Strategos, 4(2), 2020, 39-65
UDK 32
UDK 355/359
Original scientific article[1]

STRATEGOS

Vulnerability assessment of the Croatian cyberspace to information warfare campaign via means of malicious websites comments

Dalibor Gernhardt

Abstract

Influencing masses is one way of achieving military and political goals. As seen in the U.S. 2016 election campaign, adversaries are prepared to go great length to test new ways of battle. When event such as terrorist attack or natural disaster strikes, people are prone to believe anything they see without questioning the source or truthfulness of information. This work focuses on researching steps necessary to be performed by adversaries aiming to perform influence operation by method of placing malicious comments on websites. Potential adversary must evaluate target, identify most relevant websites and analyse commenting systems to make decision how to exploit them. Once adversaries choose course of action, their next step is creation of trustworthy, in this case Facebook profiles, which later can be used for malicious operations. For purpose of testing this methodology, vulnerability assessment of the Croatian webspace is performed, and Facebook as a dominant platform for writing comments is identified. In conclusion a formula for estimation of workforce required for creation and maintenance of false Facebook profiles is given. Knowledge about adversaries' action is essential for effective defence in hybrid warfare environment.

Keywords

hybrid warfare, cyberspace, websites comments, social networks, influence operation

1 Članak je primljen u Uredništvo 30. ožujka 2020. i prihvaćen za objavu 4. prosinca 2020. (The article was received by the Editorial Board on March 30, 2020 and accepted for publication on December 4, 2020.)

Introduction

Hybrid warfare is if not always, then at least very often performed by influencing masses. Although information warfare in military terms is not a new term, its usage in cyberspace is

relatively new. Both NATO and Russian military doctrine recognise cyberspace, in 2016 NATO defined cyberspace as "Domain of Operations" (North Atlantic Treaty Organisation, 2016) and Russian Military doctrine form 2010 allows shaping of favourable reaction via means of information warfare (Russian Federation, 2010). It is important to note that in west, cyber security and information security are separated, considered to be two different fields. In Russia cyber is a part of information security (Jaitner, 2015). One of many kinds of information operations is by writing malicious comments on regular news web sites with goal of shaping favourable public opinion. These comments can be bias to some topic, they can be used for: spreading false information, rumours, encouraging society division etc. Spread of false information can be explained through example of Trojan horse, where Trojan citizens themselves, after being influenced and persuaded by single individual (an agent for spreading influence – an mediator), brought in a wooden horse (bearer of misinformation) in their city, Trojans were acting like "resonance boxes" spreading misinformation from one to another until they all thought that bringing wooden horse in city is a good idea (Volkov, 2002). According to (Vojak, 2017), false information can also be accidental: viral fake news or irresponsible media, or intentional: for profit or fake news as an agent of chaos or influence. Examples of false information are described in great detail in work (Vojak, 2017).

This work aims to identify attack preparation process of adversary wanting to prepare small-scale attack modelled after case of U.S. 2016. election and influence operation conducted by Internet Research Agency (IRA). Case of IRA is well known and documented inside of U.S. but there are no works which describe how vulnerable Croatian society would be to similar, but simplified campaign, limited only to writing comments on webpages.

This work assesses vulnerability of Croatian websites on hypothetical small-scale information influence operation, modelled after IRAs campaign, where

adversary wants to shape and destabilise public trust only by mean of writing forged user comments on web portals containing news or news related web pages. For average user these comments and personas behind them need to appear as written by real persons, so it is assumed that attacker will use convincingly forged fake social networking accounts, mimicking real user behaviour.

This work is structured as follows: First, the tactics of IRA are examined, followed by a general revision of the options for placing comments on websites. In the next chapter, for the purpose of testing possible adversary tactics, Croatian news web portals are examined and their commenting system analysed. In this chapter, Facebook will be recognized as a platform that allows reaching the majority of websites. Assuming that a potential adversary will try to imitate a real Croatian Facebook user, a model of the average Croatian Facebook user is given. Since the adversaries wants to protect their fake accounts from detection, an overview of Facebook's countermeasures regarding the detection of such accounts is given. This is followed by an examination of challenges adversary faces during operation. The paper concludes with a model that estimates the size of group required to maintain malicious accounts before activation.

Scenario role model: Internet Research Agency

Majority sources regarding Internet Research Agency (IRA) are coming from court cases and from research journalism. This chapter aims to summarize available data regarding this institution and tries to give an objective description of the operation conducted. Institution known as Internet Research Agency (IRA) was situated in St Petersburg, Russia. This institution employed civilians, and performed highly complex large-scale influence campaign through usage of social media: groups, ads, user status updates, mems, and among other, by comments on mass media portals and social networks – Troll farms (Chen, 2015; United States District Court for The District of Columbia, 2018; U.S. Senate Intelligence Committee, 2019). Goal of the campaign was to divide Americans among themselves by initiating and amplifying existing divisions in society.

Tactics used by IRA can be shortly described as: high volume and multiple channels. IRA used both: fake and stolen social media profiles. Each employee was responsible for maintenance of multiple online social network (OSNs) profiles and had received daily tasks with topics for coverage, with a daily quota to meet. According to (Chen, 2015) daily targets for employees were: 10 non-political posts, 5 political posts and 150 to 200 comments on co-workers' posts to make their, fake, profiles look legit. Operators (employees) received training on local American narrative and current topics. Team leaders ensured daily targets are met. Operators were never in direct contact with IRA's "upper" departments, departments assigned with monitoring and giving objectives. IRA's large-scale operation used almost all available online social networks, internet platforms and even creating cross platform connections between forged accounts to make accounts more believable, examples are: Facebook, Instagram, Reddit, YouTube, LinkedIn, Vine, Tumblr, 4chan, 9gag, Gmail, Pinterest, Gab, Meetup, VKontakte, LiveJournal. Desired effect was to cause information overload and overwhelm target audience (Chen, 2015; U.S. Senate Intelligence Committee, 2019).

