THE GHOSTWRITER CAMPAIGN

2023

AS A MULTI-VECTOR INFORMATION OPERATION:

ATTEMPTS TO CONTROL ITS INFLUENCE &
THE LIMITATIONS OF CURRENT COUNTER-MEASURES

TABLE OF CONTENTS

SUMMARY

This report synthesizes **publicly available open-source data** to map the **evolution of the 'Ghostwriter campaign'**, tracking and tracing how it has: (1) blended information manipulation with hacking; (2) targeted a number of countries; (3) operated successfully across a range of platforms; and (4) evolved and adapted over time. In the process, the evidence collated **illuminates structural weaknesses in current information operation counter-measures capacities and capabilities**, in terms of being able to constrain the activities of a persistent and adaptive adversary. Framed in this way, this analysis addresses the need within the countering disinformation community to gain a better understanding of hostile actors' abilities to **continuously develop complex and carefully designed influence operations**; as well as learning lessons about **'what works' in mitigating their impacts.**

To date, much of the countering disinformation community's attention and the consequent policy response options, have focused on models of operational design and delivery based upon the Internet Research Agency's (IRA) interference in the US election in 2016. However, other persistent, large-scale, and well-resourced operations have been run by hostile actors that significantly differ from the IRA's playbook. Studying how these information operations have reacted to different attempts to control and limit them, and the extent to which they have evolved and adapted in response, allows us to get **a more comprehensive understanding of the dynamics and interactions between hostile action and social control interventions.**

The **Ghostwriter campaign is a cyber-enabled influence campaign** that integrates information manipulation tactics and techniques and has triggered multiple social control responses from **across several European countries**. Despite these, it has managed to **continue and expand its activity**. Ghostwriter has been active since at least 2016. Significantly, it was **not really understood as a consistent campaign until 2020.**

The Ghostwriter operation is still **active and ongoing today**. Most recently in January 2022, **Ukraine** preliminarily connected a cyber-attack against dozens of government websites to UNC1151, the state-actor believed to conduct the cyber-activity behind Ghostwriter campaign. Facebook has taken down some Ghostwriter-linked activity targeting the Ukrainian military. Attributions of different parts of the activities under Ghostwriter to **both Russia and Belarus**, raise important questions about the cooperation between the two. In addition, in terms of 'attack' methodologies, the **integration of cyber-attacks and information manipulation** is worth attending to. Notably, following Russia's invasion of Ukraine, several hacker groups have intensified their activity and are increasingly engaging in information operations.

KEY FINDINGS FROM THE ANALYSIS ARE:

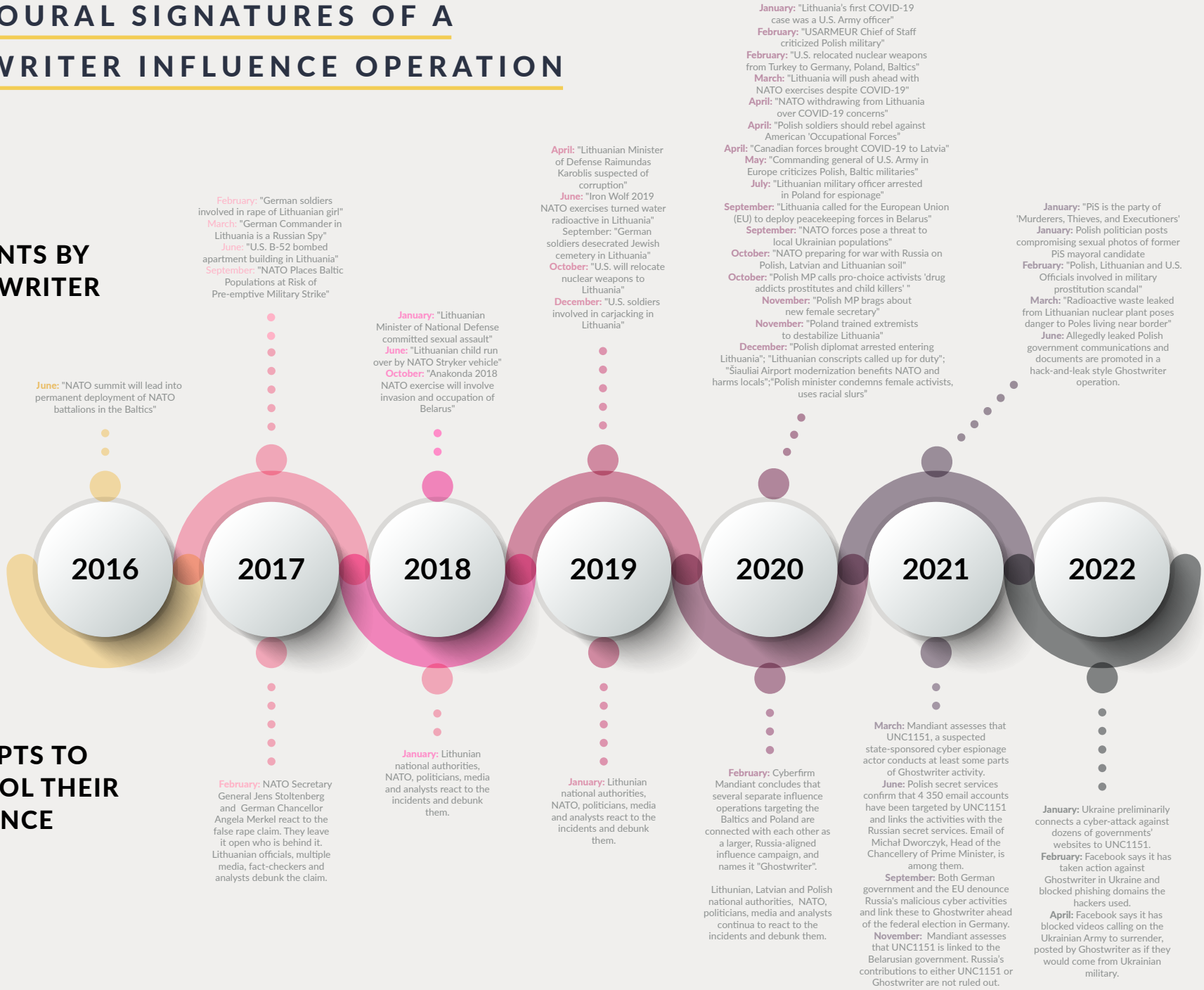
1. Based on open-source data, Ghostwriter has impacted thousands of email users, has hacked dozens of social media accounts and media websites, published hundreds of false blogposts and other falsified content, and impersonated multiple government officials, NATO representatives and journalists in Europe.
2. These operations are an ongoing threat and supported by a foreign state threat actor, either Russia, Belarus, or both. Over time, attribution has shifted, so that today the consensus is that Belarus has lead responsibility for running Ghostwriter.
3. Ghostwriter's cyber-enabled influence operations have triggered a variety of responses from governments, private cyber firms, social media platforms, media, and civil society. These have focused on strategic communication, public but partial attribution, improving cyber security, and most recently disrupting parts of the activity on Facebook and Google.
4. Fundamentally, serious gaps remain in our understanding of the scale and authorship of the operation, that has hindered the efficiency of the response and enabled continuous evolution of Ghostwriter's tactics. For example, Ghostwriter's cyber activity has been attributed to Russia's military intelligence (by Germany) and to the Russian state (by the EU). Over time, attribution has shifted, so that today Mandiant's assessment that Belarus has lead responsibility for running cyber activity behind Ghostwriter has not been challenged.
5. Social media companies have not completely removed activities linked to Ghostwriter from their platforms, and the campaign's false content remains online and available (for example on Facebook).
6. Ghostwriter has potentially had a significant cumulative impact and effect, given how its various activities have persisted over several years, across multiple social media platforms. This represents a different way of gauging the consequences of an information operation than has been previously applied to the **Internet Research Agency, where engagement figures were the primary metric used.**
7. Despite platforms' widening their threat reporting from "inauthentic behaviour" to cover a wider range of malign behaviours, there is no systematic overview of Ghostwriter's activity and related disruptions or takedowns on any of the social media platforms.
8. Criminologists use the 'term 'linkage blindness' to describe the problems that arise when different police agencies are investigating the same persistent perpetrator, and each investigator has a partial view of how and why the harmful act is being committed, but there is no-one positioned to draw all the individual pieces of knowledge together. This concept of 'linkage blindness' describes what has happened with the response to Ghostwriter, in that different governments and organisations have been looking at different facets, but no institution is positioned to take responsibility for adopting a comprehensive approach.

