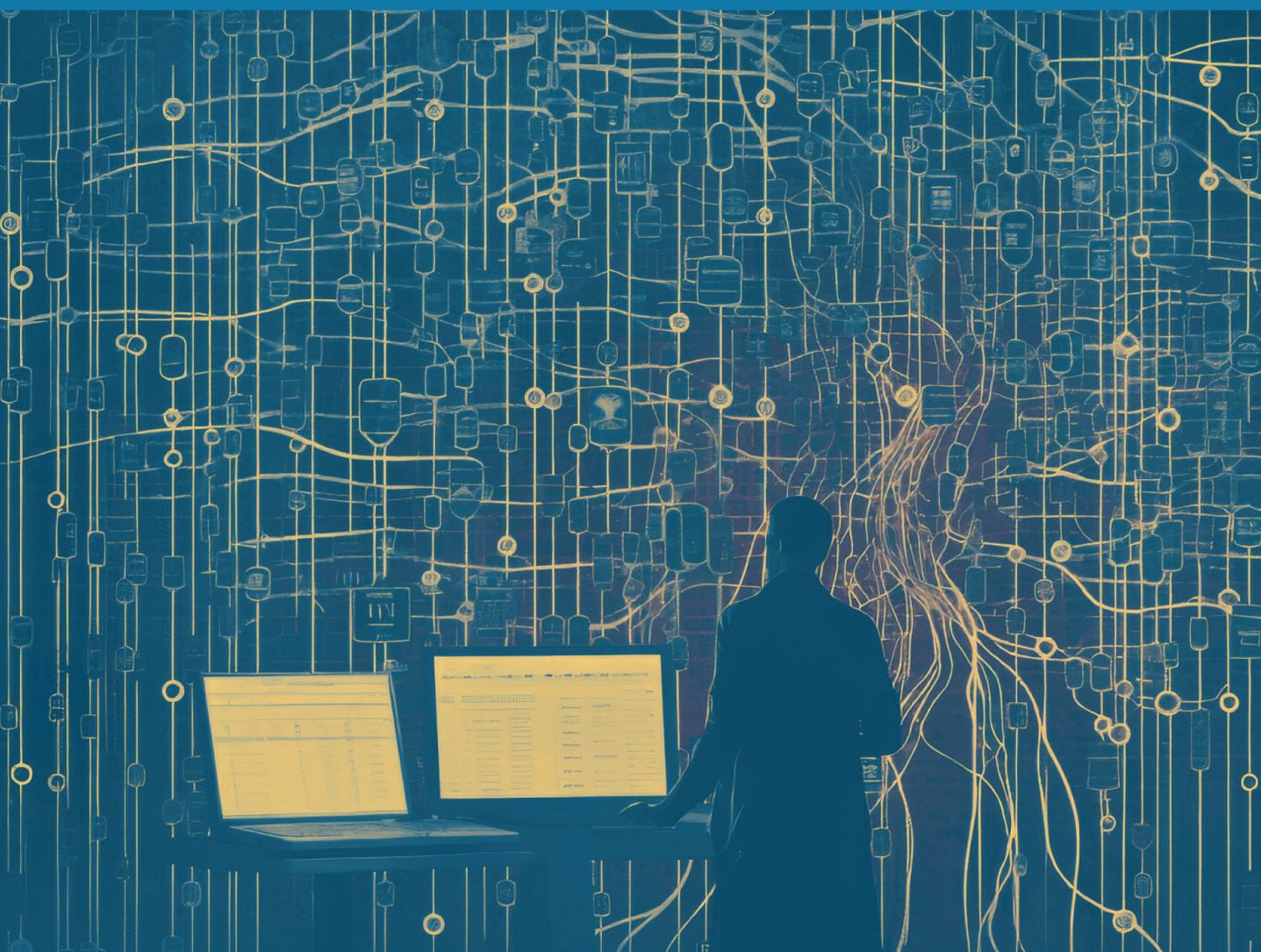
safejournalists.net

Association of Journalists of Macedonia

POLICY PAPER

COORDINATED DIGITAL ATTACKS ON JOURNALISTS AND MEDIA OUTLETS - A PRACTICAL PROTOCOL FOR RESPONSE, EVIDENCE PRESERVATION AND ACTION

