978-9934-564-73-4

MANIPULATION ECOSYSTEM OF SOCIAL MESSAGING PLATFORMS

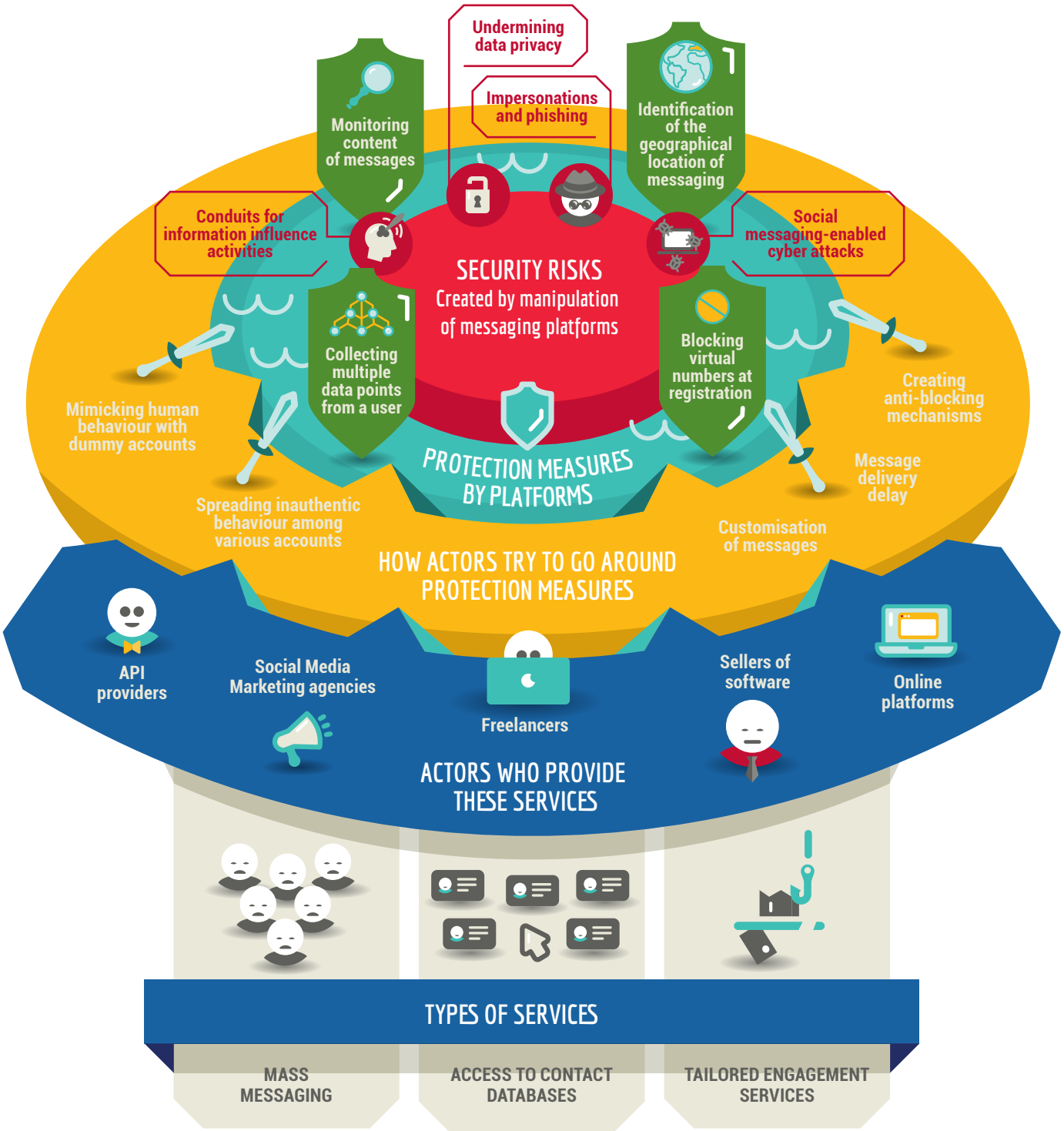
Published by the
NATO Strategic Communications
Centre of Excellence

ISBN: 978-9934-564-73-4
Authors: Rueban Manokara, Marina Paramonova
Research: Singularex
Project manager: Rueban Manokara
Copy-editing: Jazlyn Sierra Melnychuk
Design: Kārlis Ulmanis

Riga, April 2020
NATO STRATCOM COE
11b Kalciema Iela
Riga LV1048, Latvia
www.stratcomcoe.org
Facebook/stratcomcoe
Twitter: @stratcomcoe

TABLE OF CONTENTS

ECOSYSTEM OF SOCIAL MESSAGING MANIPULATION



ABOUT THE STUDY

Landscape scan of the manipulation tools and services available for WhatsApp and Telegram, finding as diverse a group of services on the open web.

Assessing the tools and services: evaluating the cost, methods and scale of manipulation, quality of manipulation tools and services, and the ability of WhatsApp and Telegram to identify and counter manipulation on their platforms.

AIMS OF THE STUDY

Mapping the online market for manipulation tools and services available for WhatsApp and Telegram.

Assessing the effectiveness of these tools and services against the protective mechanisms put up by the messaging applications.

Provide national institutions and communications practitioners with an overview of the scale and effect of manipulation on these platforms.

INTRODUCTION

Social messaging platforms started as an alternative to the Short Messaging Service (SMS), pitching themselves as faster and cheaper, with additional features such as the ability to send documents and media securely. These features granted users a level of encryption that meant no third party, including the messaging services themselves, was able to read the messages sent. Today, social messaging platforms account for a combined 4.1 billion users and social messaging has become the most frequent activity a person carries out online.[1] Similarly to major social media platforms that are being artificially inflated and manipulated for financial and political gain, social messaging platforms are equally vulnerable to the threat of exploitation.

This study maps the online market for manipulation tools and services available for two popular messaging applications: WhatsApp and Telegram. In doing so, we assess the effectiveness of these tools and services against the protective mechanisms put up by the messaging applications. This publication aims to provide national institutions and communications practitioners with an overview of the scale and effect of manipulation on these popular social messaging platforms.

In collaboration with Singularex, a Ukrainian social media analytics company, we conducted the study in two parts. The first part consisted of a landscape scan of the manipulation tools and services available for WhatsApp and Telegram. We searched the web and spoke to sellers and freelancers over a period of two months to understand what a customer, or a potential malign actor, can purchase online. Given that previous research on social media had shown no significant difference between social media manipulation services available on the dark web and the open web, we decided to focus our efforts on finding as diverse a group of services on the open web as we could.

The second part involved assessing the tools and services identified. This included evaluating the cost, methods and scale of manipulation available, the quality of manipulation tools and services, and the ability of WhatsApp and Telegram to identify and counter manipulation on their platforms.

Why WhatsApp and Telegram?

The social messaging platforms environment is diverse and highly fragmented, with clear regional favourites. For example, WhatsApp is the dominant player in Brazil and India. In China, primarily due to government restrictions and linguistic adaptations, WeChat is the most widely used messaging platform. In addition, different applications tend to specialise in different features as they aim to carve out their space in the market. For example, Facebook messenger makes it convenient to contact friends and family by syncing seamlessly with one's Facebook account. On the other hand, Signal, an increasingly popular application for a high level of encryption and security[2], has pitched itself to more security-conscious individuals.