An additional key feature of the analysis reported herein, is to capture **how the repertoire of tactics and techniques used by the Ghostwriter operators has evolved over time**. It is inferred that as they have become more experienced and confident in their abilities, the operators have extended their geographic reach into new countries. They have also become more adept at combining different methods. Some of these shifts are **'forced adaptations' in response to control measures** and the need to navigate potential inhibitors, but **others are driven more by 'experiential learning'**. The Figure below provides a visual timeline representation of the key activities that have been attributed to Ghostwriter, since its initiation. Bringing this material together in one place, helps to clarify how and why the Ghostwriter campaign is important and significant.

BEHAVIOURAL SIGNATURES OF A GHOSTWRITER INFLUENCE OPERATION

INCIDENTS BY GHOSTWRITER

ATTEMPTS TO CONTROL THEIR INFLUENCE



Based upon the data collected[1] it is possible to identify several typical elements of a Ghostwriter operation, that are akin to its behavioural profile or 'signature'. These differentiate it from other known influence operations:

- Cyber activity conducted beforehand, through which access to compromised websites or social media accounts is obtained.
- Typical targeting involves a military component, either in messaging or in the choice of targets, but does not limit itself to that. Main targets are in Central and Eastern Europe.
- Content of messages is often faked and calls for a public rebuttal.
- Distribution of faked content mixes inauthentic accounts, spoofed emails, and impersonation; as well as compromised but real websites or social media accounts.
- Operations are timed and planned to coincide before or during important political events, such as high-profile visits or military exercises. Unlike other operations, they are rarely rapid reactions to events. For example, it is likely that initially Ghostwriter activity was a response to NATO's increased presence in the Baltic region.

EXAMPLE OF A TYPICAL GHOSTWRITER INCIDENT: POLISH SOLDIERS SHOULD REBEL AGAINST AMERICAN "OCCUPATIONAL FORCES"

A typical Ghostwriter operation combines outright fabricated content, distributed via a combination of compromised websites, fake emails, inauthentic accounts on blogging and/or social media platforms. The escalation of the operation is carefully timed.

On 22 April 2020 there was a hack of the Polish War Studies Academy website and publication of a fabricated letter from its commander Ryszard Parafianowicz. The letter called on Polish soldiers to rebel against the US occupying forces. Emails providing a link to the hacked website and the letter were sent to NATO and Polish government institutions. The emails were made to look like they were coming from a former MP of the Civic Platform party and an American journalist, asking for more information about the statement.

The distribution of the material continued with publication in English on 'the Duran', an online media platform repeatedly used in Ghostwriter operations. Furthermore, articles reporting the Commander's statement appeared on three Polish language news websites: lewy.pl, prawy.pl and podlasie24.pl. The headline for the associated articles was: "A Scandalous Letter by the Rector of the War Studies Academy: PiS Politicians Are Leading Us to Disaster." The websites claimed they were hacked.

Shortly after, dozens of Facebook accounts associated with Polish outlet Niezależny Dziennik Polityczny, Independent Political Journal (NDP) shared links to the news articles. Polish secret services have accused the NDP of connections with the Russian security services.

The timing of the operation coincided with the presidential election campaign period, as well as a military exercise. When the pandemic broke out, one of the biggest military exercises in Europe since the Cold War was already taking place. The US-led multinational exercise, with NATO's involvement, "Defender Europe 2020" was modified because of the pandemic, but carried on in a scaled down form, with Poland being one of the host countries.

1 See next chapter for methodology.

The operation's messaging was aimed at stoking internal divisions in Polish politics, as well as questioning the partnership with the US, for international audiences.

The operation triggered a public reaction: The War Studies Academy warned about the hack on Twitter the same day it occurred[2]. Next day, Polish secret services and fact-checkers published an analysis of the incident[3]. The Facebook accounts sharing the link were still active when Stanford Internet Observatory investigated them in early May that year, following which they were removed[4]. NDP doesn't have a Facebook account anymore, and its Twitter account has been suspended, but it has an active Youtube account.

Polish news outlets have removed the fabricated articles. However, a Crowdtangle search shows that Facebook has not removed them from the platform completely, only the NDP-connected accounts: The Pravy.pl link has been shared to 54 public groups and pages gaining 1,8k interactions. The Lewy.pl link has been shared to 12 groups gaining 300 interactions, which rises to 4,2k interactions if both public and private post engagement is counted.

The Duran's article has been removed, but some Twitter shares of it remain. Verity Weekly republished the article and posted it on Twitter, which is also still active.

The gaps and inconsistencies remaining in the attribution of the operation are also typical for a Ghostwriter operation. Polish secret services called the operation *"congruent with disinformation activities carried out by the Russian Federation"*[5]. Whilst the cyber-security firm Mandiant attributed the operation to Ghostwriter, it has later clarified it does not attribute NDP to a certain actor, even if notes the overlaps with Ghostwriter. Facebook has not mentioned NDP or the incident in its public reports on takedowns. Russian state-owned media's reaction was also typical: it reported on the hacking attack through the angle of ridiculing Poland as 'Russophobic'[6].

2 https://twitter.com/AkademiaSzWoj/status/1252991552655409152

3 https://www.gov.pl/web/sluzby-specjalne/atak-dezinformacyjny-na-polske, https://sprawdzam.afp.com/nie-general-parafianowicz-nie-nawolywal-do-walki-z-amerykanskim-okupantem-byl-atak-hakerski

4 https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/documents/sio_disinformation_polish_military_0.pdf

5 https://www.gov.pl/web/sluzby-specjalne/atak-dezinformacyjny-na-polske

6 https://news-front.info/2020/05/06/obostrenie-rusofobii-v-polshe-oboznachilos-poyavleniem-russkih-hakerov/, https://www.rubaltic.ru/article/politika-i-obshchestvo/06052020-russkie-khakery-na-prezidentskikh-vyborakh-polsha-perezhivaet-obostrenie-rusofobii/

METHODOLOGY

This independent analysis of the campaign and responses to it, draws together the publicly available open-source evidence about 34 incidents attributed to the Ghostwriter Campaign by FireEye/Mandiant[7], as well as the official government communications, media reports, fact-checks and NGOs and think tank analysis relating to these same incidents. These documentary materials have been supplemented with nine semi-structured, in-depth interviews conducted with various representatives of governments, media and civil society, who have been directly involved in responding, exposing, or analysing the incidents. To get a comprehensive overview of the campaign and the social control responses to it, further information was collected on how Russian language media has reported the incidents.

Informed by all these materials, the timeline for the Campaign's activity extends from Summer 2016 until Spring 2021. The incidents linked to Ghostwriter by different actors continue after that in Belarus, Germany, Lithuania, Poland, and Ukraine. These are referenced in the analysis, but not included in the coding or graphs.

After the initial data collection to map the incidents, the relating data were coded by targeted country/organisation, languages used, as well as most frequently used influence techniques: website compromise, social media account compromise, impersonation, fake emails, fake blogposts, fabricated website, manipulated photos and/or videos, fake press releases/statements, falsified quotes and forged letters (See Annex 1 for definitions of these techniques). The Crowdtangle tool was also used to establish if content remains available on Facebook. In two incidents, no traces of the original messaging or responses to them could be identified. These were coded based on Mandiant's reporting.

Whilst a significant attempt has been made to collate as much detail as possible on Ghostwriter's diverse range of activities, undoubtedly, publicly available open source data can only provide a partial view. Some activities are probably partially attributed but not public, whilst others are not linked at all. Thus although it is highly likely that the incidents described and discussed in subsequent sections do not reveal the full scope of the actor's activity, they are sufficient data to construct insightful inferences and conclusions about the patterns of behaviour that provide a unique signature for the Ghostwriter Campaign.