In 2018 the United States District court for the district of Columbia filed a criminal case against IRA and its affiliates for purpose of influencing federal elections during period from 2014 to present day (United States District Court for The District of Columbia, 2018). IRA and associates were accused for "…posing as U.S. persons and creating false U.S. personas, operated social media pages and groups designed to attract U.S. audiences." (United States District Court for the District of Columbia, 2018). Among other, IRA was accused for: using stolen identities of real U.S. persons, traveling to U.S. for intelligence purpose, using computer infrastructure partly based in the U.S. to hide Russian origin, buying political advertisements on social media, staging political rallies inside the U.S., all stated by operating through a number of Russian related entities.

Organisation was headed by management group organized in different departments such as: data analysis, search engine optimisation, IT department, graphic department, and finance department. Organisation regularly evaluated its content to ensure that all posts and operations appeared as of authentic U.S. personas. Online social media accounts had

hierarchy and some accounts were used to promote another. To obscure operators' true location and to make them appear as they were operating from inside of U.S. they used computer space on servers located in the U.S. and connected them via virtual private networks (VPNs). It is estimated that IRA managed to abuse Facebook ad system by carefully placing targeted divisive ads, ads that were 10 times more effective than standard ads (Ribeiro *et al.*, 2019). Some analysis of IRA's Twitter accounts points to existence of standardized accounts types named: *Right Troll, Left Troll, Newsfeed, Hashtag Gamer, Fearmonger* depending on topics dominating on these accounts (Linvill and Warren, 2018).

According to available data, it is clear that IRA was well funded, well organised, with clear goals and working in conjunction with Russian Main Intelligence Directorate (GRU) (U.S. Senate Intelligence Committee, 2019). GRU's hacking group (advanced persistent threat group known by name APT28) role was to obtain sensitive documents by hacking "employees of major U.S. political campaigns" (U.S. Senate Intelligence Committee, 2019) and in assistance with later release of stolen documents with IRA.

Online social network profiles used by actors like one described in case of IRA are hard to distinguish from real accounts because behind them is human and these profiles to untrained eye act completely normal until activated. Organisations like IRA monitor language skills of their employees and adjust their working schedule according to time zone of targeted geographic area. Although IRA was shut down, it not likely that it is the only organisation conducting these kinds of information operations. In U.S. Congress committee report stated that ability to identify Russian activity on social media platforms was limited and reliant on social media companies to identify these threats (U.S. Senate Intelligence Committee, 2019) meaning that without close cooperation with private organisations it is very difficult to identify these threats. Other researches have also pointed out that although methods for detection of malicious accounts do exist, those methods are relying on analysing data that is only available to social networks themselves, and generally not available to other parties (Adewole *et al.*, 2017; Wani and Jabin, 2018; Hannah, 2020)online social networks (OSNs.

Options for placing comments on websites

Web sites enable their users to place comments on their pages to encourage discussions, increase traffic to websites, and even as source of information for their articles. From technical perspective, administrators can implement commenting systems using different methods. To make whole process simpler and to gain higher volume of user comments website may have no registration policy, meaning that any visitor to webpage can make comments to articles. Trade of this approach is higher time spent in moderation of user comments and comments lack of quality. Websites regularly perform moderation of user comments due to: house rules, profanities, spam, hate speech, divisive speech and fake news (Reich, 2011). Some of stated reasons for moderation, such as profanities, are easy to perform by usage of automated filters, but regardless of level of automation the whole process is both: time and resource consuming for website owner. For websites, first step of moderation is to decide how users can make comments as shown in Figure 1. Web sites have options to leave anonymous commenting without account, use their own account system, or use social media account for login (Limba and Šidlauskas, 2018)data may spread through cyber space at the speed of lightning. News portals constantly update the information available at their disposal by posting new articles. In order to attract new readers and to retain existing ones, in addition to focussing on publishing quality content, portal managers work on continuously improving their sites. These websites may have various interactive features, among them the opportunity to comment on an article. In some news portals, the number of anonymous comments is particularly high. The activities of online commenters and the issues related to their anonymity have always generated heated discussion owing to a number of reasons, including the content of the comments, the right of the commenters to remain anonymous and the extent to which the portal manager could be held liable. News portals equipped with an anonymous commenting function give rise to a culture of online bullying and hate-mongering where the cyber-criminals feel immune from punishment and existing control measures are insufficient for addressing the problem. The advocates of anonymous commenting argue that it promotes freedom of expression and portal administrators claim they can control defamatory and offensive anonymous comments by deleting

them. The article discusses the theoretical aspects of anonymity, anonymous commenting and anonymous comments. Based on a case study of the most popular news portals in Lithuania, and, in particular, on a comparative analysis of the privacy policy and the environment for commenting in three of them, the authors offer empirical data on the ratio between the number of comments and that of the commenters. The main purpose of the article is to reveal the peculiarities of anonymous comments' management of the news portals that enjoy the greatest popularity in Lithuania.","author":[{"dropping-particle":"","family":"Limba","given":"Tadas","non-dropping-particle":"","parse-names":false,"suffix":""},{"dropping-particle":"","family":"Šidlauskas","given":"Aurimas","non-dropping-particle":"","parse-names":false,"suffix":""}],"container-title":"Entrepreneurship and Sustainability Issues","editor":[{"dropping-particle":"","family":"Tvaronavičienė","given":"Manuela","non-dropping-particle":"","parse-names":false,"suffix":""}],"id":"ITEM-1","issue":"4","issued":{"date-parts":[["2018","6","29"]]},"page":"875-889","title":"Peculiarities of anonymous comments' management: a case study of Lithuanian news portals","type":"article-journal","volume":"5"},"uris":["http://www.mendeley.com/documents/?uuid=dca26e7f-c096-4a07-b34c-6a71ddc2d822"]}],"mendeley":{"formattedCitation":"(Limba and Šidlauskas, 2018. In all cases real name of user is not guaranteed.