WhatsApp is regarded as the leading mobile messaging application across 112 countries.[3] It last reported 1.5 billion active users per month and has the lion's share of market penetration in European countries such as the Netherlands at 85%, followed by Spain at 83.1%.[4] When compared to WhatsApp, Telegram is relatively modest, with 365 million downloads as of August 2019. Nevertheless, Telegram has growing ambitions with a target of 1 billion users by 2022.[5] WhatsApp and Telegram are not only interesting because of their position in the market, but also because of the inherent security risks associated with them, such as the case of fuelling violence in India[6], and the use of the platforms to spread propaganda and support the command and control structures of terrorist organisations such as DAESH.[7]

Both platforms have been in the market for a relatively long time and have developed extensive features that improve usability, but also create additional avenues for manipulation. In *Table 1 and 2* below, we trace the critical stages of development for both platforms.

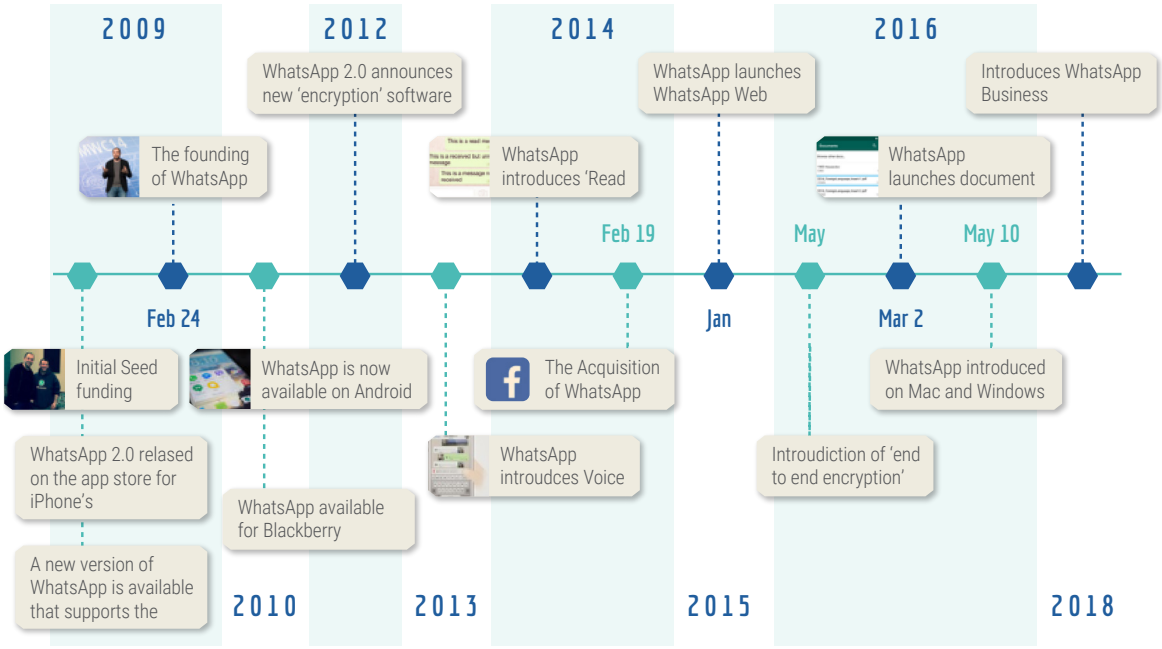


Table 1. Development of WhatsApp[8]

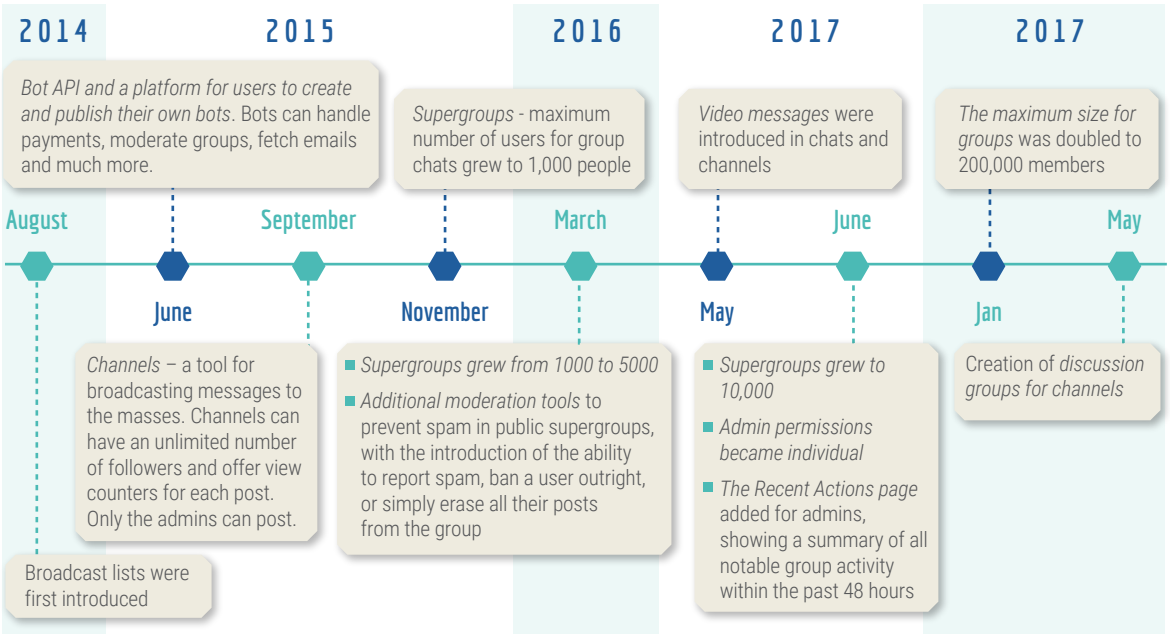


Table 2. Development of Telegram[9]

SECURITY RISKS OF SOCIAL MESSAGING

Impersonations and phishing

Millions of accounts across the various social media platforms are known to impersonate real organisations and persons, such as celebrities and politicians.[10] The motives of impersonations are disparate – from financial scams and political campaigning to simple mischief.

One recent WhatsApp impersonation[11], reported in December 2019, involved a request sent to users in Singapore, seemingly by their friends, to forward a 6-digit verification code or one-time password they received. This would subsequently lead to loss of access to the account. Scammers would then use these appropriated accounts to impersonate the users and request sensitive and confidential information from the users' friends. Unsuspecting individuals, who followed through with the instructions, soon discovered unauthorised transactions from their bank accounts.

From a national security perspective, impersonation can undermine command and control structures and reduce the morale of soldiers. For example, in 2015, Ukrainian soldiers in Eastern Ukraine received geo-targeted text messages that spread disinformation designed to degrade their will to fight. Messages were sent out to a number of Ukrainian soldiers defending the railroad town of Debaltseve[12], maintaining that the unit's commander had deserted or that Ukrainian forces were being decimated. The messages appeared to be coming from fellow soldiers with the impersonation described by one of the soldiers as "threatening and demoralising". It will not be farfetched for such techniques to be employed, at scale, on the increasingly popular social messaging platforms.

Social messaging-enabled cyber attacks

As much as 90 percent of data breaches occur due to human errors.[13] Human actions such as enabling a macro, downloading a file, or simply clicking on a link, open doors for successful cyber-attacks. This leads to malicious actors preying on oblivious or sometimes plainly complacent individuals to infiltrate critical cyber infrastructure as they present easier targets.

One of the most conspicuous examples of such an infiltration is the 2018 phone hack of Jeff Bezos, the founder of Amazon and the owner of The Washington Post.

According to an investigation[14], Bezos had allegedly received a WhatsApp message from the crown prince of Saudi Arabia with a malicious file that subsequently infiltrated the phone, giving access to large amounts of data that included some private – and sensitive – information.

Similarly, an unsuspecting military commander opening a link sent through a social messaging platform may compromise potentially sensitive security-related data.