To guide the enquiry, the following research questions were defined:

- How has the Ghostwriter campaign evolved and adapted over time, and what evidence is there to connect the multi-modal activities that have been conducted, in order to produce a more comprehensive assessment and understanding?
- Why has the Ghostwriter campaign been able to circumvent and limit the impacts of the interventions and interdictions directed towards it? How has the interplay between the responses and the hostile actor evolved?
- What are the implications of the insights and evidence generated via this case study for our understanding of contemporary (dis)information operations and attempts to control and constrain their impacts more generally?

7 https://www.fireeye.com/content/dam/fireeye-www/blog/pdfs/unc1151-ghostwriter-update-report.pdf

EVOLUTION OF GHOSTWRITER AS AN INFLUENCE CAMPAIGN

Analysis commenced by mapping out all publicly available information about the incidents attributed by Mandiant; as well as the responses to them, covering the period Summer 2016 through until Spring 2021. Based on this temporal mapping, three key phases of evolution where the campaign's behaviour changed were identified:

- **Phase 1:** summer 2016-January 2020 is the period involving early tactics, consistent and repeating incidents, mainly targeting NATO's presence in the Baltics.
- **Phase 2:** the campaign expands to targeting Poland in 2019/2020.
- **Phase 3:** late 2020 it starts exploiting hacked social media accounts more frequently, also extending its geographic reach, and utilising a greater repertoire of tactics.

Consistent with what is known about researching other forms of recurring deviant and transgressive behaviours, when researching actors conducting disinformation operations, investigating their earliest activities is often especially revealing. For it is when they are setting out on their careers, that they tend to focus upon their core motivations and use tactics that they feel especially confident with. So it is in the case of Ghostwriter, which in its early stages was almost exclusively:

- Targeting NATO's presence in the Baltics;
- Using Lithuanian, Latvian and English languages, but in parallel, also Russian in a coordinated manner;
- And employs blog platforms and forums, hacked media websites and email distribution as the way to convey its messages in the Baltics, and in Russia.

NATO'S PRESENCE IN LITHUANIA AND POLAND AS THE MAIN TARGET (2016-2020)

The first incidents tied to the Ghostwriter campaign coincided with the NATO Summit in Warsaw in Summer 2016. The official decision of NATO's enhanced forward presence in Estonia, Latvia, Lithuania and Poland was announced, including that four battalion-sized battle groups were to be established in early 2017[8]. Subsequently, a pattern emerged whereby attributed incidents repeated several times a year, all coinciding with NATO's military exercises, or high-level NATO visits to the countries.

The evolution of Ghostwriter's tactics during this time targeting Lithuania, focused on improvement of the quality of falsified content: writing, formatting, visuals and knowledge of local context. It also expanded its hacking attempts to target several media outlets at the same time, as described in the following case study.

8 https://www.nato.int/cps/en/natohq/official_texts_133169.htm?selectedLocale=en

CASE STUDY:

LITHUANIAN INTERNET PORTAL WAS ATTACKED 7-8 TIMES IN LESS THAN TWO YEARS

Brigita Sabaliauskaitė, former editor of internet portal Kas vyksta Kaune, interview in June 2022:

"It was in 2018 when Lithuania's military strategic communication team called me and asked if we had published fake news. I was in huge stress and ran to the office. The false message had been hidden already some days ago within the content of our news portal, it was difficult to find it. It had no shares on Facebook and about 200 people had read it.

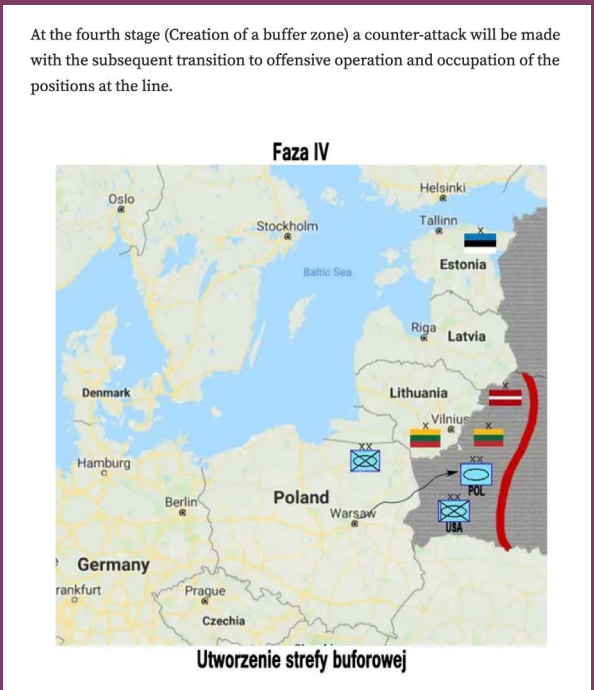


Figure 1. Screenshot of a false message posted by Ghostwriter persona Paul Black on Medium.com

The claim was that Poland and Lithuania will take part in a NATO excersise and will attack Belarus. The fake was prepared and published beforehand to wait for the exercise to start, and then the attackers could start sharing the fake.

Later on, we investigated what had happened. We still don't know who exactly was behind it. The National Cyber Security Centre helped us to close the vulnerabilities in our systems, but eventually only replacing our old publishing system with a new one helped and the hacks stopped. Before that, we were attacked 7-8 times during 1,5 years.

First time it took 10 days for us to notice it and take the false piece offline. In the last cases, it took only 30 minutes. The last incident was a false claim that Lithuania's first COVID-19 case was a US army officer.

In the beginning the quality of the fakes was poor. They used different font than our publication, and the photos were of different format. Later on, there was a clear improvement in the quality of the fakes. They had very good knowledge of the local context, used native Lithuanian language, and good photos. During a time when there were problems with water supplies in Kaunas, they spread false claims that a nuclear bomb had been dropped to Kaunas during a NATO exercise, and that is why the water had been cut off. Which means they had very good knowledge of local context and current affairs.

Later on, the incidents didn't target only us, but grew into a series of incidents with different targets. We were only one target and part of a much bigger campaign."

The below graphic shows how the use of Lithuanian language in Ghostwriter's messaging was consistent throughout the analysed period. Targeting audiences in Poland only started years later during 2019/2020, and again a similar gradual improvement in the quality of Polish language usage can be observed there[9]. Russian language was routinely used when the campaign was only starting to operate, but its role faded during the later phases.

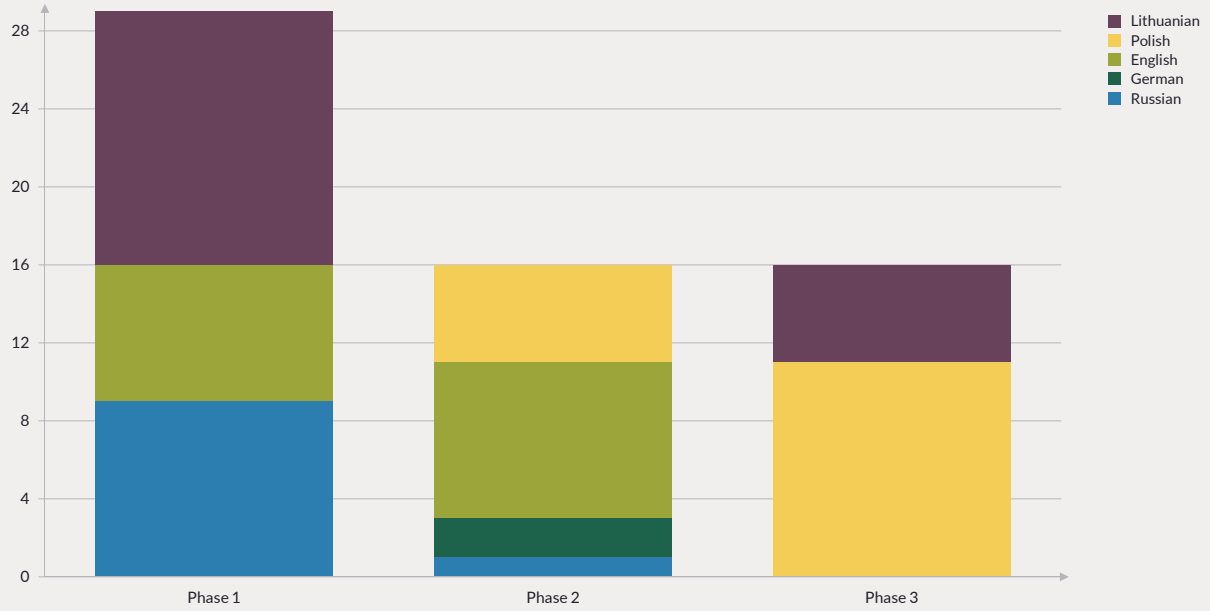


Figure 2. Languages used in each Ghostwriter phase.
**One Ghostwriter incident may involve several languages.*

