	#cybersecurity	#virus	#cybersecurity	#worm	#android
#cyberattack	#cyberattack	#infosec	#infosec	#iot	#security
#security	#infosec	#gozeus	#malware	#symantec	#botnet
#eternalblue	#malware	#gameoverzeus	#security	#php	#infosec
#malware	#security	#zeus	#miraibotnet	#malware	#cybercrime
#wannacrypt	#goldeneye	#cybersecurity	#nanog69	#freeantivirus	#banking
Gameover_ZeuS	**TorrentLocker**	**VBMania**	**Shamoon**	**Duqu**	**sKyWIper**
#gameover	#torrentlocker	#hereyouhave	#shamoon	#duqu	#skywiper
#zeus	#ransomware	#virus	#virus	#stuxnet	#flamer
#cryptolocker	#malware	#vbmania	#malware	#worm	#flame
#cybercrime	#cryptolocker	#security	#windows	#security	#stuxnet
#securityaffairs	#bitcoin	#spam	#aramco	#twitter	#duqu
#malware	#decrypter	#worm	#security	#malware	#infosec
#security	#infosec	#email	#flame	#virus	#malware
#botnet	#security	#pgvirus	#saudi	#infosec	#theflame
#infosec	#securityaffairs	#viruses	#hacking	#iran	#cyberwar
#fbi	#eset	#networkworld	#it	#symantec	#iran
Tiny_Banker_Trojan	**ZeroAccess**	**Reveton**	**Stuxnet**	**BASHLITE**	**CryptoLocker.F**
#tinba	#zeroaccess	#reveton	#stuxnet	#bashlite	#symantec
#malware	#rootkit	#ransomware	#trojan	#shellshock	#latestthreats
#trojan	#malware	#malware	#security	#busybox	#threat
#tinbapore	#security	#fbi	#cyberwar	#wwwjoweeomicilcom	-
#cybersecurity	#tdl3	#virus	#cyber	#bash	-
#conficker	#tdl4	#citadel	#military	#malware	-
#infosec	#malwares	#infosec	#social	#infosec	-
#gozi	#infosec	#ransom	#cyberwarfare	#africa	-
#	#twitter	#security	#cybercrime	#security	-
#zeus	#analysis	#scareware	#scada	#caboverde	-
Xafecopy	**WaleDac**	**Kenzero**	**NGRBot**	**Fusob**	**Alureon**
#xafecopy	#waledac	#kenzero	#ngrbot	#fusob	#alureon
#trojan	#botnet	#yfmwomensday	#trojan	#hcacronyms	#jmu
#malware	#microsoft	#porn	#worm	#saynotodrilling	#chesapeakehall
#xafecopytrojan	#malware	#trojan	#malware	-	-
#india	#operationb49	#virus	#mcafeepic	-	-
#androidpic	#cybercrime	#japan	-	-	-
#mobile	#infosec	#hentai	-	-	-
#security	#spam	#japanese	-	-	-
#phone	#securitytrends	#zar	-	-	-
#cybersecurity	#stormworm	#gamers	-	-	-

Table 3: The table represents the most important hashtags associated to the Twitter networks of 25 viruses and ransomware attack popularized in Twitter since 2010

The study of the most frequent hashtags (see Tab. 4) reveals that Ransomware attacks such as WannaCry and Petya are often associated in tweets and hashtags. Another name for WannaCry is WannaCrypt an alternative term pointing to the encryption threat. Also the name ”eternalblue” that reports information about the actual exploit is in the WannaCry list. Mirai is associated with the powerful

botnet and the ddos that it can create. Linux.Darlloz is correctly seen as a worm targeting the linux platform. In general the classification of the each attack using their hashtags is pretty accurate. Each malware is recognized as virus, ransomware, botnet, worm and so on and all the synonym names are present like for wannacry. The hashtag list proved to be a good classification tool to understand the nature of each malware and providing a quick classification.