Conduits for information influence activities

Information influence activities are harmful forms of communication orchestrated by actors or their representatives.[15] These activities shape perceptions and seek to undermine national institutions and values, such as democracy and freedom of expression. Like social media, social messaging platforms can be conduits for information influence activities.

In light of the recent coronavirus – officially named COVID-19 – outbreak, technology companies and platforms have pledged to promote truth and combat coronavirus-related misinformation. Facebook, which owns WhatsApp, pledged to support the work of the global public health community, especially by limiting the spread of misinformation and harmful content.[16] Concrete steps were taken such as working with Singapore's government to use the WhatsApp Business API to respond with health information to people that agreed to receive updates about the developing coronavirus situation.[17] Reddit, a community chat online platform, also recently "quarantined" one of its communities for posting false information related to the outbreak.[18]

Nonetheless, there were still notable instances of misinformation being spread through the digital domain, including through social messaging platforms. For example, a message claiming that the cure for the virus is a mix of garlic and boiling water was widely circulated on WhatsApp.[19] Such misinformation can undermine public safety and related national institutions, in this specific case – the authority and guidelines of the Department or Ministry of Health.

Social messaging platforms are also known to be used by violent extremist groups to spread propaganda and recruit vulnerable individuals. The inherent anonymity and encryption of these platforms has catalysed the shift of Jihadi groups from social media platforms to social messaging platforms, particularly Telegram channels.[20]

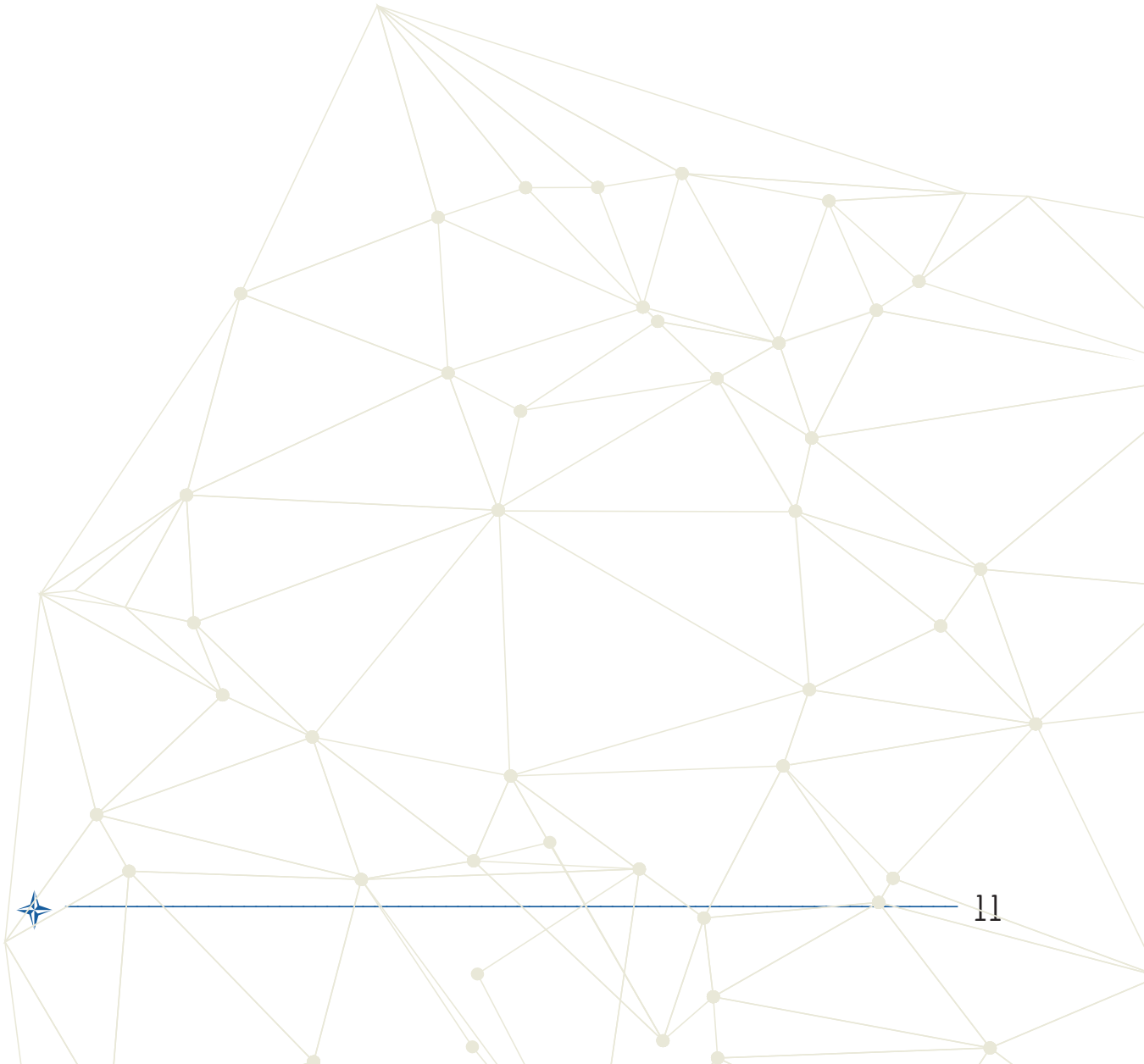
For a while, Telegram had indeed been the preferred platform for various extremist groups.[21] For instance, it was used by DAESH to spread propaganda regarding the 2015 Paris attacks and recruit perpetrators for the Christmas market attack in Berlin in 2016, among others. Additionally, DAESH claimed responsibility for the terror attacks

through Telegram – as in the case of the so-called lone-wolf terrorist attack in London in March 2017, as well as the Manchester Arena attack in May 2017.[22] In 2019, a significant number of accounts affiliated with DAESH and al-Qaeda were deleted from the platform – a sign of Telegram's efforts to eliminate terrorist activities by employing an algorithm introduced for these purposes.[23] While this is the case, there is still evidence of recruitment and propaganda being spread on private channels on Telegram.[24]

Undermining data privacy

Social messaging platforms request access to a host of features on one's phone, such as microphone, camera, and the Global Positioning System (GPS). This wide access to the various features on one's phone, correspondingly increases threats to data and digital privacy.

For example, the Israeli spyware tool Pegasus was allegedly used by some governments between April and May 2019, undermining the security of around 1,400 phones which belonged to lawyers, journalists and human rights activists.[25] The devices were hacked through video- and voice-call requests on WhatsApp – even when ignored or declined – and Pegasus spyware gained access not only to the users' WhatsApp activity, but also to their location and general phone use. The risk of such breaches has led international organisations such as the United Nations to ban its employees from communicating on WhatsApp.[26]



> We found service providers advertising on Google and selling manipulation software on online marketplaces and forums.

THE GLOBAL AND OPEN MARKETPLACE

As the number of social messaging platform users continues to grow, with an estimated 25% increase by 2022[27], it is logical to anticipate that the market of manipulation tools and services will evolve in tandem. During this study, while we expected service providers to be tucked away in remote corners of the open web, we found manipulation products and services for both WhatsApp and Telegram to be widely available and easily accessible. For instance, we found service providers advertising on Google and selling manipulation software on online marketplaces and forums.

Unlike social media manipulation services, which are observed to be predominantly Russian[28], the manipulation service providers for social messaging platforms are not clustered within a single geographical location. Rather, they are active throughout the world with providers as likely to be based in Indonesia or India as in the US. However, the availability of manipulation services in a certain location is understandably related to the app's popularity in that location. For example, a notable number of manipulation service providers for WhatsApp are based in India due to the overwhelming popularity of WhatsApp in the Indian subcontinent. In addition to differences in geographic areas of popularity, WhatsApp and Telegram have distinct characteristics that influence their respective places within the social messaging platform manipulation ecosystem.