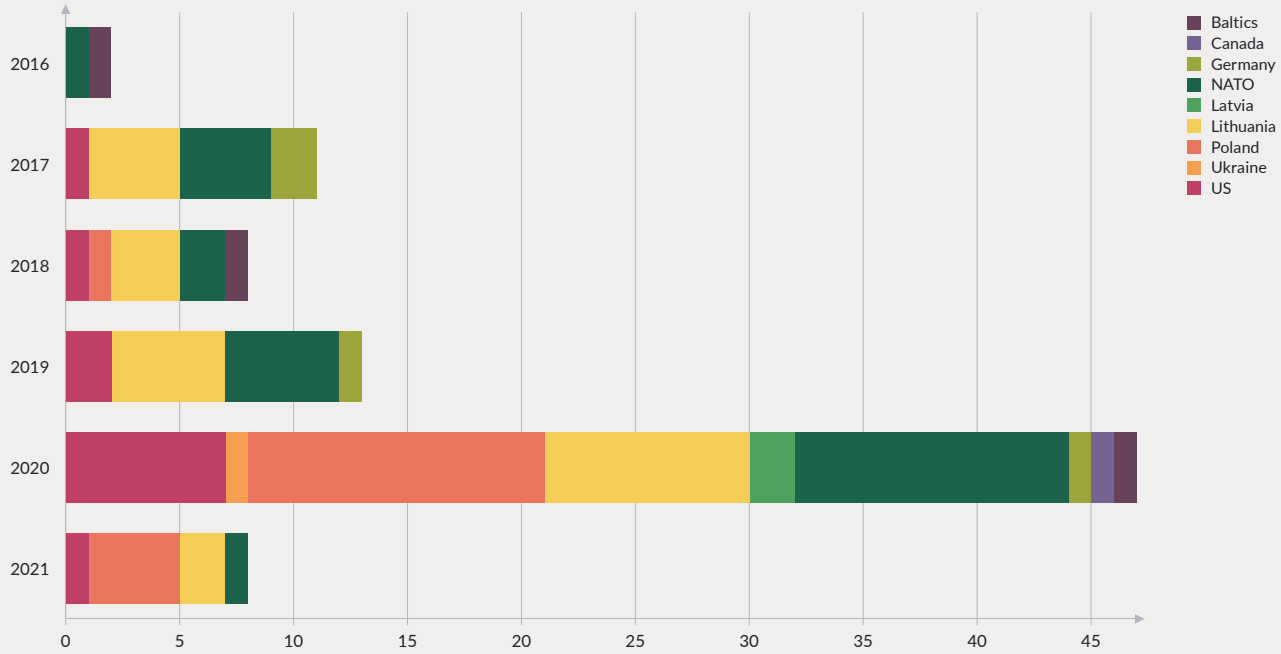


Figure 3. Targets of Ghostwriter influence operations.
**One Ghostwriter incident may involve several languages.*

NATO's presence in the Baltic countries and especially in Lithuania is the most often targeted focus of influence operations under Ghostwriter. Poland became a prominent target in 2020, whilst Polish-Lithuanian relations were a focus as well, and NATO continued as a main target.

9 https://cert.pl/posts/2022/07/techniki-unc1151/

CAREFUL TIMING OF THE OPERATIONS

One of the characteristics of the operations conducted by Ghostwriter is that they are especially carefully timed. They typically occur in the lead-up to and during big military exercises or high-level visits. The timeline below highlights real events that have coincided with the incidents, and the background image shows an example of suspected Ghostwriter messaging relating to this.

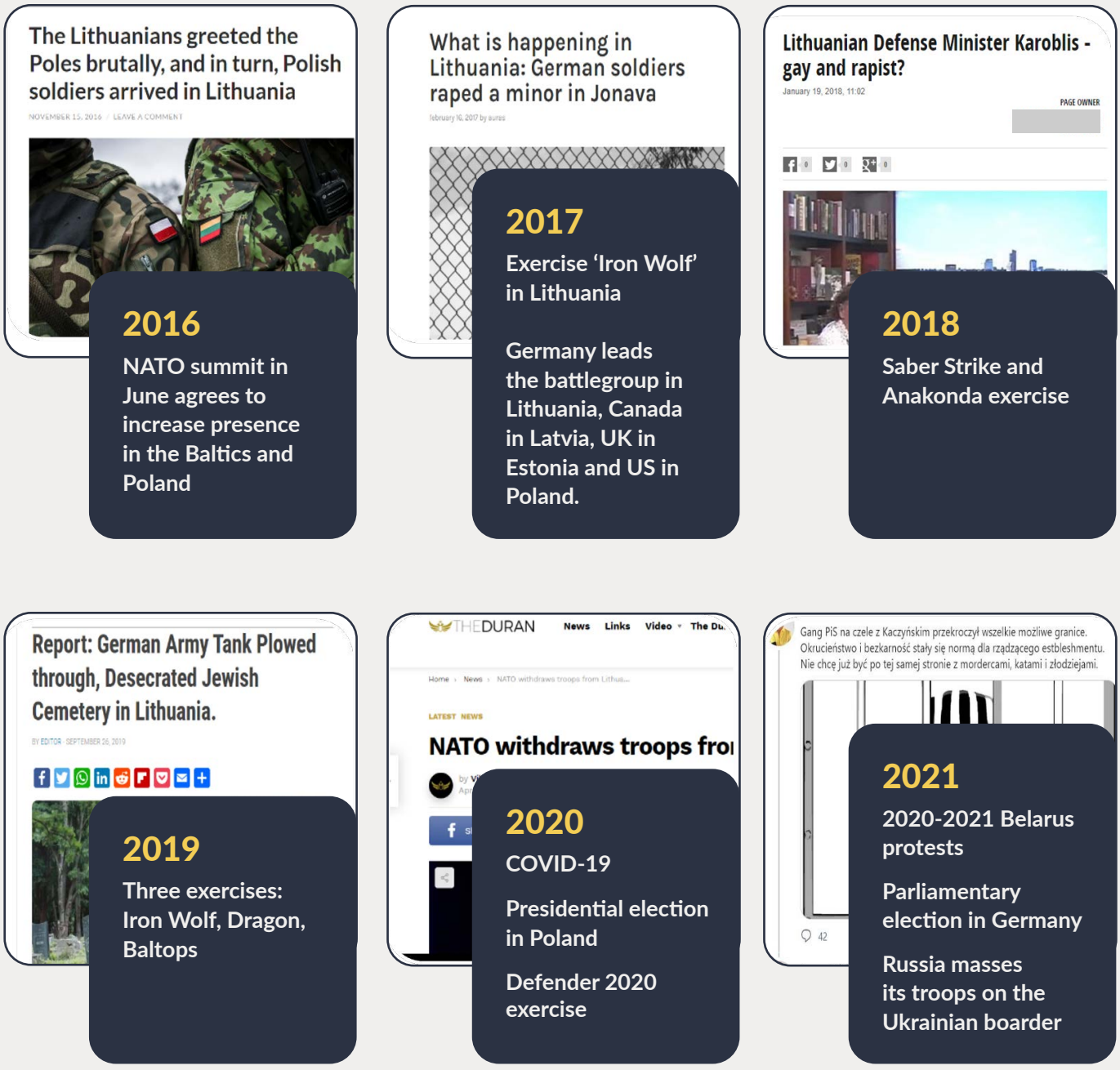


Figure 4. Timeline of real events coinciding with the Ghostwriter incidents.