url	freq	type
twitter.com	18348	social network
unresolved	11229	unresolved
www.youtube.com	3797	social network
thehackernews.com	3263	hacking
www.facebook.com	2273	social network
www2.themsphub.com	1878	tools
www.bbc.co.uk	1289	news
threatpost.com	1284	hacking
www.zdnet.com	1209	hacking
www.symantec.com	1207	hacking
cm.gy	1072	news
www.theguardian.com	1032	news
www.infosecurity-magazine.com	1018	hacking
www.wired.com	952	news
www.bleepingcomputer.com	890	hacking
www.bbc.com	859	news
www.cnet.com	856	news
www.carbonblack.com	841	tools
or-argent.eu	823	advertising
www.linkedin.com	807	social network
securityintelligence.com	755	hacking
www.forbes.com	749	news
www.reuters.com	746	news
blogs.technet.microsoft.com	715	hacking
engage2demand.cisco.com	705	hacking

Table 4: The most important web sites and their category related to all the attacks ordered by frequency with at least 700 occurrences

The distribution of number of tweets per user and of number of retweets per each attack is reported in Fig. 8 and Fig. 7. Again the linearity in log-scale indicates the rarity of users tweeting thousands of tweets and the abundance of those with few tweets; retweets follow the same exponential logic. A bit of variability is present when the size of the attack is not large.

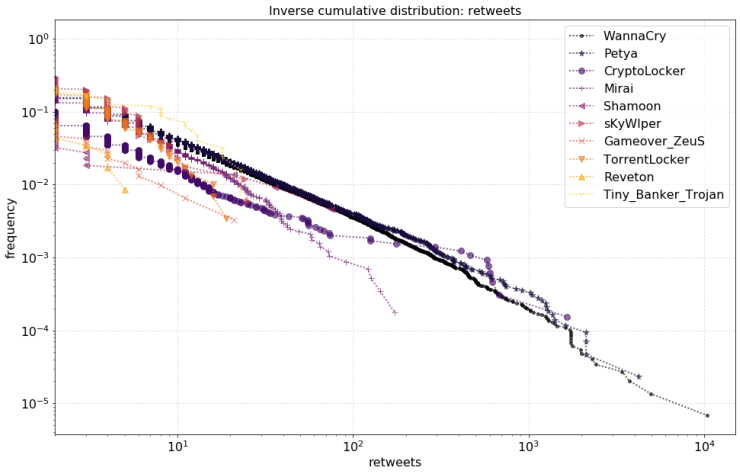

Figure 7: The distribution of the retweets per each attack. The plot is in log-log scale and shows an identical behavior for each malware: tweets with many retweets are exponentially rarer than those with just few retweets. From the plot we removed the attacks with less of 100 retweets.

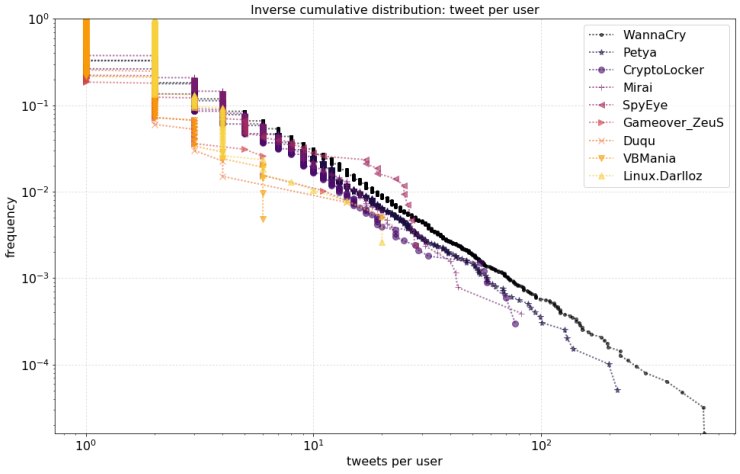

Figure 8: The distribution of the tweet numbers per user for each malware. The plot is in log-log scale and shows a similar behavior: users tweeting large numbers of tweets are much rarer than those twitting just one or two tweets in all the dataset. From the plot we removed the attacks with less of 100 tweets.