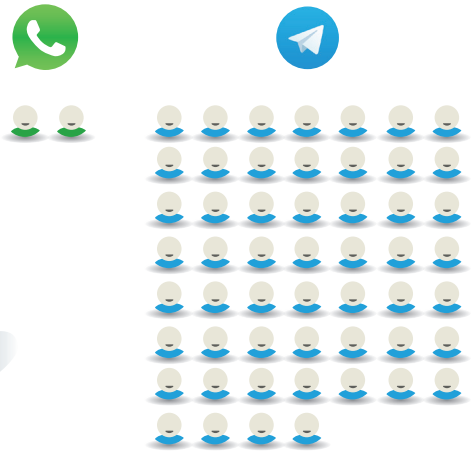
The differences in the inherent features of the platforms reflect the variety of manipulation services offered. Unlike WhatsApp's limit on the number of members in a group set at 256, Telegram allows for groups of up to 200,000 users and channels that enable broadcasting to an unlimited number of recipients. These supergroups and channels on Telegram are also often segregated by interest groups such as investing, shopping, or travel, which individuals can join through an invitation link. Overall, the marketplace clearly strives to replace the traditional telemarketing industry with three services dominating the selection: mass messaging, sale of contacts databases, and tailored engagement services.

Unlike **WhatsApp's** limit on the number of members in a group set at **256**,

Telegram allows for groups of up to

200,000 users and channels

that enable broadcasting to an unlimited number of recipients.



These supergroups and channels on Telegram are also often segregated by interest groups such as investing, shopping, or travel, which individuals can join through an invitation link.

Overall, the marketplace clearly strives to replace the traditional telemarketing industry with three services dominating the selection: **mass messaging, sale of contacts databases, and tailored engagement services.**

Types of manipulation services provided

Mass messaging

Mass messaging services dominate the market. Such services offer the ability to send messages automatically to a number of recipients, either directly to individuals or to groups and channels. The limit on the number of users in a group on WhatsApp is an inherent obstacle for mass messaging, as finding groups still open for joining (i.e. that have less than 256 members) is not an easy process. Nevertheless, despite tighter restrictions on WhatsApp, and possibly due to its higher popularity, mass messaging services were equally easy to find for both Telegram and WhatsApp.

Access to contact databases

Selling access to contact databases is another service that is widely offered. While often tied together with mass messaging services, these databases are also sold separately and often tailored to a specific audience. The most common databases on offer are commercially driven, including collections for gambling or online shopping. Nonetheless, how the contacts are collected or how accurate the segmentation of the databases is remains unknown.

We also found that it is relatively simple to compile a database of contacts manually from open WhatsApp groups and public Telegram groups and channels. For example, on WhatsApp, the only prerequisite to accessing the phone numbers of members of a group is being a member of that group yourself. On Telegram, the extraction of contacts is slightly more complex. Telegram does not reveal the phone numbers of a group's members. To get access to the usernames of members some basic web scraping is required, and even then the process is laborious, with only a select number of members per group identified on each attempt. WhatsApp, therefore, seems to be weaker at protecting the identities of users than Telegram.

Tailored engagement services

Another category of services available on the market are tailored engagements. While more commonly available for Telegram due to its nature, these services typically offer a six to eight-month campaign to increase the reach of content. These services offer the ability to purchase channel followers (with the opportunity to select a target audience – for example, based on country and gender), buy votes, likes or dislikes on a channel's posts, or get access to a service that artificially increases the number of views on a post to boost its popularity, among other features.

Actors providing services

The manipulation service buyer has multiple options to choose from. The most widely accessible service, mass or bulk messaging, can be acquired not only through sellers of mass messaging software, but also through API providers, social media marketing (SMM) agencies, freelancers, as well as online platforms. Prices for the services vary significantly even within the same category

Software for bulk messaging is the most common and cheapest manipulation service available on the market. Once purchased, the software typically allows an individual to send an unlimited volume of messages. However, the price, sophistication, and number of features included vary significantly between different software.

of product. While manipulation services for social messaging platforms are more costly than social media manipulation services (e.g. buying bots, likes or comments on Facebook/Instagram/YouTube), they are still relatively affordable. For example, we were able to buy a software that allowed for an unlimited number of messages on WhatsApp and Telegram for 26 EUR.

Sellers of software

Software for bulk messaging is the most common and cheapest manipulation service available on the market. Once purchased, the software typically allows an individual to send an unlimited volume of messages. However, the price, sophistication, and number of features included vary significantly between different software. While some programs have a simple interface with limited functions, others have additional features like speed and delay control, or the support of multiple accounts. In one case, bitcoin was the only currency accepted by the provider.

API providers

These providers offer access to their APIs and related documentation, such as codes that automate bulk messaging, both on WhatsApp and Telegram. These APIs allow the sender to integrate mass mailing services into a site or an application, or even customise or enhance the bulk messaging service. While also fairly cheap, API services are not easy to use. For example, use with WhatsApp requires an approval by WhatsApp itself, alongside the mandatory status of a business account. Telegram APIs, on the other hand, are easier to access and allow for the platform to be used for a variety of functions such as charts and polls.

Social Media Marketing (SMM) agencies

In addition to bulk messaging services, SMM agencies offer 'value-added' services such as target audience selection. Examples of the most widely available criteria for target audience selection include gender, age, and ethnicity. Some providers within this

> We were able to find freelancers offering to send messages to over 50,000 users, without being blocked and flagged as spam for as low as 30 EUR per day of hire.

category of actors specialise in a specific audience, such as those interested in cryptocurrency or online shopping. There is occasionally a geographical dimension to target messaging, with an option to choose a specific country or even a region within a country. The cost of the services provided by SMM agencies, naturally, is higher as they offer both better reliability and the selection of a target audience.

Freelancers

There are forum discussions dedicated to manipulation services for social messaging platforms. These discussions often include people offering or looking for manipulation services, as well as reselling purchased software for cheaper. Potential customers post an advertisement there, and in response, freelancers offer their services. We were able to find freelancers offering to send messages to over 50,000 users, without being blocked and flagged as spam for as low as 30 EUR per day of hire.

Online platforms

Online platforms are online interfaces that offer mass messaging through back-end processes and are, in principal, similar to software. Users can upload their contacts and messages, coupled with a set of instructions, onto websites that send out the messages accordingly.

When taking into account the sheer popularity of social messaging platforms like WhatsApp and Telegram, it is not surprising that an extensive and diverse marketplace of manipulation services for the platforms exists. It is also clear that this ecosystem, dominated by offers of bulk messaging services, is widespread and immediately accessible around the world through various types of providers and at a relatively affordable price.

Features used to circumvent protective mechanisms

When testing mass messaging services for WhatsApp and Telegram, we found a specific – and often rather limited – set of features that varies only slightly from provider to provider. Most of these features attempt to circumvent the protective mechanisms of the apps, while some aim to enhance the convenience of the mass messaging process. Overall, we found the list of features for mass messaging on Telegram to be significantly shorter and

Figure 1. Example of a dashboard of time delay

less sophisticated than the selection of features for sending messages in bulk on WhatsApp.