CASE STUDY:

"GERMAN SOLDIERS DESECRATED A JEWISH CEMETERY" – GHOSTWRITER'S ACTIVITY EXPANDS

Živilė Didzgalviene, advisor, Strategic Communication Department of the Lithuanian Armed Forces. Interview in June 2022:

The false claim was about German soldiers desecrating a Jewish cemetery. It was in September 2019, and the operation focused on perfect timing. Lithuania's president and foreign minister were travelling to New York for the UN General Assembly, and separately held meetings with the Jewish community there. In addition, the US had just announced it will increase its troops in Poland and Lithuania. So several audiences were targeted: the Jewish community in the US was targeted to disrupt the meeting, question Lithuania's policies in the eyes of the US decision makers, and to discredit NATO forces in Lithuania.

A special website was also created, where photoshopped photos of the cemetery were published. Emails were sent to the government to spread the message further. Lithuania's Jewish community was the first one to see it and flagged it to us. We started calling media to inform them about the fake.

The next day it continued. A new fake was published on the hacked Kas Vyksta Kaune news portal that the Lithuanian Government and the Ministry of Defence are hiding the truth about German soldiers' behaviour in the cemetery. Jewish press picked it up, and a petition was created to "condemn the act". We asked our Jewish community to help, and they informed editors in chief in the Israeli press about the fake. In the end, the petition was removed, articles referencing the fake were taken down, and the Lithuanian media did not pick it up. The President's meetings in New York took place and the US soldiers were deployed.

Figure 5. Screenshot from French-language Infos-Israel.news that corrected its reporting about the incident. Machine translated into English from French.

The end result was good, and the operation didn't manage to achieve damage. But as an operation it was well prepared. There might have been intelligence gathering beforehand, as the President's meeting agenda was not public. The same with sending fakes via emails to government employees – it is possible for the attackers to follow the forwarding route of an email exchange, and then conclude who oversees the response and how it connects with the president's office. It is advanced manipulation.

It is better to be proactive than reactive. If you wait for 1-2 hours, the damage might have been done."

FROM HACKING WEBSITES TO HACKING SOCIAL MEDIA ACCOUNTS

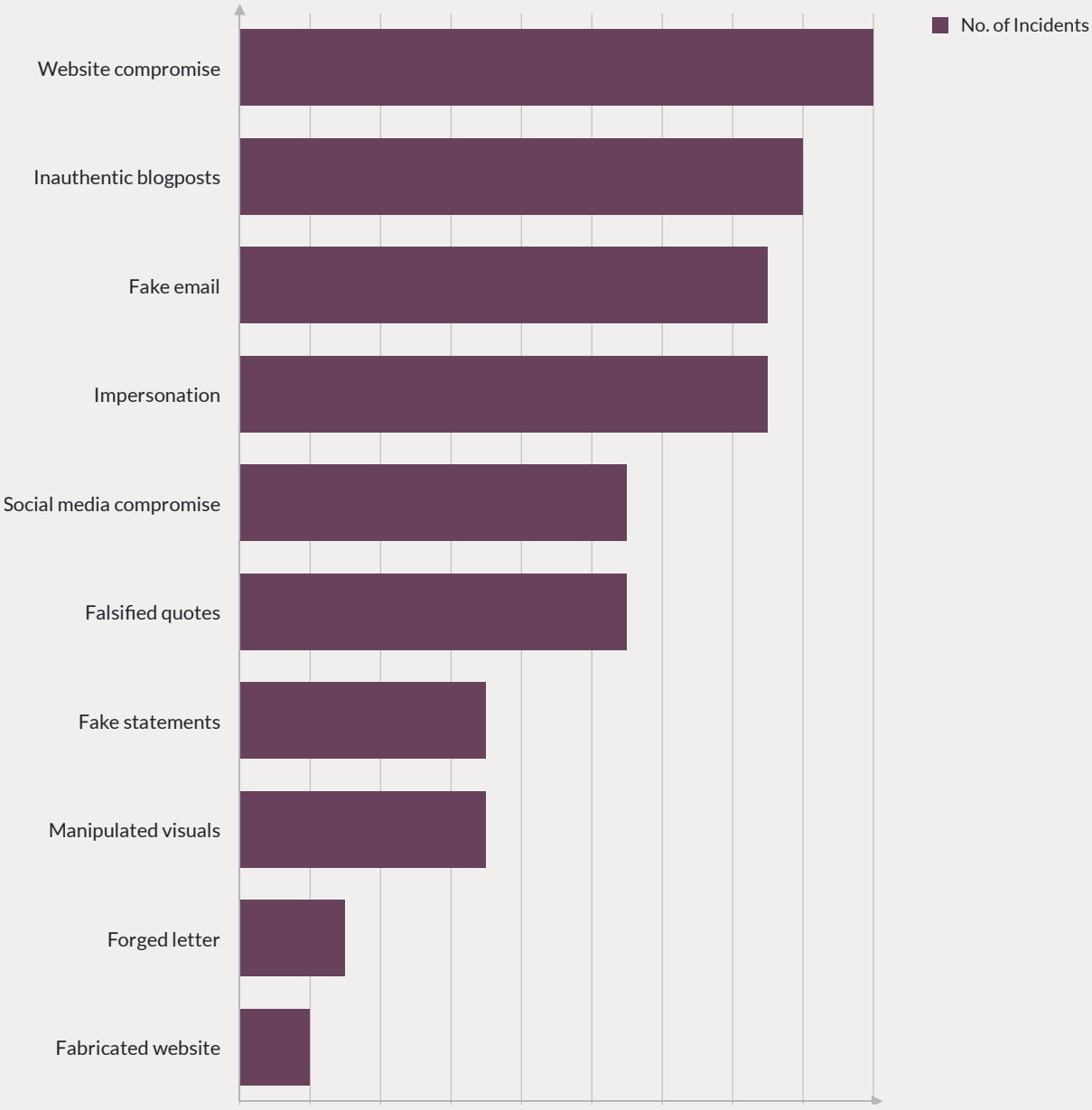


Figure 6. Techniques most often used by Ghostwriter.

The above graph shows different types of influence techniques Ghostwriter has exploited in the campaign (See Annex 1 for definitions). In the first phase (2016-2020) it relied mostly on hacking media websites, and using blogposts by inauthentic personas, as well as fake emails to distribute and deliver the messages to audiences. In the second phase (2019-2020), the use of blogging platforms and spoofed emails continued. Only in the third phase (2020-2021), did the campaign start exploiting compromised social media accounts to convey the messages.

CASE STUDY:

GHOSTWRITER EXPANDS TO SIMULTANEOUS ATTACKS IN POLAND AND LITHUANIA

Anna Gielewska, vice-chairman at the Reporters Foundation (Poland); Head of Investigations at VSQUARE.

”In the autumn 2020, news broke out that several Polish politicians had been hacked and the accounts were posting some ridiculous content. Literally, no one treated it seriously at the beginning. Those politicians were either drunk and not really hacked or - if hacked, that was some joke. That was the perception back then – partly because the targets were PiS politicians and represented the government. But when we started following it up and investigating it, we realised that these attacks were really complex: for example, one fabricated website of Lithuania’s State Nuclear Power Safety Inspectorate reporting about a nuclear disaster in Lithuania, then two hacked websites of Polish institutions where false information was published, and Twitter accounts of an expert and governmental officials hacked and used to further distribute the fake. It became clear this was happening on a large scale, and simultaneously in Poland and Lithuania.

Figure 7. Screenshot of an analysis by Vsquare[10] on the operation that targets both Lithuania and Poland, involved several hacked government agencies’ websites, and hijacked Twitter accounts.