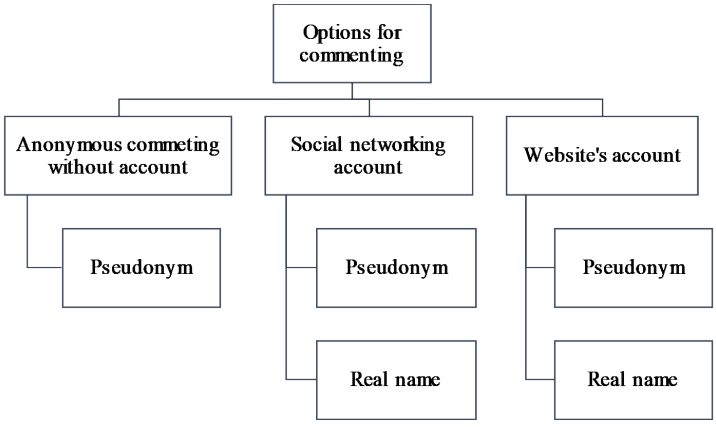

Figure 1. Model for selecting options for commenting (Limba and Šidlauskas, 2018)data may spread through cyber space at the speed of lightning.

News portals constantly update the information available at their disposal by posting new articles. In order to attract new readers and to retain existing ones, in addition to focussing on publishing quality content, portal managers work on continuously improving their sites. These websites may have various interactive features, among them the opportunity to comment on an article. In some news portals, the number of anonymous comments is particularly high. The activities of online commenters and the issues related to their anonymity have always generated heated discussion owing to a number of reasons, including the content of the comments, the right of the commenters to remain anonymous and the extent to which the portal manager could be held liable. News portals equipped with an anonymous commenting function give rise to a culture of online bullying and hate-mongering where the cyber-criminals feel immune from punishment and existing control measures are insufficient for addressing the problem. The advocates of anonymous commenting argue that it promotes freedom of expression and portal administrators claim they can control defamatory and offensive anonymous comments by deleting them. The article discusses the theoretical aspects of anonymity, anonymous commenting and anonymous comments. Based on a case study of the most popular news portals in Lithuania, and, in particular, on a comparative analysis of the privacy policy and the environment for commenting in three of them, the authors offer empirical data on the ratio between the number of comments and that of the commenters. The main purpose of the article is to reveal the peculiarities of anonymous comments' management of the news portals that enjoy the greatest popularity in Lithuania.","author":[{"dropping-particle":"","family":"Limba","given":"Tadas","non-dropping-particle":"","parse-names":false,"suffix":""},{"dropping-particle":"","family":"Šidlauskas","given":"Aurimas","non-dropping-particle":"","parse-names":false,"suffix":""}],"container-title":"Entrepreneurship and Sustainability Issues","editor":[{"dropping-particle":"","family":"Tvaronavičienė","given":"Manuela","non-dropping-particle":"","parse-names":false,"suffix":""}],"id":"ITEM-1","issue":"4","issued":{"date-parts":[["2018","6","29"]]},"page":"875-889","title":"Peculiarities of anonymous comments' management: a case study of Lithuanian news portals","type":"article-journal","volume

":"5"},"uris":["http://www.mendeley.com/documents/?uuid=dca26e7f-c096-4a07-b34c-6a71ddc2d822"]}],"mendeley":{"formattedCitation":"(Lim ba and Šidlauskas, 2018

Websites tend to use social networking account login systems to broaden the audience. Users do not have to go through registration process, remember usernames and passwords. All user needs to do is to login using existing social networking account and it is immediately set to place comments. This approach is beneficial for both web pages and users: web site administrators do not have to worry about user's registration process, email addresses, forgotten passwords, securely storing users' passwords etc. Using social media plugin for login or commenting is also a measure of reducing spam and increasing creditability of comments because each user is signed with, assumingly, real name and surname, often has profile picture attached. For every placed comment it is visible who is the author, and the author's public social media profile is available to every other user resulting in more credible comments. Web site administrators leave user verification to another party – social networking site. Additional benefit is that web pages get data about visitors needed for advertisement purposes. On the other hand, some users care more about privacy and they are not willing to share their identity. It is important to note that online social networking sites have full access to user activity.

Analysis of Croatian's news web portals

In this hypothetical scenario, adversary wants to influence public by placing comments on websites. At least to the author, there is no known data or methodology on how to prepare for this kind of operation, nor how it is done in practice. In this situation, best method is to try to think like adversary would. As one of first steps of any successful operation is intelligence preparation, adversary is likely to ask following questions:

- how many relevant news related web sites exist in targeted area,
- how many of them enable their readers to comment on articles, and
- how do websites manage user accounts.

For purpose of testing this approach, in following text Croatian webspace will be analysed. To identify most influential news web pages, first step is to identify valid data source regarding websites visits. In case of Croatia, largest publicly available dataset is available via Gemius.com2 service. For purpose of identification of most relevant Croatian websites at the time of writing this article data for month of November of 2019 was used (Gemius S.A., 2020). The goal was to reduce list of websites to only relevant sites and to discard minor websites with low traffic and activity. From initial list of web sites, only sites with 100 000 unique visits during observed month were analysed. Furthermore, Gemius public dataset is not limited to websites containing news, so websites focused on: fashion, sports, cars, health concerns, family or children advices are excluded from analysis because they are not relevant, as they have low possibility of political influence via method of placing comments. After applying described filters, from initial list of 54 web portals with more than 100 000 visits, 18 of them were dismissed due to non-political content. During initial review it was noticed that some popular portals are missing from the list because they do not participate in Gemius program. To objectively supplement Gemius dataset, an alternative list of most visited websites called SimilarWeb3 was used (SimilarWeb LTD, 2020b). In alternative list, only basic information is offered free of charge, and furthermore, SimmilarWeb uses different visitor counting methodology meaning that visit data present in the SimilarWeb dataset is not directly comparable to visit data in the Gemius dataset. After data review, two websites (index.hr and forum.hr) were added into analysis, but due to different visitor data methodology their visit numbers will not be considered in a latter discussion. Real adversary would likely have access to data behind paywall and thus could use single data source, nevertheless the methodology for determining relevant websites is unaffected. In total, 38 Croatian news web sites were analysed (Table 1.) with over 200 million

2 Gemius dataset is created by collecting cookie and browser data for advertising purpose and, among other information, gives information of unique visits during months period. Data is available for following countries: Belgium, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Moldova, Romania, Serbia and Slovakia.