Message delivery delay

This feature broadly entails setting a random or fixed time delay between messages sent. With a fixed delay option, users indicate how long a delay should be and when it should occur vis-à-vis the number of messages to be sent. Random delays assign a random period of delay between each message. An example of this is illustrated in *Figure 1*. The recommended delay period and type of delay, however, differ widely between sources. For example, some providers recommend a fixed time delay (i.e. message interval set at 7 to 10 seconds), while others recommend a random time delay between 30 and 60 seconds – with strict advice against the fixed delay. Additionally, the time delay feature is more prevalent on WhatsApp than Telegram. The absence of a community standard for time delay is possibly indicative of the robustness of the messaging platforms' ability to detect when time delay is used as a method to circumvent the terms of use of the messaging platforms.

Mimicking human behaviour with dummy accounts

In its essence, the most fundamental way to avoid getting blocked is to mimic human behaviour. Accordingly, the dummy feature is designed to imitate the activities of an ordinary user in the intervals between mass mailings. To emulate such an activity, in addition to the main account from which the mass messaging occurs, several connected accounts are set up that share mutual 'friends'. These accounts are then used for the imitation of authentic behaviour – the same 'persons' seem to appear in the chat

history of the main account, seemingly reducing a messenger's suspicion about the account's other activities.

Customisation of messages

The idea behind customisation is to make the content in sent messages more individual and therefore less suspicious. Service providers use this feature to automatically randomise and vary, at least slightly, the content and headers of texts in a message. In practice, the feature often just adds a random character or emoji at the end of a text, potentially attracting even more suspicion from the apps' protective algorithms.

Spreading inauthentic behaviour among various accounts

The multiple accounts feature uses several accounts during a single session of mass messaging. The sending of messages alternates between several accounts to reduce the likelihood of being detected by the social messaging platforms' protective mechanisms. As the inauthentic behaviour is spread among various accounts, every individual account, in theory, attracts less suspicion. We found that this feature is not widely available, with only one manipulation service provider observed to offer it.

Use of anti-blocking mechanisms

Some providers claim to have developed anti-blocking algorithms (often marketed by individual providers as an 'enhanced anti-blocking algorithm'). While not much is known about this feature, it is broadly described as an amalgamation of common features that purportedly reduce the chance of getting blocked.

Toggling the internet speed

This feature involves toggling the internet speed of the platform used to mass message. For example, we observed one provider that allows users to set the connection speed to very slow, slow, normal, fast, or very fast. However, the relevance of this feature is not described in any detail and the connection between internet speed and successful circumvention of Telegram or WhatsApp's protective mechanisms is questionable. Two possible explanations for the inclusion of this feature might be that toggling the speed may allow one to either imitate a patchy mobile internet network or diversify the speed of sending messages.

Usage of number/user filter in databases

This feature allows a user to check whether a given phone number is registered on WhatsApp or Telegram and extract viable contacts from public databases that store different kinds of numbers. While the feature does not affect if one gets blocked or not, it reduces the potentially extensive list of recipients (e.g. extracted from public forums/chat groups/channels) to only relevant numbers. This feature is aimed at improving efficiency by reducing the time needed for mass messaging, rather than at circumventing the protective mechanisms implemented by messaging applications. An example of this feature is illustrated in *Figure 2*.

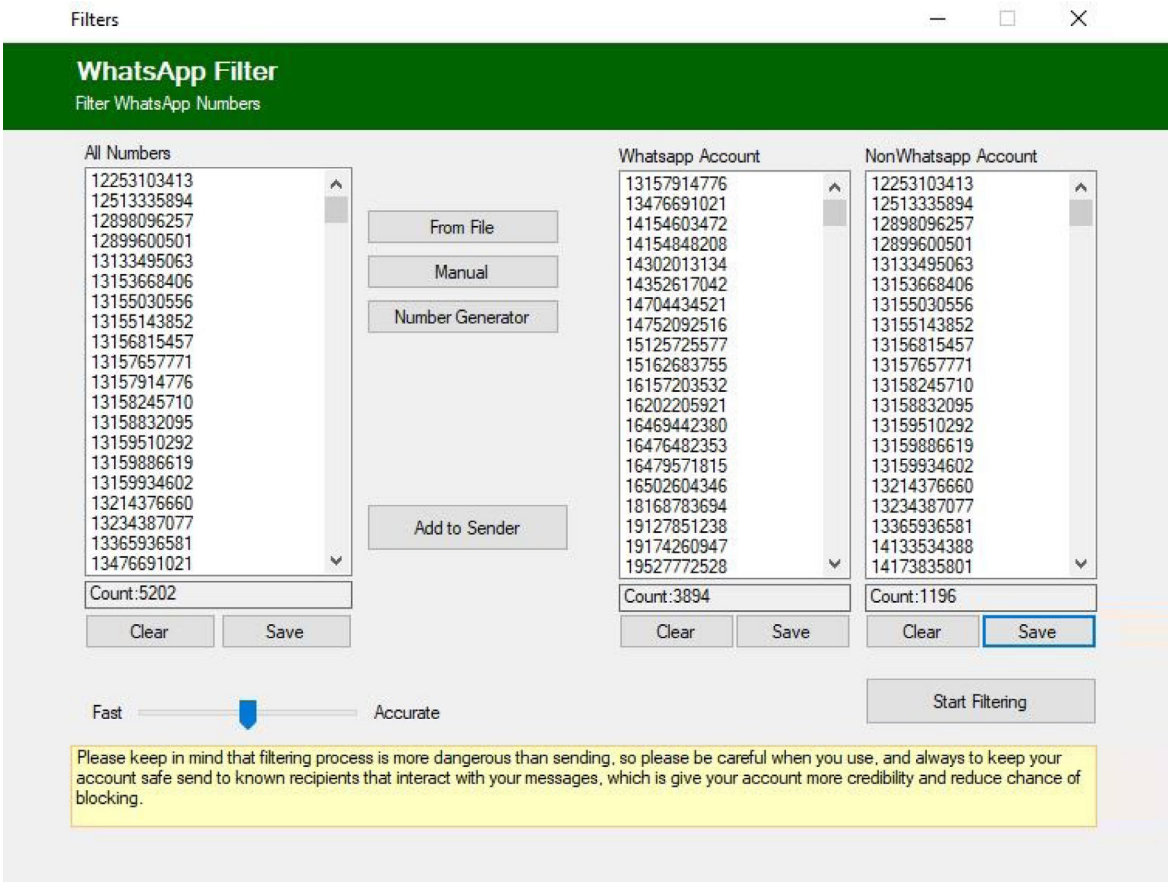


Figure 2: Example of the number filter feature

EFFECTIVENESS OF MASS MESSAGING SERVICES

We set out to test the effectiveness of manipulation service providers in circumventing the protective mechanisms of WhatsApp and Telegram. To do so, we researched the market of mass messaging services for the apps and tested a selection of these services. After using the acquired manipulation service, we assessed how it operated against protective measures and how long it took for the instant messaging apps to block us. Multiple challenges associated with mass messaging became evident during the testing process, many of which appear difficult to overcome. These challenges shaped our original plans for each stage of the testing.

Blocking virtual numbers at registration

To start using the aforementioned manipulation services, we needed phone numbers linked to WhatsApp and Telegram accounts. The numbers used for setting up new accounts were of two types – virtual phone numbers and newly purchased SIM cards.

At first, it seemed that purchasing virtual numbers, which are cheap and available for a variety of countries, was an innovative way of creating accounts. However, accounts set up using virtual numbers were blocked almost immediately, likely because they have been reused (and resold) multiple times. Attempts to create a new WhatsApp/Telegram account using a different virtual number on the same device also led to us being blocked at the point of registration, presumably because data points from the devices had already been registered by the messaging platforms during previous failed attempts. Examples of virtual numbers that can be linked to WhatsApp accounts for sale are illustrated in *Figure 3*.