SKOPJE, FEBRUARY 2026



Imprint
Policy Paper: Coordinated Digital Attacks on Journalists and Media Outlets - A Practical Protocol for Response, Evidence Preservation and Action

Author
Martin Petrovski

Editing
Dragan Sekulovski

Publisher
Association of Journalists of Macedonia (AJM)

Project
“SafeJournalists Network”

Proofreading in Macedonian
Maja Katarova

Translation from Macedonian to English
Olivera Celeska

Translation from Macedonian to Albanian
Ferikan Iljazi

Design
AJM

Skopje, February 2026.

NOTE

This publication is designed to provide practical guidance to newsrooms and journalists on managing coordinated digital attacks, with a focus on safety, evidence preservation, and effective escalation. It does not constitute legal advice. In high-risk cases, especially when there are threats, doxxing, or suspicion of unauthorized access, it is recommended to consult with legal counsel and report to the competent institutions.

CONTENTS

LIST OF ACRONYMS

- *2FA* - Two-Factor Authentication
- *CPJ* - Committee to Protect Journalists
- *EU* - European Union
- *EUR-Lex* - official database of EU law (portal for EU legislation and acts)
- *FB* - Facebook
- *FIMI* - Foreign Information Manipulation and Interference
- *IG* - Instagram
- *OSCE* - Organization for Security and Co-operation in Europe
- *SOP* - Standard Operating Procedure
- *X* - X (formerly Twitter)

INTRODUCTION

Coordinated digital attacks on journalists and media outlets often look like a “sudden surge” of harassment – mass false reports to take down profiles, organized insults and threats, impersonation, compromising profiles or publishing personal data for the purpose of intimidation. These attacks are rarely random. They are designed to create panic, disrupt operations and undermine trust in the editorial staff. In some cases, such attacks can also be linked to broader operations of Foreign Information Manipulation and Interference (FIMI), but the guidelines below are practical and applicable regardless of the origin of the attack.

This paper lays out practical steps that newsrooms can immediately implement - how to recognize a coordinated attack, what to do in the first 60 minutes and in the first 72 hours, how to collect and preserve evidence without violating privacy, what actions to take with the platforms, how to request support from AJM and when there are grounds to contact competent institutions.

The goal is for the response to be calm, consistent, and safe. The key principle is “do no harm” – to protect people, minimize the processing of personal data, and not take steps that unnecessarily increase the visibility of the attack or expose the team to additional risks.

Target audience of this paper

This paper is intended for editors, journalists, social media profile administrators, editorial management, and persons responsible for security and legal issues in the media.

Methodology

This paper was prepared based on practical cases and experiences from AJM’s work with journalists and editorial teams in dealing with digital incidents, as well as relevant guidelines for digital security, evidence preservation, and official rules on platforms for reporting and restoring access.

How to use this paper

- If an attack is ongoing, first use the “Rapid Response” page and immediately start an incident log[1].
- Then follow the standard operating procedure (SOP) in stages (first 60 minutes, first 24 hours, 48 - 72 hours).
- Collect evidence systematically and use the “evidence package” template for platforms and institutions.
- If there are threats or “doxxing”[2], apply the emergency security protocol and act without delay.
- When necessary, contact AJM for support, guidance and coordination of next steps.

Key take-aways

- ✓ A coordinated attack is treated as an incident - with roles, a timeline and clear decisions, not as a “chaos of comments”.
- ✓ Evidence is collected immediately, but carefully - with a minimum of personal data and with access control.
- ✓ Action is taken in stages - platform, AJM, institutions - according to the risk and damage.
- ✓ People’s safety takes priority - especially when there are threats, doxxing or compromised profiles.

1 Incident log - a brief chronological record of events and steps taken.

2 “Doxxing” - publicly sharing personal data without consent, with the aim of intimidating, harassing, or causing harm.