3 SimmilarWeb is web analytic platform which gathers worldwide data regarding websites traffic, referral sources by usage of different methods.

views during one observed month. Each site was visited, and following topics were assessed:

- Are visitors allowed to post comments?
- If comments are enabled, what method is used:
  - proprietary commenting system (website account),
  - Facebook login or Facebook commenting plug in,
  - Google login,
  - no registration required.

Table 1. Commenting system on most visited Croatian web sites as December 2019, visit data compiled form (Gemius S.A., 2020)

	Website	Visits [millions]	Commenting enabled	Social networks plugins		Website account	Anonymous comments
				Facebook	Google		
1	24sata.hr	40.82	Yes	Yes	Yes	Yes	No
2	net.hr	26.29	Yes	Yes	No	No	No
3	vecernji.hr	23.66	Yes	Yes	No	Yes	No
4	dnevnik.hr	20.92	Yes	Yes	No	No	No
5	rtl.hr	16.87	Yes	Yes	No	No	No
6	tportal.hr	15.73	Yes	Yes	No	No	No
7	telegram.hr	7.67	Yes	Yes	No	No	No
8	dnevno.hr	7.16	Yes	Yes	No	No	No
9	novilist.hr	4.62	Yes	Yes	No	No	No
10	express.hr	3.73	Yes	Yes	No	Yes	No
11	poslovni.hr	3.54	Yes	No	No	Yes	No
12	dalmatinskiportal.hr	3.27	No	/	/	/	/
13	direktno.hr	2.72	Yes	Yes	No	No	No
14	glasistre.hr	2.69	No	/	/	/	/
15	n1info.com	2.53	Yes	No	No	No	Yes

16	dalmacijadanas.hr	2.44	Yes	No	No	No	Yes
17	srednja.hr	2.07	Yes	Yes	Yes	Yes	Yes
18	hrt.hr	2.00	No	/	/	/	/
19	klik.hr	1.60	No	/	/	/	/
20	sibenik.in	1.30	Yes	Yes	No	No	No
21	podravski.hr	1.21	No	/	/	/	/
22	zagreb.info	1.17	Yes	Yes	No	No	No
23	epodravina.hr	1.03	Yes	Yes	No	No	No
24	istarski.hr	0.87	Yes	No	No	Yes	No
25	bug.hr	0.85	Yes	No	No	Yes	No
26	prigorski.hr	0.74	No	/	/	/	/
27	glas-slavonije.hr	0.73	Yes	Yes	No	No	No
28	zagorje.com	0.69	Yes	Yes	No	No	No
29	mirovina.hr	0.65	Yes	Yes	No	No	No
30	057info.hr	0.63	Yes	No	No	Yes	No
31	sbplus.hr	0.54	Yes	No	No	Yes	No
32	teen385.com	0.37	Yes	No	No	Yes	No
33	ebrod.net	0.36	Yes	Yes	No	No	No
34	lider.media	0.28	Yes	Yes	No	No	No
35	mojfaks.com	0.22	Yes	Yes	No	No	No
36	studentski.hr	0.18	Yes	Yes	No	No	No
/	index.hr*	n/a	Yes	Yes	No	No	No
/	forum.hr**	n/a	Yes***	No	No	Yes	No
Total:		202.15	32	23	2	11	3
Percentage of sites:			84.21%	68.75%	5.23%	28.95%	7.89%

**index.hr does not use same visit data methodology as other sites (Index promocija d.o.o., 2015),*

***forum.hr does use Gemius service, but visit data is not publicly available,*

****forum.hr is a popular bulletin board, not a news website as such.*

In total, 84% of observed websites enable users to post comments. Users with online social networks account can post comments on 69% of observed sites. Due to uneven visit distribution, number of websites enabling commenting option is not a valid measure. By comparing website views instead, it is possible observe real reach of social network accounts in Croatian websites. In Figure 2, a ratio between visits to websites in regard of options for commenting: Facebook commenting, other mean of commenting or no commenting is shown. By analysing ratio between sum page views on all websites enabling comments and sum page views of all websites that enables login via Facebook platform, we come to result that out of total 190 million views on sites with comments enabled, 179 million views or around[4] 94% are websites using Facebook platform. All results of websites evaluation are shown in Table 1.

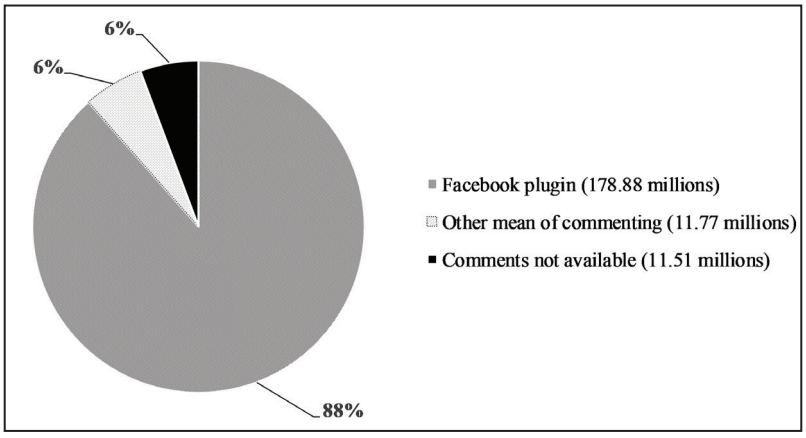

Figure 2. Distribution of commenting options compared to total website visits in observed time period, number in brackets shows number of visits in observed month

4 Directly comparable visit data for index.hr and forum.hr is not available, other publicly available data states that index.hr has around 50 million visits and forum.hr around 5 million visits per month (SimilarWeb LTD, 2020a). Since index.hr does use Facebook plugin and forum.hr does not, when we take visit numbers into account, then we are dealing with roughly 230 million visits with Facebook plugin and 17 million using different means, looking the data this way total reach by Facebook plugin is higher (230 millions), but percentage is slightly lower at approximately 93%.