The attempts at mass messaging using physical SIM cards were slightly more successful but, in general, were blocked within the first 20-50 messages. To avoid getting blocked quickly, some additional precautions were taken. Instead of creating accounts with new SIM cards, which are more prone to attracting suspicion and therefore get blocked faster, we used SIM cards which had been active for a while, without engaging in any kind of inauthentic behaviour. Additionally, before engaging in bulk messaging, the activity of an ordinary user was imitated, i.e. chatting with other users. The first messages were also only sent to recipients within the same general geographical region. Finally, each time an account was blocked, a new number *and* a new device were used. These measures

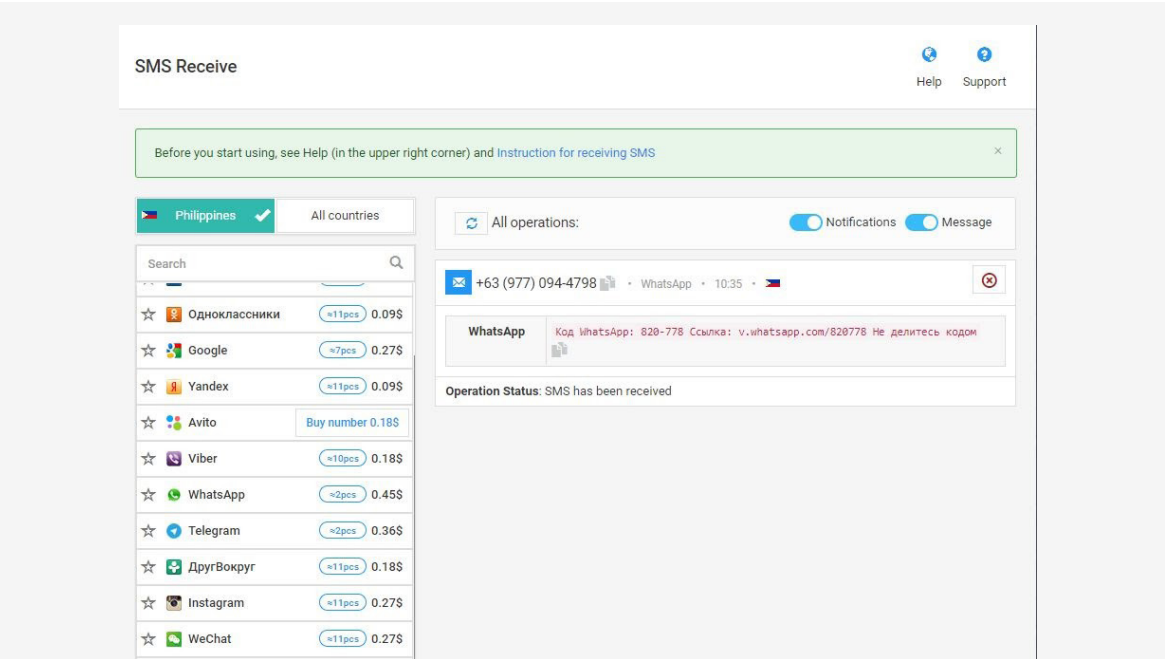


Figure 3: Virtual numbers (that can be linked to WhatsApp accounts) for sale

did make some difference during the early stages of testing. However, taking into consideration that after each time a block occurred, a new SIM and device were required, the resources needed for messaging on such a scale were unsustainable.

Collecting multiple data points from a user

After numerous attempts to avoid being blocked on the same device using different numbers, it became abundantly clear that both WhatsApp and Telegram collect multiple data points, or metadata, from a user that allow them to fingerprint a user by their unique identifiers. WhatsApp, for instance, has permission to view network connections, retrieve running apps, read Google service configuration, phone status, and identity, as well as user accounts on the device.[29] This means a simple change of SIM and reinstallation of the application on a device likely will not be enough to bypass WhatsApp and Telegram's security measures.

The use of proxies can, in theory, be a solution. Proxy use is explicitly recommended by some manipulation service providers as a way to avoid being blocked. A proxy or a proxy server is an intermediary server that acts as a getaway between a local network and a larger-scale network, usually the internet[30] (i.e. between end users and the applications they run[31]). Theoretically, this means that with the use of proxies, the excessive changing of devices would not be necessary – because

> The problems we experienced while attempting to send messages to other regions were so pronounced that eventually we had to limit the geographical scope of the experiment. Overall, it is clear that both WhatsApp and Telegram are very effective at detecting inauthentic activities carried out across geographical regions. In a security context, this translates to domestic and regional actors being more likely to succeed in carrying out mass messaging campaigns than foreign actors.

to WhatsApp and Telegram, it will seem like a certain request (or, in this case, bulk messaging) is coming from a proxy server rather than from your device.

Yet the problem we encountered with proxies was the general difficulty of integrating them into the majority of services offered by the manipulation providers. Only one provider allowed us to install a proxy within the interface of software; in this case, we started to experience difficulties with the mass messaging service itself after the proxy was installed. Furthermore, multiple proxies were required to send messages at scale, making it impractical for an average user due to the additional money and time required.

Identification of the geographical location of message recipients

The third observation during our testing was WhatsApp and Telegram's ability to quickly recognise the geographical location of messaging. Both apps assess the probability of a phone with a certain country code connecting through a cell network in a geographical area different from the area presumed by this country code. The geographical location of the message recipients also makes a difference, as the apps' algorithms analyse the likelihood of the cross-country mailing being authentic. During our testing, we had higher success rates messaging users that resided within the same geographical region than between regions.

The problems we experienced while attempting to send messages to other regions were so pronounced that eventually

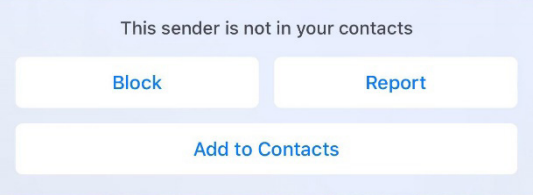
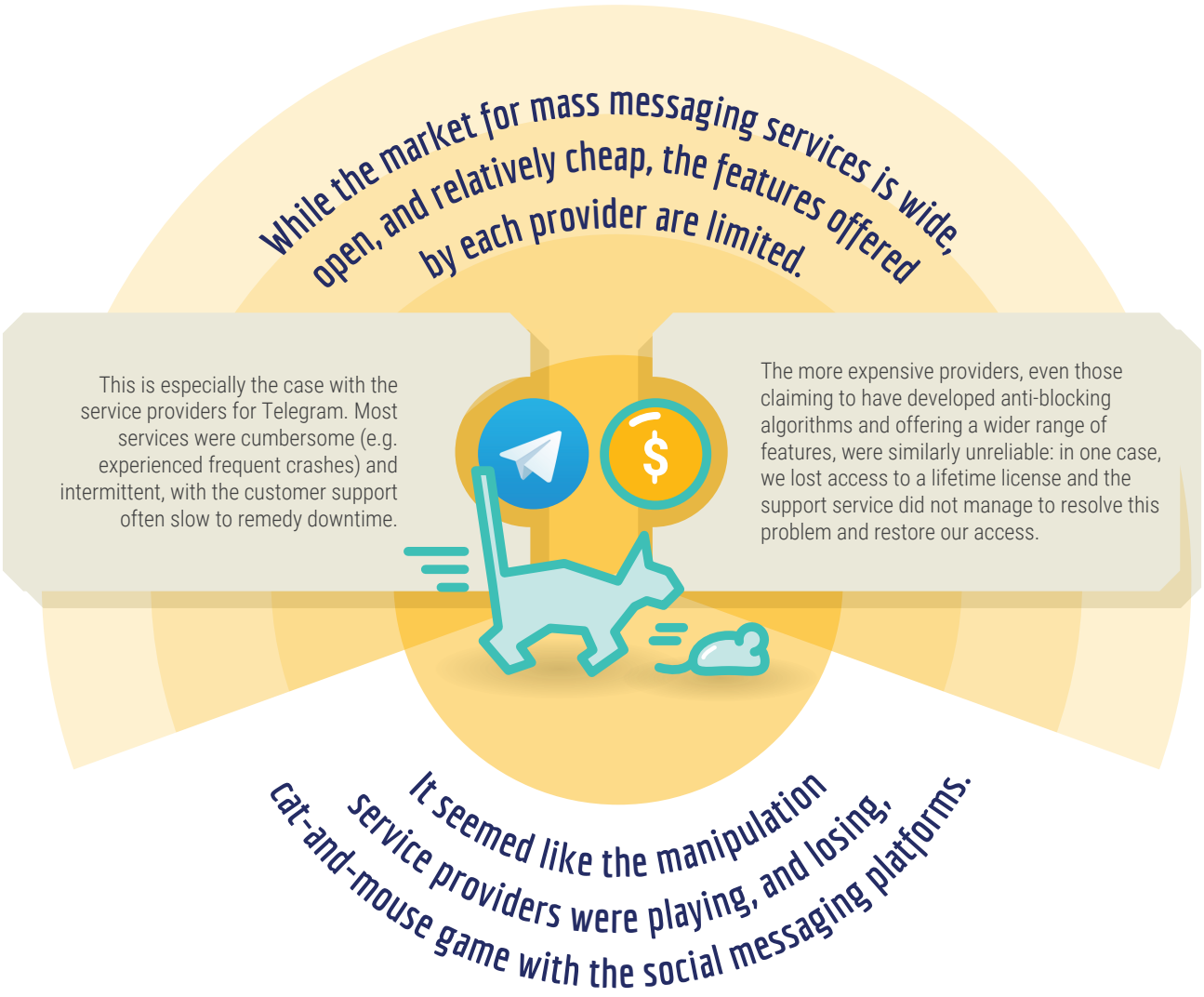