WHAT CONSTITUTES A COORDINATED DIGITAL ATTACK?

A coordinated digital attack is an intentional and organized activity directed at a journalist, editor, or media outlet, with the aim of causing disruption, self-censorship, discrediting, or physical intimidation. The attack may appear to be a “spontaneous” wave, but in practice it often has recurring patterns, synchronization, and a clear goal.

Coordinated attacks differ from “regular” online criticism in that they are aimed at controlling visibility and access to an audience. This means that the goal is not just to leave a comment, but to take down a profile, clog communication channels, impose a false narrative, or threaten personal safety.

Why this is important for newsrooms

When an attack is coordinated, the response takes on a structure. Improvisation often leads to three problems: loss of evidence, mishandling of platforms, and increasing risk through unnecessary public amplification of the attack. That is why this paper treats a coordinated attack as an incident with roles, an incident log and clear decisions.

PART A - HOW TO RECOGNIZE THE PROBLEM

How to recognize that it is a coordinated attack

There is no single “sure” sign, but the combination of multiple signals is a strong indicator that an attack is organized and that the protocol should be activated.

- There is a sudden increase in comments, messages, or reactions in a short period of time, with a repetitive tone and identical wording.
- Multiple profiles post the same or very similar accusations, links, or visual materials at the same time.
- Mass false reporting of profiles, posts, or pages is being activated in order to restrict or suspend them.
- New profiles with minimal history are emerging, which coordinately attack or “reply” to one another to create the impression of broad support.
- The attack moves from one platform to another, with the same message or the same targeting.
- There are attempts to impersonate, hack, or gain administrator access.
- Personal data, private photos, addresses, telephone numbers, or family information are published or threatened to be published.
- “Context erasure” is used through cropped quotes, unsourced screenshots, or fake headlines that are then “spun” as evidence.

Key terms

Mass false reporting is a coordinated practice of multiple accounts reporting content or a profile in order for the platform to automatically limit reach, remove content, or suspend a profile.

Impersonation is the creation or use of a profile that poses as a journalist, media outlet, or editorial staff member, for the purpose of manipulation, fraud, discrediting, or collection of personal data.

Compromised profile means that an attacker has gained unauthorized access to a profile or to management tools, such as administrator rights, the email associated with the profile, or management panels.

Doxxing is the public disclosure or threat of disclosure of personal data, for the purpose of intimidation, incitement to harassment, or creating a risk to physical safety.

Coordinated Harassment is organized targeting through insults, threats, sexualized comments, manipulated photos, or incessant messages, with the aim of exhausting and intimidating.

The most common scenarios of attacks on journalists and media outlets

In editorial practice, it is useful to group attacks according to what they attempt to "destroy": visibility, credibility, identity, access, or safety.

Attacks on visibility	occur when the goal is to limit reach, take down content, or suspend a profile through mass false reporting and abuse of platform mechanisms.
Attacks on credibility	occur when the goal is discrediting through false claims, manipulated statements, «screenshot evidence» without a source, or coordinated spin on a topic.
Attacks on identity	occur when there is impersonation, fake profiles «on behalf of the editorial team,» or attempts to deceive sources and audiences.
Access attacks	occur when there is phishing, profile hijacking, email compromising, or attempts to gain administrator rights.
Security attacks	occur when there are threats, doxxing, targeting of family members, or calls for a physical confrontation, and they should always be considered as highest priority.

PART B - WHAT TO DO AFTER RECOGNIZING AN ATTACK

If you recognize several of the listed signals, proceed to the practical steps below and then follow the “Rapid Response” and SOP in stages.

Basic rule for decision-making

If the attack affects access to profiles, involves threats or doxxing, or creates a risk to physical safety, the response is handled as a high-risk incident. If the attack is “just” a wave of comments, but is coordinated and impacts visibility or trust, the protocol is still activated, with a focus on evidence, controlled communication, and appropriate action on platforms and other relevant channels.

Roles and coordination in the editorial office

When an attack is coordinated, the newsroom works as a team with clear roles. The goal is to reduce chaos, not lose evidence, and make decisions that reduce risk.