In conclusion, for potential adversary aiming to place comments on Croatian websites, Facebook social media account is identified as most viable option. By using single Facebook account adversary can have access to 94% of totally viewed webpages which enable users to make comments (in observed time period). In this way adversary can use one account to access almost all webpages without necessity dozens of different accounts. Main disadvantage of this centralised approach can be viewed in following: if Facebook account is compromised, then attacker will no longer have access to all websites where account was used to post comments.

Facebook as dominant platform

Modelling of average Facebook user in targeted area

As shown in previous chapter, Facebook is identified as a dominant platform in Croatian society, especially for enabling placement of user comments on websites. Adversary aiming to exploit commenting system via Facebook accounts wants to create convincing Facebook accounts, mimicking real user daily behaviour with goal of making Facebook automated fake account detection more difficult. Adversaries next step is modelling of average Facebook user in targeted area. Most of publicly available data regarding Facebook users is aimed for marketing purposes. Most public reports do not state methodology of information collecting, they focus on different users' habits and present statistics of users living in different countries, using different time periods thus making direct comparison difficult. Not all data is available for all geographical regions, making Facebook user modelling using publicly available data difficult.

To make assumption on average Croatian Facebook user following publicly available data was used:

- average Croatian comments other users status update five time a month - 0.16 status comments per day (Arbona d.o.o., 2019),
- average Croatian user makes one status share per month (Arbona d.o.o., 2019),

- average Croatian user likes one page per month (Arbona d.o.o., 2019),
- average Croatian 17 times per month visits advertised web pages (Arbona d.o.o., 2019),
- data regarding average Facebook usage time for Croatian user is not publicly available, but we can approximate it using knowledge that during 2019. average U.S. Facebook user older than 18 years spent 37 minutes[5] a day using Facebook (Droesch and eMarketer, 2019),
- in addition, as beginning of 2019 Croatia had 1.9 million of Facebook users, 1.1 million of Instagram users and 168.5 thousand of Twitter users (Hootsuite, 2019).

According to available data, results reveals that although most websites offer for their users to use Facebook for placing comments, average Croatian Facebook user is not very active, and for potential adversary it would be relatively easy to mimic real user behaviour. One person could maintain relatively large number of fake profiles.

Facebook measures against fake accounts

Facebook itself is aware of platform exploits and it is fighting against "Inauthentic Behaviour", and in one such report stating that Facebook removed 50 networks of such accounts worldwide (Rosen *et al.*, 2019). As one measure filters detecting accounts and pages who repeatedly share confirmed misinformation[6] are installed without noting if this measure is language dependent. Most of these measures are aimed toward to 2020 US elections although Facebook representative in one report stated that they removed 36 suspicious French accounts posting about various political topics

5 Measured time includes all time spend on social platform regardless of user multitasking, meaning that social platform has been accessed for 37 minutes in average while user could have done other things simultaneously, for example having opened multiple tabs in web browser (Droesch and eMarketer, 2019).

6 Authors of the report do not distinguish between terms misinformation (*false or wrong information which are accidentally transmitted and which can lead to damage – not sent with intent) and disinformation (false or wrong information sent with intention of influencing opinion in favour of person who is sending information)* (Tuđman, 2012)

(Gleicher and Facebook, 2018). It is noted that that recent accounts used to spread influence are more sophisticated and harder to detect than those in case of IRA (Facebook, 2018). Attribution of malicious accounts is difficult because actors behind them (like IRA or some other groups) improve their techniques once being uncovered. Facebook attributes actors in four general categories based on: (1) Political motivations, (2) Coordination, (3) Tools, Techniques and Procedures (TTPs) and (4) Technical Forensics (Stamos and Facebook, 2018). In report (Rosen and Facebook, 2019) Facebook estimates that 5% monthly active accounts are fake accounts without stating are they automated, bot accounts or connected in any measure with politically motivated accounts. During a period from January to September of 2019 5.4 billion, mostly automated fake accounts have been disabled (Facebook, 2019).

Facebook does not publicly disclaim technical details on how are they discovering fake accounts, giving only few technical examples such as: blocking certain IP addresses and even ranges if they detect large number on new accounts originating from that IP, network or location, removing accounts using suspicious email addresses and actions, removing accounts if other users report it as fake (Facebook and Schultz, 2019). As previously noted, in U.S. Congress committee report it is stated that detection of malicious accounts is reliant on social media companies themselves and governments have limited influence (U.S. Senate Intelligence Committee, 2019).

Adversary

Adversary challenges

As Facebook has been identified as a dominant platform, in this chapter focus is given on adversary challenges regarding creation of fake Facebook profiles. Adversary can create many fake profiles at once, but if he does so, it is likely that this action will trigger suspicion. To create profiles which are both: convincing and difficult to automatically detect, it is likely that they will be gradually created over time, long before information operation itself.

These accounts can be created and maintained for years, and weaponized when needed. From different types of possible accounts types described in detail (Adewole *et al.*, 2017; Wani and Jabin, 2018)online social networks (OSNs, in this case we will focus on manual creation of accounts. These accounts can be created either by using non existing identities, by using stolen personal information or by usage of existing but stolen social network accounts[7]. Adversary aims to craft and maintain accounts in such way to deter automatic detection. For example, for Facebook it would be easy to detect large number of accounts that have been inactive for years and then suddenly activated at once. As result, adversary must maintain activity on forged, fake accounts even when not in use for they intended purpose. Adversary will likely use both fake and stolen identities for creation of social network profiles. At time of writing, Facebook procedure for reporting usage of stolen identity states that victim itself must report identity theft and provide evidence for identity theft in form of a personal document, otherwise request won't be considered by Facebook (Facebook, 2020). Knowledge about this relatively complicated procedure makes usage of stolen identities by adversary more likely. Adversary will likely exploit fact that detection of fake profiles is reliant of social network providers meaning that attacked side will have problems defending against such attacks as visible in examples presented in (Hannah, 2020).