Early Warning signals

To collect data from the Twitter Search we specified a time interval and a set of keywords related to each RansomWare/Virus attacks. Each interval was selected according to the informations reported in the specialized press and the corresponding Wikipedia pages of each malware. We decided to further

explore the possibility of using Twitter to detect some early warning about some malware attack. Our aim has been to get the first tweet mentioning the threat and to compare it with the information obtained from specialized press.

The task is rendered more complex by the possibility of having some misleading hashtag or username confused with the malware word. For instance, for 2008 we found a user called "vbmania" having nothing in common with the virus. We also noticed that Twitter for the years around 2009-2011 is somehow less precise in the answers given back by the search. A small fraction of tweets was in those years totally unrelated with the query. For this reason we carefully filtered out the inaccurate answers using the keyword selected as Bag of Word. The BoW is created selecting hashtags (i.e #CryptoLocker) or direct citations (CryptoLocker) for each malware but filtering out user matches such as "@vbmania". The results of the early signal search made clear that Twitter is behind schedule in reporting the news about a new attack. We confirmed the accuracy of the time intervals for the majority of the malware, with the only exception of very old malwares (such as Kenzero and WaleDac) were the first tweet appeared several months after the official discovery. The interest on malware is clearly growing with time and has become more evident for the recent attacks of 2016, 2017 with ransomware. A notable exception that we report as anecdotal evidence of the anticipatory capabilities of Twitter is the "TorrentLocker" malware. Although being officially discovered in 2014 a single tweet using the hashtag TorrentLocker and correctly referring to a malware that exploits the BitTorrent platform appeared in 2012. We believe that this single tweet is a coincidence or a lucky name collision not a real anticipatory event. Here the text of the tweet:

> @thinksnews:
>
> Netjups: World's First Bitorrent / Cyberlock Hybrid - next attack vector or target for sopa ? zite.tow/wfhbjO #torrentlocker
>
> 08:33 - 10 mar 2012

Discussion and future research

In this paper we presented an analysis of 25 malware attacks in a period of 7 years, from 2010 to 2017 (present) as they appeared in the Twitter platform. We estimated the impact of each attack in terms of total tweets, size of the mentioning network. We analyzed temporal overlap and the usage of the most frequent hashtags as a semantic tool. We also investigated the possibility of some leak on Twitter exposing in advance information about cyber threats. We saw that the networks of users mentioning other users in their tweets on malware tend to form structures with nearly constant average degree. The topological properties of those graphs are interesting: an average of two users are mentioned by each user no matter the size of the network. We also report a strict proportionality from the total number of tweets and media coverage (as extracted from the tweets).

The role of the media is evident in the development of the tweet volumes: the preparation phase, the peak and the decay that characterize all major attacks. After 100 days from the outbreak of the infection there is still signal. The volume of tweets talking about the malware is a proxy for the perceived importance of an attack. The recent Wannacry and Petya incidents received a broad attention while others, like TorrentLocker, were barely present in the social media platform. The threat of the ransomware is by far the most scary for the users.

The tweets themselves can be used to extract more information about each malware: the most frequent hashtags are an easy way to classify each malicious code. This is an interesting feature of the tweets about malware. Usually, in fact, tweet content is noisy, the hashtags reported in the political debate can incorporate sarcasm and other jokes. From our analysis the hashtags about malicious code are

instead more informative and less emotive. In particular, from the short messages and their hashtags, the classification of each threat as worm, botnet, ransomware, virus and the targeting platform (windows7, linux, macos) is generally very clear. Finally the hashtags reveal other informations about the presence of synonyms and exploits (for instance EternalBlue and WannaCry) that can be missing from the initial Bag of Word used to collect the tweets.