Later the social media account hacks continued. In June 2021 the Head of the Polish prime minister’s office, Michal Dworczyk’s mailbox was hacked. He had been using his private email for sensitive issues. A fabricated post was published on Dworczyk’s wife’s hijacked account, promoting a Telegram channel where the leaked content was published.

The content was well adapted for Polish audiences and created by people with deep understanding of the Polish political backstage and how to trigger polarisation. In parallel, Belarusian government propaganda made use of the email content. Another Telegram channel in Russian was also set up already in February”.

10 https://vsquare.org/privet-you-have-just-been-hacked/

ATTEMPTS TO CONTROL AND LIMIT GHOSTWRITER'S ACTIVITY

In the preceding sections, we highlighted how Ghostwriter's influence tactics have evolved. Whilst it relies on similar, characteristic behavioural patterns over the course of the years, the complexity of its operations increases over time. At the end of our timeline in 2021, it had started using longer and more complex distribution chains to escalate the operation, engaging simultaneous targeting of countries (Poland and Lithuania), and had moved to use new social media platforms (Twitter and Facebook account hacking, Telegram distribution).

In this section, we move to highlight the main control strategies that have been used to try and limit the impact of these incidents, as well as the changes in Ghostwriter's tactics these strategies have induced.

STRATEGIC COMMUNICATION: RAPID, PUBLIC REACTION

Lithuania and specifically NATO's presence there was the main target of Ghostwriter operations especially in the beginning. Interviews with Lithuanian Armed Forces strategic communication staff and analysts, as well as journalists, indicate a good level of preparedness and awareness to tackle cyber-enabled influence operations. Lithuania's approach relies on effective monitoring of the domestic information space, rapid assessment, coordination and reaction capabilities, as well as straightforward cooperation between the government, media and civil society. The overarching idea being that each of the parties should flag potential incidents or signals of a campaign to one another.

Among the interviewees, there was a shared understanding of the necessity to react fast. This capacity for a rapid, public reaction has most likely managed to reduce Ghostwriter's ability to escalate the spread of false messages at the later stages of an operation.

The issue with this model is that it is not easily replicable in other countries. This is because the Government's role in monitoring the domestic information space is frequently politically sensitive, and cooperation between government, media and civil society not always straightforward. As Brigita Sabaliauskaitė, former editor of Kas Vyksta Kaune described in an interview, some of the friction that arises between media and governmental organisations were also present in Lithuania:

"We were a news outlet startup which has grown since. We didn't get any state support for our investment in new security systems. I explained to our government officials how difficult it was to find resources. The cyber centre said that they could as well close our portal because we were threatening Lithuania's national security. I just answered that we are also victims in this war."

An additional consequence of rapid public debunking is that it makes the organisation being victimised responsible for presenting evidence behind the call-out. Moreover, it can amplify the story itself. Symptomatic of which, some of the journalistic reporting saw visible public reactions and debunking the incidents as problematic:

*"A news story that was horse***t and not worth reporting is now being commented on in a very serious way by influential public figures and officials with thousands of followers, who are even providing links to it. And thus, whether we like it or not, it becomes newsworthy."*[11]

11 https://eng.lsm.lv/article/features/commentary/opinion-the-fake-news-catch-22.a357224/

INVESTMENTS IN CYBER SECURITY DETERRED FURTHER ATTACKS

Repeated hacks by Ghostwriter eventually pushed the Lithuanian media to invest in their cyber security. News portal Kas Vyksta Kaune was repetitively hacked over 1,5 years, but investing in a new publishing platform prevented further attacks. As a repeat victim of Ghostwriter's activities, by 'target hardening' to reduce their vulnerabilities, they were able to help mitigate the threat.

Lithuanian TV channel TV3 had a similar experience. It was hacked in 2018 and managed to remove the false article from their website in minutes. The fake story was about Lithuanian Defence Minister Raimundas Karoblis allegedly harassing journalist Ridas Jasiulionis:

"After the article was removed, I called the defence minister as well as the journalist and apologised for what happened."

TV3's editor in chief Artūras Anužis recalls his reaction. The TV channel started cooperating with Lithuanian national cyber security authorities to find out what had happened. After a year of investigation, it was still unclear who was behind the hack, although it was clarified that hackers had access to TV3's administrative system for a longer period:

"We cleaned the system, verified all the accounts we had in there, and introduced a two-level identification for logins. After that, we haven't had similar attacks", Anužis says.

Journalist Vaidas Saldžiūnas, Defence Editor of news portal Delfi, agrees most media outlets simply don't have sufficient money to invest in cyber security. Those media outlets that did manage to find the resources, have been saved from further attacks. For the government, he gives 6/10 points in tackling the threat:

"Governments are slow to react, analyse and learn. Time, skills, and the nature of attack along with diverse people working and becoming victims are the limitations and always will be, unless there's a huge attack", he adds.

Saldžiūnas himself was targeted with an impersonation attack. In October 2019, an email in his name was sent to several recipients, including the President's office and an air force base. The fake message claimed the Lithuanian President asked the US to deploy nuclear weapons in Lithuania. The email created a false impression that it was Saldžiūnas asking the respondents to explain what is going on:

"The institutions, despite in doubt I wrote the letter, had some time and, I'd say, stupidity to open the link in that letter. As previous examples have shown, it might have been infected with malware. For example, the parliament responded to me in a week after sharing that spoofed email among themselves."

While Saldžiūnas believes that the awareness of the threat and general online skills of users have improved during the past few years, he also emphasises that to be effective Ghostwriter need just one weak human link:

"The Ghostwriter incidents are stupid, primitive or mildly sophisticated stories with clear disinformation signs all over them. This is just the outer layer, the real danger is who clicks the link to the attachments, and who is the weak link to be identified for potential future ops."

According to the interviewees, the continuous need to respond to Ghostwriter's attacks have not only increased awareness among journalists and government officials, but also among local audiences that were targeted. Brigita Sabaliauskaitė, former editor of news portal Kas Vyksta Kaune, describes how the outlet's

reaction time decreased from ten days to 30 minutes. She also insisted on informing the readers every time a new incident had occurred:

"After several incidents, we were criticised of being either losers or liars – how could there be so many attacks on us? But I didn't change my mind and we continued informing our readers. Soon the readers themselves learned to find and flag us if an incident had happened, so we used this to practice some media literacy. But I don't think they stopped attacking us only because we changed our security systems. They changed their strategy."

Sabaliauskaitė describes in an illustrative way the dynamics between Ghostwriter's attacks and the responses. Once the most vulnerable news portals had invested in their cyber security in Lithuania, the attacks prioritised a different location and different platforms – they moved to hack Polish politicians' social media accounts, as well as media and governmental websites.

DEPLATFORMING EFFORTS

Public reaction by big tech platforms such as Facebook and Google to Ghostwriter came only after Russia invaded Ukraine in February 2022, six years after the first known operation within the campaign. According to Anna Gielewska, Head of Investigations at VSQUARE: "Facebook only reacts when the milk is spilled",

Lithuania's rapid response probably reduced Ghostwriter's possibilities to grow on social media platforms. Even if the campaign's main objective was not to spread the false messages as widely as possible, it did attempt to gain a presence on social media platforms. Later, this enabled the campaign to exploit new vulnerabilities in Poland via hacking real, existing social media accounts.

Ghostwriter's characteristic patterns of behaviour differ significantly from those of the Internet Research Agency. As the major public pressure towards the platforms came largely from the need to tackle the interference in the US election in 2016, the IRA's tactics are the ones that initially started directing social media platforms' policies against "inauthentic coordinated behaviour", or more generally on takedowns, demotion and labelling of content. The effect the IRA had on American users was often described and best understood in terms of the operation's reach, whereby: "Over 30 million users, between 2015 and 2017, shared the IRA's Facebook and Instagram posts with their friends and family, liking, reacting to, and commenting on them along the way", as one of the first reports revealing the scale of IRA's activities in the US stated[12].