Depending on adversary intentions, it can perform small scale operation by placing targeted comments on selected topics at targeted geographical area and thus attract less attention and suspicion from intended victim, potentially making this kind of operation very effective. If adversary engages in larger operation it is more likely to gain higher attention, potentially leading to state officials starting to investigate links of comments and user accounts possibly leading to compromise of whole operation.

Adversary wants to have reliable social media accounts, accounts which will not trigger suspicion when inspected by both: average user and some automated detection system. Those accounts need to be specially tailored

7 According to (Ablon, Libicki and Abler, 2014) on black market stolen social network profiles can have greater value than stolen credit card information.

to meet specific mission requirements. As discussed in earlier chapters, adversary must find a way to mitigate commonly used detection methods and it is likely to deal with following challenges:

- as each IP address reveals its approximate physical (geographical) location, accounts need to obscure their physical location to deter automatic detection (one mean of achieving this is by using multiple VPN connections to multiple densely populated areas to avoid correlation of accounts due to location proximity and behaviour),
- avoiding IP and possibly geographical location contact between accounts to deter automatic correlation between accounts, accounts need to consistently login form similar locations with goal of not attracting unwanted attention,
- avoiding correlation between accounts time of activity (i.e. accounts are often active at the same time),
- how to mimic real social media user behaviour with limited data regarding real users' behaviour available,
- evaluate and choose best approach to add friends to social media accounts: connecting fake accounts in networks or avoid contact and try to add real people to its friend list,
- adjust hours of activity of fake profiles to those of targeted time zone users,
- how to achieve a synergistic effect of all accounts,
- how to validate effectiveness of operation.

Estimate on personnel necessary for account maintenance

Putting technical challenges aside and knowing these kinds of operations have been performed in past, goal of this chapter is to estimate personnel resources necessary for maintenance of dormant (relatively inactive – at this phase not used for writing malicious comments) accounts. Due to large number of possible adversaries' modus of operandi, only a preparation phase (consisting of creation and maintenance of social network accounts) is analysed. By comparing publicly available data about average Croatian

Facebook user with data regarding accounts used by IRA as shown in Table 2, it is evident that potential adversary could not directly apply IRAs model due to significantly different level of user activity. Comparing to accounts used by IRA it is visible that average Croatian user is significantly less active and potential adversary could maintain more fake accounts per real person compared to IRAs case of six accounts per employee.

Table 2. Comparison on IRAs social media accounts and average Croatian Facebook user, data derived from (Seddon, 2014; Chen, 2015; Arbona d.o.o., 2019)

IRAs social media accounts	Average Croatian Facebook user
10 non-political, five political posts per day	one per month
150–200 comments per day	five per month
50 comments to news articles	*no data available*
Other, not directly comparable, available data	
Each blogger maintains six Facebook accounts Each blogger maintains 10 Twitter accounts with up to 2000 followers and 50 tweets a day	Average Croatian user likes one page per month

If adversary wants to have convincing Facebook accounts at its disposition, they need to be maintained 24/7 and especially tailored for specifics of certain geographical area. For getting representative value, we will examine much workforce is necessary to perform 24/7 operation. Let's introduce following variables:

T_{da} = [*minutes per profile*] – targeted daily activity (T_{da}) for single social network account,

A_{wf} = [*number of persons*] – available workforce (A_{wf}) maintaining accounts,

W_m = [*minutes*] – amount of daily working time in minutes (W_m) shown per employee, for example in case of 8 working hours, excluding brakes, this value equals to 480 minutes.

To address necessity of accounts to be active on every day of month regardless off weekends, additional factor is introduced C_w – Corrected week. If single employees have 5-day workweek then simple ratio of workweek days and actual week length will enable us to compensate for weekends, this factor can be treated as a constant (1):

$$C_w = \frac{\text{Numbers of workdays in week}}{\text{Numbers of days in week}} = \frac{5}{7} \tag{1}$$

As accounts must be maintained in convincing looking 24/7 operation, so we need to model different amount activity during day and night period. For example, we can assume 90% daily, 10% activity during night, in other words, if adversary has 20 persons at its disposal 18 persons will be working during daytime and two employees will always be unavailable during daytime because they perform night account maintenance (temporarily ignoring other factors as they are all modelled as independent). We can approximate this by factor of Daily activity (D_a):

D_a = [*no dimension*] – ratio of daytime and night activity, for example, a value of 0.9 means that accounts spend 90% of time during daytime.

In previous steps we have reduced all variables to daily level. Now by simple multiplication of variables it is possible to roughly approximate adversary's number of Fake profiles (F_p) with a given constraints. To recap, number of Fake profile depends on available workforce (A_{wf}), ratio between day and night accounts usage (D_a), multiplied by weekend correction factor (C_w of 5/7) and it is limited by average time that adversary wants to spend working on each account – $\frac{W_m}{T_{da}}$. If we assume there is no overlap between accounts activity (one employee works on one account at any given time) and multiply each of factors stated above, we can approximate number of fake profiles (F_p) as linear function of multiple variables (2):

$$F_p = A_{wf} \times D_a \times C_w \times \frac{W_m}{T_{da}} \quad (2)$$

Given formula (2) takes into consideration necessity for both day and night maintenance during every day in a year and as results gives number of profiles that could be maintained with given parameters and resources. Results depend on available workforce (operators behind profiles), ratio of day to night activity and total targeted daily activity of fake profiles.

We can put this formula to use by assumption that adversary wants to perform squad to company sized operation while spending 20, 30 or 40 minutes daily per fake Facebook profile and by using following values for variables:

$A_{wf} = 10 - 120$ [*number of persons*] – squad to company sized operation,

$D_a = 0.9$ [*no dimension*] – assuming accounts spend 90% of time during daytime,

$C_w = \frac{5}{7}$ [*no dimension*] – five workdays in a week,

$W_m = 480$ [*min*] – assuming each employee works for full 8 hours,

$T_{da} = 20, 30, 40$ [*min/profile*] – comparing 3 cases of activity time.