From our analysis we can infer the overlap of the user communities discussing the different threats especially if the attacks are closer in time and typology. We observed that the family of 2015 cryptolockers were discussed from a common set of users, while the intersection of the users of Petya and Wannacry is smaller even if they are often associated in each tweet. In our study, we also considered the possibility that Twitter might offer some early warning signal about new threats. We explored the presence of tweets about each case and we can exclude anticipatory contents in Twitter. The discussions on the microblogging platform is mostly related to comment about the news as reported by the (specialized) press. In some case the news became viral and the hashtags trending topics for one or two days.

We foresee in the analysis of malware in Twitter two possible interesting directions. Judging the interest and the impact of each attack by the volumes, establishing for instance if the Ransomware is seen as a menace of greater intensity compared to Botnets and DDOS. Also the campaign of information about the risk of each malware can spread through the social networks and rise awareness for the threat.

Another usage is related to the semantic possibilities of these Twitter messages, from a simple screening of the hashtags associated to each attack, we were able to classify the attacks and to gather further details about the exploits, the target platforms, the operating systems. The usage of the user mention network can further increase the quality of the filtering: using the tweets from the most central nodes can eventually lead to gather better information about the threat. Other future directions of research are given by the explanation of network formation:

- are the professional users the drivers of the network growth?
- are the large press agencies with their user base those who triggers the reaction with a cascade of retweets?
- are the mainstream news those that start the reaction of an army of individual users with each of them posting a couple of tweets?
- are the tools of temporal networks of interest in studying the evolution of the network and its efficiency?
- why the networks are disappearing after a period of time?

In conclusion we believe that this unique study tracing the impact of malware in Twitter can give the experts in malware an estimation of the awareness on each cybercrime threat and make available to network scientists a clear case of temporal network formation.

Acknowledgements

G.C. acknowledges EU grant SoBiGData, CoeGSS and OpenMaker. RDN acknowledges the support of the FilieraSicura project.

Author contributions statement

All the authors conceived the experiment(s), M.P. conducted the experiment(s), M.P. analysed the results. All authors reviewed the manuscript.

References

[1] Foteini Alvanaki, Sebastian Michel, Krithi Ramamritham, and Gerhard Weikum. See what's enblogue: Real-time emergent topic identification in social media. In *Proceedings of the 15th International Conference on Extending Database Technology*, EDBT '12, pages 336–347, 2012.

[2] Fabio Ciulla, Delia Mocanu, Andrea Baronchelli, Bruno Gonçalves, Nicola Perra, and Alessandro Vespignani. Beating the news using social media: the case study of american idol. *Arxiv*, page 6, may 2012.

[3] Cath Everett. Ransomware: to pay or not to pay? *Computer Fraud & Security*, 2016(4):8 – 12, 2016.

[4] Salvatore Gaglio, Giuseppe Lo Re, and Marco Morana. A framework for real-time twitter data analysis. *Computer Communications*, 73(Part B):236 – 242, 2016. Online Social Networks.

[5] Cristobal García, Paul Chauveau, Javier Ledezma, and Maria Pinto. What can Social Media teach us about protests? Analyzing the Chilean 2011-12 Student Movement's Network evolution through Twitter data. *ArXiv*, aug 2013.

[6] Julio Hernandez-Castro, Edward Cartwright, and Anna Stepanova. Economic analysis of ransomware. *SSRN*, 2017.

[7] Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. Social Networks that Matter: Twitter Under the Microscope. *SSRN Electronic Journal*, pages 1–9, dec 2008.

[8] F. Jin, Edward Dougherty, Parang Saraf, Yang Cao, and Naren Ramakrishnan. Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, SNAKDD ́13, pages 8:1–8:9. ACM, 2013.

[9] Michael Mathioudakis and Nick Koudas. Twittermonitor: Trend detection over the twitter stream. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, SIGMOD '10, pages 1155–1158, 2010.