A similar challenge with defining the impact of a foreign influence operation can be seen in how Ghostwriter incidents have been assessed during recent years. Especially in the first years of the campaign – when there was no clear understanding that these ostensibly separate incidents were connected – the challeneg of assessing the potential harm was especially pronounced. The Atlantic Council's DFRLab assessed in 2020 that the overall engagement with Ghostwriter false narratives about NATO and COVID 19 remained rather low on social media.[13.]

Like the social media platforms, governments and institutions were claiming that influence operations have only limited impact: NATO assessed in 2020 that one of the campaigns was carefully planned and coordinated, but failed to gain significant interest online[14].

12 https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/The-IRA-Social-Media-and-Political-Polarization.pdf

13 https://medium.com/dfrlab/fact-checkers-identity-stolen-to-spread-disinfo-about-nato-and-covid-19-111a2eef70a0

14 https://www.nato.int/cps/en/natohq/177273.htm

Similarly, another DFRLab analysis uses Facebook's CrowdTangle tool to assess the engagement figures for an incident where the false claim was about a radioactive leak in Lithuania endangering people's lives in Poland.:

"However, engagement on these posts was almost close to zero; a third-party fact-checker promptly labeled the story as false information, and this label appeared on posts containing the links in the Facebook groups."

Framing the IRA's large engagement figures as the primary way to measure the impact of an influence operation may have played a role in helping to undermine the cumulative threat and risk Ghostwriter poses to social media platforms over several years.

However, when we look at how Ghostwriter has escalated activity on the social media platforms, it has been able to exploit multiple vulnerabilities. For example, by getting targets clicking on phishing emails, then getting login information to their private emails and social media accounts, thereby creating an opportunity to hijack the accounts to spread false messages.

Although Facebook has updated its threat reporting from inauthentic coordinated behaviour policies to also cover cyber espionage, mass reporting, inauthentic amplification and brigading, there is no systematic overview for Ghostwriter's activity and related disruptions or takedowns on Facebook, or any other platform. Therefore, it is impossible for users to get reliable information on what has been the scale of Ghostwriter's activity on Facebook, or which accounts have been hacked. Facebook has recently started covering cyber espionage in its quarterly threat reports.[15]

In the report from August, Facebook notes it has acted against hacker groups in South Asia and Pakistan, and that it "took down accounts, blocked their domain infrastructure from being shared on our services, and notified people who we believe were targeted by these malicious groups". The users are required to rely on Facebook's own reporting about the threat and the response to it, and verifying these statements is very difficult. Like in the case of IRA, providing historical data to the public of a certain threat actor's activity would help to build more systematic ways of disrupting their current malign actions.

Relatedly, according to a whistleblower,[16] Twitter has seriously neglected its cyber security policies which could allegedly open the door to foreign spying or manipulation, hacking and disinformation campaigns. For example, thousands of the company's employees had access to some of the platform's critical controls, and if a user cancels an account, the account data is not reliably deleted. This leads us to a related problem when our aim is to understand the evolution and limitations of responses to Ghostwriter.

ATTRIBUTION: GRADUAL BUILD-UP OF PARTIAL ATTRIBUTION

Multiple actors have so far attributed different elements of several cyber-enabled influence operations to Ghostwriter or UNC1151. This attribution to Ghostwriter has developed and been built gradually. While parts of Ghostwriter's cyber activity have been attributed to Russia's military intelligence (by Germany) and to the Russian state (by the EU and Poland), as well as to Belarus (Mandiant and Google), there is only limited public knowledge of who is conducting the influence operations and content production that underpins Ghostwriter's 'footprint.'

15 https://about.fb.com/wp-content/uploads/2022/08/Quarterly-Adversarial-Threat-Report-Q2-2022.pdf

16 https://edition.cnn.com/2022/08/23/tech/twitter-whistleblower-peiter-zatko-security/index.html

One framework to understand attribution has been proposed by Helsinki Hybrid COE and NATO Stratcom COE, which helps in understanding differences between different actors that have attributed Ghostwriter:[17]

	Technical Evidence	Behavioural Evidence	Contextual Evidence	Legal & Ethical Assessment
Open Source	Web domain ownership, IP addresses, economic ties	Account activity, page activity, posting/cross-posting, sharing, follows, network	Media content, discourse and narratives, linguistics, political context, cui bono	Risk of litigation; research ethics; personal risk of becoming a target
Proprietary Source	Data collected by platform backend	As above, with more extensive platform data	As above and data on previous takedowns with suspected links	Protecting political and commercial interests; data protection
Classified Source	SIGINT; proprietary source data acquired by warrant	As above and SIGINT, HUMINT	As above and classified geo-political assessments	Actor-specific stragety; protecting political interests; data protection

Figure 8. Attribution framework by Helsinki Hybrid COE and NATO Stratcom COE.

In the case of the IRA, the attribution was also built gradually over time. It has been designated to a specific entity (IRA and related organisations), and they have been held responsible for a series of linked actions, as detailed in US Treasury's sanction designations[18] as well as on social media platforms' takedown reports[19]. Ghostwriter differs from this, as specific incidents have been attributed to one actor only by a private company, Mandiant, and its attribution covers only the cyber espionage part of the malign activity. Unlike in the IRA's case, the ultimate source of the content creation remains unclear.

Comparing the ways to attribute that different actors have practiced in relation to Ghostwriter with the above table, highlights the discrepancies and weaknesses of the current system for countering information operations. First, most attributed information falls under classified sources, to which the public does not have access. Most of the public callouts have been seemingly designed to deter further attacks. Germany connects Ghostwriter's cyber activity targeting Germany to Russia's military intelligence. The EU talks about some EU Member States that "*have observed malicious cyber activities, collectively designated as Ghostwriter, and associated these with the Russian state*." It does not specify which member states are engaged. Moreover, these callouts have not been followed up with further measures even if the Ghostwriter operation is an ongoing threat.

Secondly, proprietary sources (private companies) require the public and their users to accept their attribution at face value. While Mandiant did raise public awareness about Ghostwriter by connecting the different incidents as part of one campaign, some relevant information is still non-public and has not triggered other parties (governments or social media platforms) to do the same. Facebook started talking publicly about Ghostwriter on 27th of February 2022, three days after Russia's invasion of Ukraine. It does not explain how Facebook detects Ghostwriter's activities, state who is behind this threat actor, or what Facebook has done to counter it. No visual examples are published about the attributed posts, albeit the following description is given:

"We detected attempts to target people on Facebook to post YouTube videos portraying Ukrainian troops as weak and surrendering to Russia, including one video claiming to show Ukrainian soldiers coming out of a forest while flying a white flag of surrender."

Thirdly, open-source attribution has been a limited addition in terms of linking the different actors – foreign and/or domestic. Ghostwriter's efforts to obscure the origins of its malign activity has been rather successful.

17 https://stratcomcoe.org/pdfjs/?file=/publications/download/Nato-Attributing-Information-Influence-Operations-DIGITAL-v4.pdf?zoom=page-fit

18 https://home.treasury.gov/news/press-releases/jy0899

19 https://about.fb.com/news/2018/04/authenticity-matters/, https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html

It has been openly running the operations in Russian language and on Russian blogging platforms and has not tried to avoid all connections to Russia. State media in Russia, however, were careful not to directly amplify Ghostwriter's fakes. Instead, such media have focused on amplifying the rebuttals of the incidents, accompanied by ridiculing Western governments, accusing them of Russophobia.

The nature of this partial attribution has contributed to the campaign's successes in Poland. There is distrust that politicians' social media accounts would have really been hacked, and especially the leaked emails of Polish prime minister's adviser, Michal Dworczyk, gained weight in the domestic debate. According to one EU official working in countering disinformation:

"The hack and leak campaign has been very embarrassing for the government".

Currently, there is still no common consensus and clear attribution for who was behind the hack and leak part[20] of this operation:

"After we obtained technical data on hacked Dworczyk's mailbox, we asked Mandiant's expert for additional analysis. For our request, Mandiant's expert did assess with high confidence the hack was part of UNC1151 activity, which also serves the "Ghostwriter" operation." Anna Gielewska, Head of Investigations at VSQUARE, says.