In case of $A_{wf} = 10$ *people* and $T_{da} = 40$ *min/profile*:

$$F_p = A_{wf} \times D_a \times C_w \times \frac{W_m}{T_{da}} = 10\ people \times 0{,}9 \times \frac{5}{7} \times \frac{480\ min}{40\ min/profile} \approx 77\ profiles \quad (3)$$

By using presented methodology, we come to conclusion that with given parameters, 10 operators can continuously maintain about 77 profiles. Complete results of this hypothetical scenario are shown in Figure 3, due to multiple variable influencing results, this result can be used only as a rough estimate of personnel necessary for maintenance of dormant account. By using selected parameters, we can estimate one platoon sized force (30 people) can maintain between 232 (for 40 minute a day activity) and 464 (20 minutes of day activity) of fake Facebook accounts. Due to fact that average

Croatian Facebook user is inactive compared to U.S. user (Table 2) time spent online on single account could be even shorter that assumed 20, 30 and 40 minutes per day, resulting in higher number of accounts.

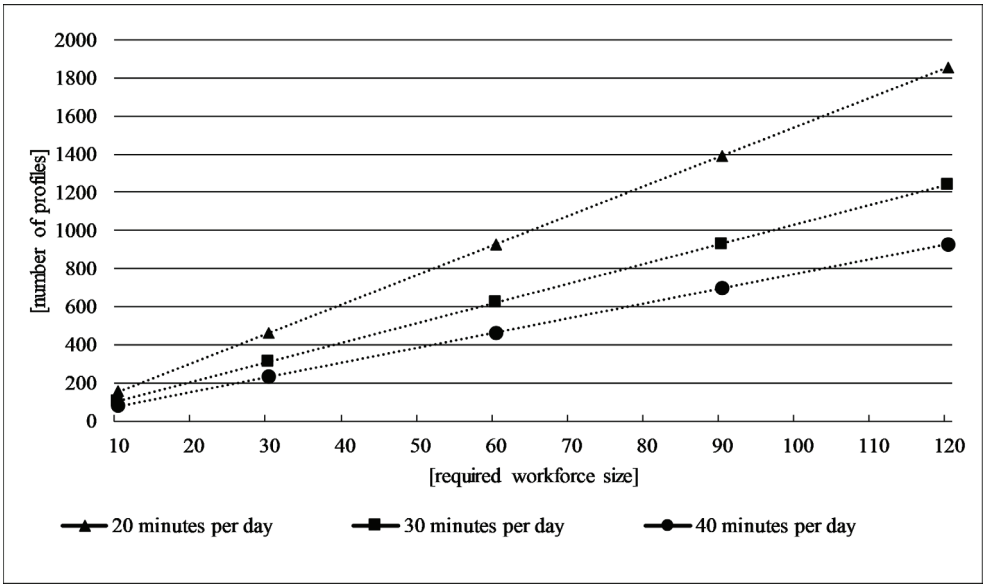

Figure 3. Number of dormant Facebook accounts depending on available workforce and account activity of 20, 30 and 40 minutes per day

This model concentrated only on personnel necessary for working on accounts, not on management, logistic or IT staff necessary for creating the conditions for operation. Once adversary deploys these accounts (starts placing malicious comments) it is difficult to give an estimate on number of personnel required due to different nature of every operation, desired goals, effect etc.

Conclusion

For defence against hybrid warfare, knowledge about threats, adversaries, motives, goals, techniques, tactics and procedures is essential. Without better understanding of the problem, it is not possible to employ adequate defence. As the example of IRA shows, adversaries are prepared to make great efforts to achieve their goals by using cyberspace opportunities. This paper focuses on one segment of cyber information warfare and attempts to identify steps that must be taken by adversary. In case of placing malicious comments that is to identify key websites in some countries webspace, identify how and when comments should be written and how to exploit them later. In test case of Croatia, it is shown that through usage of Facebook accounts as primary vector for placing comments it is possible to reach 94% of Croatian viewed pages meaning that potential adversary will most likely try exploit Facebook platform. Results of evaluation will vary from country to country and adversary must assess each country individually. Once adversary identifies mode of operation, adversaries' goal is to mimic real users in targeted area and creation of falsely trustworthy accounts which are dormant until used for information operation itself. As shown, adversary wanting to mimic Facebook user, faces multiple unknowns which can be used to reveal malicious activity. Analysis shows adversary aiming to create fake Facebook profiles, while not in use, can maintain dozen accounts per one operator. Knowledge about adversary steps is essential for planning defence against hybrid threats. For future work it is necessary to test same or similar methodology to another countries revealing differences between countries and examining wider range adversaries' procedures.

References

Ablon, L., Libicki, M. and Abler, A. (2014) 'Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar', *Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar*. doi: 10.7249/rr610.

Adewole, K. S. *et al.* (2017) 'Malicious accounts: Dark of the social networks', *Journal of Network and Computer Applications*, 79, pp. 41–67. doi: 10.1016/j.jnca.2016.11.030.

Arbona d.o.o. (2019) *Facebook i Instagram u Hrvatskoj: zanimljive statistike - 2019.* Available at: https://www.arbona.hr/blog/drustveni-marketing/facebook-i-instagram-u-hrvatskoj-zanimljive-statistike-2019/2832 (Accessed: 16 January 2020).

Chen, A. (2015) *The Agency*, *The New York Times Magazine*. Available at: https://www.nytimes.com/2015/06/07/magazine/the-agency.html (Accessed: 15 January 2020).

Droesch, D. and eMarketer (2019) *Instagram's New Explore Ads Signal Potential Changes to Organic Reach*, *eMarketer inc.* Available at: https://www.emarketer.com/content/instagrams-new-explore-ads-signal-potential-changes-to-organic-reach (Accessed: 16 January 2020).

Facebook (2018) *Removing Bad Actors on Facebook*. Available at: https://about.fb.com/news/2018/07/removing-bad-actors-on-facebook/ (Accessed: 15 January 2020).