Figure 4: WhatsApp's proactive prompt to report messages



we had to limit the geographical scope of the experiment. Overall, it is clear that both WhatsApp and Telegram are very effective at detecting inauthentic activities carried out across geographical regions. In a security context, this translates to domestic and regional actors being more likely to succeed in carrying out mass messaging campaigns than foreign actors.

Monitoring content of messages

WhatsApp famously prides itself on its end-to-end encryption. In the case of Telegram, end-to-end encryption is not enabled by default and only applies when using secret chats. Some mass messaging service providers, therefore, insist that the suspicious content (i.e. use of links and

references to channels or names) should be minimised, if not excluded altogether. It is also likely that the content of the message matters simply because it affects how likely one is to get reported by the recipients of the message. A message littered with emojis, links, and bad language from an unknown user, is far more likely to be treated as spam. In addition, reporting suspicious content is made easy, as both social messaging platforms proactively give recipients the option to report content when messages are received from an unknown source. (illustrated in *Figure 4* above)

Quality of service providers

Our final observation was the general poor quality of the manipulation service providers themselves. While the market for mass messaging services is wide, open, and relatively cheap, the features offered by each provider are limited. This is especially the case with the service providers for Telegram. Most services were cumbersome (e.g. experienced frequent crashes) and intermittent, with the customer support often slow to remedy downtime. The more expensive providers, even those claiming to have developed anti-blocking algorithms and offering a wider range of features, were similarly unreliable: in one case, we lost access to a lifetime license and the support service did not manage to resolve this problem and restore our access. It seemed like the manipulation service providers were playing, and losing, a cat-and-mouse game with the social messaging platforms.

Services purchased that worked were also generally slow to send out messages. The fastest bulk messaging provider took about 7 seconds to send one message across both WhatsApp and Telegram. In the digital world, where social media bots deliver comments in milliseconds, this seemed like aeons.

Overall, during each stage of the testing period it became clear that mass messaging services are incapable of delivering what they promise to, and it would take a relatively sophisticated and sufficiently motivated actor to send messages on a large scale across both WhatsApp and Telegram.

CONCLUSION

This study sheds light into the open, global, and cheap marketplace for manipulation available on social messaging platforms. While the marketplace is not as mature as the one for social media manipulation, it does offer tools and services that a malign actor can easily purchase to reach a selected audience. Through the study, we discovered that:

- **The manipulation of social messaging platforms, especially mass messaging, is an information influence and cyber-security risk**. One third of private companies admitted to having suffered a breach that involved a mobile device in 2019, a 5% increase from the previous year.[32] Cyber-attacks often start with phishing attempts, including getting an unsuspecting user to click on a malicious link. Mass messaging directly to an individual through social messaging platforms, a more trusted medium, increases society's vulnerability to cyber-attacks in general. Similarly, messaging platforms such as SMS have previously been exploited for information operations by adversaries in times of crises. Social messaging platforms may be equally vulnerable to this exploitation.

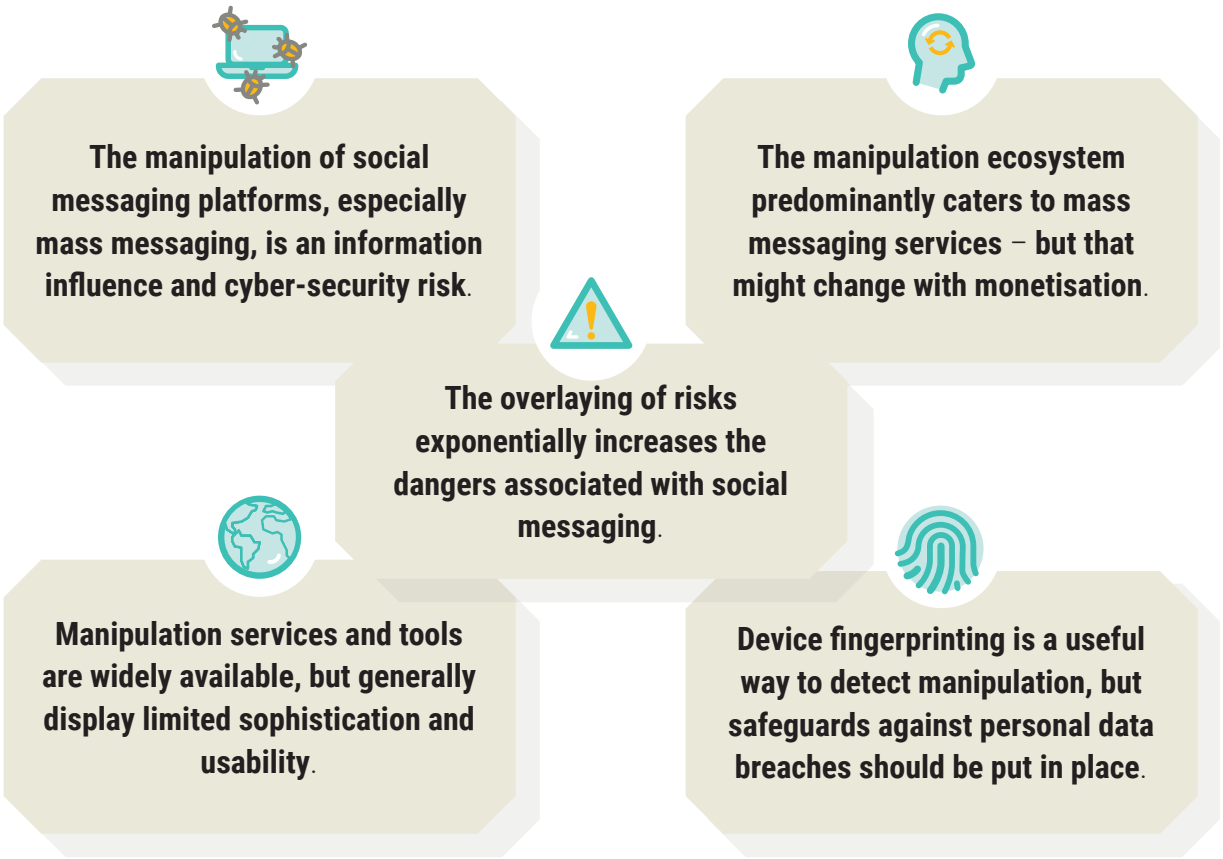
- **The manipulation ecosystem predominantly caters to mass messaging services – but that might change with monetisation**. The orientation to mass messaging can be attributed to the fact that WhatsApp and Telegram are still relatively advertisement-free, making the incentive to move beyond telemarketing limited. With plans for both platforms to monetise various parts of their applications, the scale and avenues for manipulation are expected to expand correspondingly.