The Polish secret services and later government did attribute the attack to Russia[21]:

"But many respected independent media did not buy that announcement. Because PiS was targeted, they questioned the accusation", Gielewska continues.

Many of those interviewed for this research, agree that the operations are probably run by several different groups. With different parts of the operation, from cyber-attacks, potential target surveillance and content creation is probably conducted by different networks, potentially not fully aware of each others' activities. The interviewees conclude that parts of the operations require local language and political knowledge, suggesting proxies in each of the countries are involved.

Lithuania has not attributed separate incidents but treats them as foreign hostile state operations. As Tomas Ceponis, senior specialist, Counter hybrid response group at the Ministry of National Defence of Lithuania, states:

"We look at the attribution differently and ask: To whom is it useful? Are there other indicators that are aligned with Russia and Belarus? Are there complex resources behind it? State actors like Russia conduct complex operations with several levels from preparation of webpages, social media accounts, cyber-attacks, creating video, audio and visual content, gathering intelligence, using fake email addresses, then in the final stages escalate on social media the spread. These operations are not a joke",

Ultimately, he concludes:

"If they succeed in one place, they will repeat it elsewhere".

A rapid stratcom response, such as debunking or removing false content, is possible based on threat assessment without attribution. However, to move from refuting falsehoods and taking down false content into raising the cost for the adversarial behaviour, more comprehensive attribution is required:

"But sometimes we don't respond. It is better not to show your opponent all what you know. These operations are designed for the long-term effects. We can also aim at building ambushes in the future, when the opponent might wait results in five years", an analyst from Lithuania's Armed Forces stratcom team summarised.

20 https://vsquare.org/behind-the-hack-and-leak-scandal-in-poland/

21 https://apnews.com/article/russia-ukraine-poland-judiciary-warsaw-a4e37e00c14e337f853ec1c9384d4b26

CONCLUSION AND RECOMMENDATIONS

Ghostwriter has, despite the responses from states, international organisations, private companies, and civil society, managed to expand and evolve its malign activity. Especially significant in its evolution was the move to use phishing emails, acquire politicians' and others login information, and publish false information through hacked social media accounts. These moves proved especially impactful in Poland.

The responses to date have largely focused on rapid callouts alerting about false information. These alerts and warnings have come as coordinated responses from the EU, as well as NATO. Germany also used diplomatic channels warning Russia from continuing its cyber-attacks ahead of the German federal election[22]. In addition, responses have focused on updating security systems against cyber-attacks and hacks, and blocking the phishing attempts. As a positive side-effect, some local audiences of hacked media in Lithuania have developed better skills in identifying and reacting to false information.

Different proposals have been developed to create more effective policies focusing on deterring influence operations. Drawing on criminology, disruption, displacement, and deterrence could be used in countering influence operations[23]. Policy interventions have been crafted for the EU that could be used to build cumulative deterrence against influence operations and foreign interference[24]. From these many available options, designating sanctions or prosecuting those responsible for Ghostwriter have not been used as part of the response. The EU has not put its cyber sanctions toolbox into use and designated sanctions on those responsible for Ghostwriter's activity. Potentially, this is due to a lack of intelligence sharing, or because of a lack of robust knowledge of the specific operators behind it[25].

No countries, at least based on public knowledge, have used offensive cyber capabilities to target Ghostwriter's activities. Likewise, none of the social media platforms have public reporting on Ghostwriter's activities on their platforms equal to that provided for the IRA, where the public pressure for accountability led the companies to change their policies.

At least part of the problem that has inhibited the development of effective responses to the Ghostwriter Campaign derives from what criminologists refer to as 'linkage blindness'. Developed to describe an intelligence problem that arises when multiple police forces are unwittingly investigating individual incidents that are part of a connected series, their 'blindness' relates to how they are unable to perceive the key connections and patterns. This captures very neatly the problem that has arisen in terms of understanding the scope, scale, and evolution of Ghostwriter's various individual operations. Different countries have researched and responded to those attack vectors directed at them. Similarly, individual platforms have also highlighted particular aspects, but are understandably principally concerned with defending their surfaces. As a result, it is difficult to discern who precisely has the responsibility for piecing together the various bits and pieces of the jigsaw to construct a more holistic analysis and understanding.

Likewise, given the multi-vector nature of the attack methodologies, that innovatively blends hacking with information manipulation, it would seem reasonable to suggest that a multi-dimensional social control response is warranted. But again, it is not clear where the centre of gravity for coordinating such a response lies.

22 https://www.auswaertiges-amt.de/de/newsroom/regierungspressekonferenz/2480282#content_4

23 https://carnegieendowment.org/2020/10/28/using-criminology-to-counter-influence-operations-disrupt-displace-and-deter-pub-83058

24 https://carnegieendowment.org/2020/09/30/eu-s-role-in-fight-against-disinformation-developing-policy-interventions-for-2020s-pub-82821

25 https://www.lawfareblog.com/after-year-silence-are-eu-cyber-sanctions-dead

Framed by such considerations, this report has sought to take a first step in developing a 'richer picture' understanding of Ghostwriter, based upon publicly available evidence. In so doing, we conclude with the following set of recommendations:

RECOMMENDATIONS:

- Cyber-enabled influence operations require more holistic understanding on the part of Western governments, social media platforms and civil society. Currently, cyber and influence operations are understood as separate fields, with distinct sets of expert knowledge. At the same time, adversaries often don't make similar distinctions between the two. It may be more convincing to the public to both debunk/deny and present 'hard' cyber evidence together, as the former is more reliant on people's trust in institutions/media.
- The tactics used by the Internet Research Agency have shaped, in a profound way, democratic responses to foreign interference and influence operations. Expanding understanding to operations like Ghostwriter, that are using different playbooks, is crucial for shaping future and more effective responses.
- Ghostwriter's partial successes in evolving and expanding its activity is likely due to a mix of limitations in attribution; intelligence sharing; varying level of understanding; requiring responses to the threat in different countries to involve different political actors, as well as social media platforms.
- The social media platforms' transparency practices related to takedowns and policies regarding specific threat actors should be developed to cope with multi-vector attack methodologies. Platforms could also consider publicising historical data from a specific threat actor's activity and its disruption to enable further research and preparedness for future attacks.
- Targets of website, email and social media hacks have significant trouble in communicating that they have become a victim of an attack. Social media platforms, service providers and government officials should make public statements to support the users and targets of the attacks to confirm that a hack has happened, or in the best case, warn the potential targets beforehand. Cost is a barrier for news media in upgrading their websites to reduce their vulnerabilities. More funding could be directed specifically to this aim, including for pen-testing. Prebunking could be considered as an option ahead of important political events, visits, and NATO exercises in Central and Eastern Europe.
- It is important that monitoring the evolution of malign influence operations is done continuously over time. Treating Ghostwriter's operations as separate, unrelated incidents, rather than part of a linked series, has contributed to weaker situational awareness and threat perception. To recover from these deficits has taken both time and effort and has probably increased the impacts and harms delivered by Ghostwriter's activities.

ANNEX 1

GLOSSARY OF GHOSTWRITER INFLUENCE TECHNIQUES:

Website compromise: threat actor has gained access to the website and typically publishes false content

Inauthentic blogposts: Dissemination of false messages on blogging platforms by suspected inauthentic accounts

Fake email: Dissemination of false messages via fake sender address, typically impersonating a real person or institution

Social media compromise: threat actor has gained access to a real person's social media account, typically publishes false content

Fake statements: Fabricated press-releases and other statements, typically made to look like coming from a governmental institution

Impersonation: Using real person's identity as part of the campaign, typically in fake emails or as authors of false content

Falsified quotes: Content includes false quotes from a real person

Manipulated visuals: Content includes photo or video manipulation

Forged letter: Content includes fake letters

Fabricated website: A whole website is fabricated, making it resemble the real one but typically changing the URL slightly

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Security, Crime and Intelligence Innovation Institute

Sefydliad Arloesedd Diogelwch, Troseddu a Chudd-wybodaeth

This report was funded by:

RESET: Digital for Good
www.reset.tech