Facebook (2019) *Community Standards Enforcement Report*. Available at: https://transparency.facebook.com/community-standards-enforcement#fake-accounts.

Facebook (2020) *Facebook Help Centre - Report an impostor account*. Available at: https://www.facebook.com/help/contact/295309487309948 (Accessed: 3 March 2020).

Facebook and Schultz, A. (2019) *How Does Facebook Measure Fake Accounts?* Available at: https://about.fb.com/news/2019/05/fake-accounts/ (Accessed: 2 March 2020).

Gemius S.A. (2020) *Domains metrics Croatia*. Available at: https://rating.gemius.com/hr/tree/8 (Accessed: 10 January 2020).

Gleicher, N. and Facebook (2018) *Analysis of French Language Material*. Available at: https://about.fb.com/news/2018/11/last-weeks-takedowns/ (Accessed: 13 January 2020).

Hannah, S. (2020) *ONLINE INFLUENCE AND HOSTILE NARRATIVES IN EASTERN ASIA*. Riga.

Hootsuite (2019) *Digital 2019: Croatia*. Available at: https://datareportal.com/reports/digital-2019-croatia?rq=croatia (Accessed: 15 January 2020).

Index promocija d.o.o. (2015) *Index najčitaniji i u rujnu*. Available at: https://www.index.hr/vijesti/clanak/index-opet-najcitaniji-hrvatski-portal/845355.aspx?fb_comment_id=944976845574705_945053858900337 (Accessed: 30 November 2020).

Jaitner, M. L. (2015) 'Russian Information Warfare: Lessons from Ukraine', *Cyber War in Perspective: Russian Aggression against Ukraine*. Tallinn: NATO Cooperative Cyber Defence Centre of Excellence Tallinn, Estonia, pp. 87–94. Available at: https://ccdcoe.org/uploads/2018/10/Ch10_CyberWarinPerspective_Jaitner.pdf.

Limba, T. and Šidlauskas, A. (2018) 'Peculiarities of anonymous comments' management: a case study of Lithuanian news portals', *Entrepreneurship and Sustainability Issues*. Edited by M. Tvaronavičienė, 5(4), pp. 875–889. doi: 10.9770/jesi.2018.5.4(12).

Linvill, D. L. and Warren, P. L. (2018) *Troll Factories: The Internet Research Agency and State-Sponsored Agenda Building*.

North Atlantic Treaty Organisation (2016) 'Warsaw Summit Communiqué'. North Atlantic Treaty Organisation.

Reich, Z. (2011) 'User Comments', in *Participatory Journalism*. Oxford, UK: Wiley-Blackwell, pp. 96–117. doi: 10.1002/9781444340747.ch6.

Ribeiro, F. N. *et al.* (2019) 'On microtargeting socially divisive ads: A case study of Russia-linked Ad campaigns on Facebook', *FAT 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, pp. 140–149. doi: 10.1145/3287560.3287580.

Rosen, G. *et al.* (2019) *Facebook Helps to Protect the 2020 US Elections*. Available at: https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/.

Rosen, G. and Facebook (2019) *An Update on How We Are Doing At Enforcing Our Community Standards*, *Community Standards Enforcement Report*. Available at: https://about.fb.com/news/2019/05/enforcing-our-community-standards-3/ (Accessed: 16 January 2020).

Russian Federation (2010) 'Military doctrine of the Russian Federation'. Available at: http://kremlin.ru/supplement/461.

Seddon, M. (2014) *Documents Show How Russia's Troll Army Hit America*, *Buzz Feed News*. Available at: https://www.buzzfeednews.com/article/maxseddon/documents-show-how-russias-troll-army-hit-america (Accessed: 15 January 2020).

SimilarWeb LTD (2020a) *SimilarWeb comparison: forum.hr Vs. index.hr*. Available at: https://www.similarweb.com/website/forum.hr/?competitors=index.hr (Accessed: 30 November 2020).

SimilarWeb LTD (2020b) *Top Websites Ranking Croatia*. Available at: https://www.similarweb.com/top-websites/croatia/ (Accessed: 20 November 2020).

Stamos, A. and Facebook (2018) *How Much Can Companies Know About Who's Behind Cyber Threats?* Available at: https://about.fb.com/news/2018/07/removing-bad-actors-on-facebook/#whos-behind-cyber-threats (Accessed: 15 January 2020).

Tuđman, M. (2012) *Programiranje istine. Rasprava o preraspodjelama društvenih zaliha znanja (Programming the truth. Treatise on the rearrangement of social stock of knowledge)*. Zagreb: Hrvatska sveučilišna naklada.

U.S. Senate Intelligence Committee (2019) *Report of the select committee on intelligence United States senate on Russian active measures campaigns and interference in the 2016 U.S. election.*

United States District Court for The District of Columbia (2018) *Case 1:18-cr-00032-DLF, UNITED STATES OF AMERICA v. INTERNET RESEARCH AGENCY LLC and Defendats.*

Vojak, B. (2017) 'Fake News: The Commoditization of Internet Speech', *California Western International Law Journal*, 48(1), pp. 123–158.

Volkov, V. (2002) *Dezinformacija, od trojanskog konja do interneta*. Beograd: Naš dom.

Wani, M. A. and Jabin, S. (2018) 'A sneak into the Devil's Colony- Fake Profiles in Online Social Networks', *arXiv*.

About the author

Dalibor Gernhardt (dalibor.gernhardt@morh.hr) graduated Electrical Engineering from the Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, Croatia. Since graduation in 2011 he has worked at different positions in Republic of Croatia Armed Forces, Ministry of Defence of Republic of Croatia, currently at Dr. Franjo Tuđman Croatian Defence Academy Janko Bobetko Center for Defence and Strategic Studies Int 2018 he started Postgraduate doctoral study programme area of technical sciences, scientific fields of electrical engineering and computing at University of Zagreb, Faculty of Electrical Engineering and Computing. His research interests include computer system and information security.