- **Manipulation services and tools are widely available, but generally display limited sophistication and usability**. Although the existing ecosystem is extensive, the services and tools adopt rudimentary methods of circumventing the protective mechanisms of the platforms, and it is clear that the social messaging platforms are currently capable of detecting and disrupting large-scale coordinated mass messaging attempts. This is a race where social messaging platforms are currently in the lead.

- **Device fingerprinting is a useful way to detect manipulation, but safeguards against personal data breaches should be put in place**. We found that the social messaging platforms have robust mechanisms for detecting

manipulation that can in some part be attributed to the large amounts of metadata the platforms take from a user. In contrast, previous research has shown that social media companies are falling behind in this race. While elements of digital fingerprinting can be adapted to improve protection on social media platforms, concurrent efforts should be made to reduce the data privacy risk and mitigate the possible negative effects of a breach of collected metadata.

- **The overlaying of risks exponentially increases the dangers associated with social messaging**. This study has reinforced that, due to the sheer number of tools and services available, it is certainly possible for motivated malign actors to leverage social messaging platforms to manipulate the perceptions of vulnerable audiences. Given its inherent exclusivity, social messaging platforms are also arguably regarded as 'safer' spaces for communication. Unfortunately, this feeling of a 'safer' space makes one increasingly susceptible to information influence and impersonation attempts. 'Hacked' or misappropriated accounts, especially of influential individuals, can result in exponentially greater effects of information influence activities.



Endnotes

1 Montaque, T. (January 3, 2019) 'Here's How Messaging Is Positioned to Dominate in 2019' Adweek.

2 Hill, S. and S. Chandler (January 31, 2020) 'The best text messaging apps for Android and iOS'. *Digital Trends.*

3 Dove, J. (March 25, 2020) 'What is WhatsApp?'. *Digital Trends*

4 Linaker, E. 'What's Up WhatsApp? My Top Tips For Enhancing Your Business Profile in 2019' *Medium.com*

5 Monfex Research Department (September 10, 2019) 'The TON Coin is the Next Big Thing for Social Media' *Monfex.*

6 McLaughlin, T. (February 12 2020) 'How WhatsApp Fuels Fake News and Violence in India'. *Wired*

7 Katz, R. (September 1, 2019) 'A Growing Frontier for Terrorist Groups: Unsuspecting Chat Apps'. *Wired*

8 Whatsapp. 'Whatsapp Blog' . *Whatsapp*

9 Telegram.com (August 15, 2019) 'Celebrating 6 years of Telegram' *Telegram.com*

10 Confessore, N. and G.J.X. Dance 'On Social Media, Lax Enforcement Lets Imposter Accounts Thrive'. *The New York Times.*

11 CNA (February 19, 2020) 'Police Warn of new scam involving takeover of WhatsApp accounts'. *CNA.*

12 Satter, R. and D. Vlasov (May 12, 2017) 'Ukraine Soldiers Bombarded by 'Pinpoint Propaganda' Texts'. *AP News.*

13 Spadafora, A. (May 8, 2019) '90 percent of data breaches are caused by human error'. *Techradar.pro*

14 Kirchgaessner, S. (January 22, 2020) 'Jeff Bezos Hack: Amazon boss's phone 'hacked by Saudi crown prince'. *The Guardian.*

15 Swedish Civil Contingencies Agency 'Countering Information Influence Activities: A Handbook for Communicators'. MSB.se

16 Jin, K (April 7, 2020) 'Keeping People Safe and Informed About the Coronavirus'. Facebook.com

17 Lim, S. (February 10, 2020) 'Why misinformation is a clear and present danger during the coronavirus outbreak'. thedrum.com

18 Thomas, Z. (February 1, 2020) 'Coronavirus: How Facebook, TikTok and other apps tackle fake claims' BBC.com

19 Latto, R. (February 17, 2020) 'A Coronavirus Hoax message is circulating on WhatsApp – here's what to do if you receive it'. Derry Journal.

20 Stalinsky, S. R. Sosnow, M. Khayat, M. Al-Hadj, R. Green et. Al (December 23, 2016) 'Germany-Based Encrypted Messaging App Telegram Emerges as Jihadis' Preferred Communications Platform – Part V of MEMRI Series: Encryption Technology Embraced By ISIS, Al-Qaeda, Other Jihadis – September 2015- September 2016: Section 2 MEMRI Research Documents Jihadi Use of Telegram. *MEMRI.*

21 Tan, R. (June 30, 2017) 'Terrorists' love for Telegram, explained'. *Vox.com*.

22 Counter extremism project (May 2017) 'Terrorists on Telegram'. *Counterextremism.com*

23 Katz, R. (December 5, 2019) 'Telegram has finally cracked down on Islamist terrorism. Will it do the same for the far-right?' *The Washington Post*.

24 Airbus Defence and Space (2020) Mapping Extremist Communities: A Social Network Analysis Approach. *NATO StratCom COE.*

25 Adegoke, Y. (October 30, 2019) 'A WhatsApp hack used Israeli spyware to target Rwandan dissidents'. *Quartz Africa.*

26 Reuters (January 23, 2020) 'U.N. says officials barred from using WhatsApp since June 2019 over security'. *Reuters.com*

27 Clement, J. (October 8, 2019) 'Global number of mobile messaging users 2018-2022'. *Statista.com*

28 NATO StratCom COE and Singularex (2018). 'The Black Market for Social Media Manipulation' *NATO StratCom COE*.

29 Google Play Store: 'Whatsapp Messenger'

30 Petters, J. (March 29, 2020) 'What is a Proxy Server and How Does it Work?'. *Varonis.*

31 Indiana University (November 15, 2018) 'About Proxy Servers'. *The Knowledge Base Indiana University*.

32 Verizon (2019) 'It's time to tackle mobile security'. *Enterprise.verizon.com*

Prepared and published by the
NATO STRATEGIC COMMUNICATIONS
CENTRE OF EXCELLENCE

The NATO Strategic Communications Centre of Excellence (NATO StratCom COE) is a NATO accredited multi-national organisation that conducts research, publishes studies, and provides strategic communications training for government and military personnel. Our mission is to make a positive contribution to Alliance's understanding of strategic communications and to facilitate accurate, appropriate, and timely communication among its members as objectives and roles emerge and evolve in the rapidly changing information environment.

Operating since 2014, we have carried out significant research enhancing NATO nations' situational awareness of the information environment and have contributed to exercises and trainings with subject matter expertise.

www.stratcomcoe.org | @stratcomcoe | info@stratcomcoe.org