It is best to assign roles in advance and for each role to have a replacement. When this is not possible, roles are assigned as soon as it is determined that an attack is underway.

- **Incident Coordinator -** leads the response, sets priorities, and approves public communication.
- **Platform Administrator -** verifies access, enforces settings, files reports and appeals procedures, and monitors status.
- **Evidence and Log Officer -** collects evidence, labels and archives it, maintains an incident log and prepares an evidence package.
- **Communications Officer -** prepares short, controlled messages for audiences and partners, without “blowing up” the attack.
- **Security and Legal Officer -** assesses the risk of threats and doxxing, proposes protective measures, and coordinates addressing institutions when there are grounds for that.

The editorial team applies one clear rule: decisions go through the coordinator, and everyone else adheres to the agreed-upon plan. This prevents parallel activities that create contradictory reports, loss of evidence, or additional exposure.

Incident log

An incident log is a simple record that allows you to track what happened, when, and how the editorial team responded. The log is useful for handling platforms, for institutions, and for internal analysis after the attack is over.

An incident log is kept from the very beginning and is updated in real time or every 30 - 60 minutes in the first hours. The log has limited access and is shared only with people who have a role in the incident.

The following fields are sufficient for practical use:

- The log includes the date and exact time of the event or activity.
- The log lists where the event is taking place, including platform, profile, and link.
- The log describes what was observed, for example mass false reporting, threats, impersonation, or compromised access.
- The log states what action was taken and by whom.
- The log has a reference to evidence, for example a folder number, file name, or link to archived material.
- The log has a status, for example “in progress”, “reported”, “under appeal” or “closed”.

Rapid response - the first 60 minutes

This section is intended for situations where an attack is in progress and the newsroom needs to stabilize the situation immediately. If there are imminent threats or doxxing, people’s safety has absolute priority.

1. Activate a procedure

Immediately designate an incident coordinator and at least two other roles, even if the same person temporarily covers multiple tasks. Open an internal coordination channel and agree on a rule that decisions go through the incident coordinator.

2. Do a quick risk assessment

Check for threats of violence, doxxing, posting addresses, phone numbers, or targeting of family. If there is, immediately switch to emergency safety protocol and act without delay.

3. Preserve evidence before any changes take place

Before deleting, blocking, or changing settings, first save the evidence. In the first 60 minutes, it is most important to capture a "snapshot" of the attack while it is in progress.

The minimum evidence package at this time includes:

- Each screenshot displays a full screen with date and time, username, and context of the post or profile.
- Each video screenshot shows how to open the link and how to reach the content in question, to preserve context.
- Each log includes a direct link to the content or profile, and when possible, a post or profile identifier.
- Each suspicious profile that is part of the attack is recorded with a link and a short note about its role in the attack.
- Every notification or email from the platform is saved as evidence, including subject, date, and content.

4. Stabilize profiles and access

The platform administrator checks for signs of compromised access and implements basic protection measures.

- Passwords are changed from a secure device and unique and strong passwords are used.
- Two-factor authentication (2FA) is activated or confirmed to be active for all key accounts and emails.
- Active sessions are reviewed and unknown logins are logged out.
- Administrator roles are checked, and unnecessary access is removed while the incident is ongoing.

5. Limit damage without exacerbating the attack

The communications officer should avoid impulsive public reactions. If a public message is necessary, it should be brief, factual, and aimed at protecting the audience and the team.

In the first 60 minutes, priority is given to measures that reduce exposure without creating additional conflict. Moderation is done carefully and only after evidence has been preserved, so as not to lose any evidentiary trail.

6. Make a report and follow the procedure

The platform administrator initiates a report based on the type of attack, for example impersonation, compromised profile, doxxing, or harassment. The evidence officer provides a structured package that is used in reports and is linked with the incident log.

7. Seek support from AJM when there are grounds

When the attack is coordinated, when there is a security risk, or when the platform response is ineffective, the editorial team may contact AJM for guidance and coordination of next steps. The contacts and the format of a short report are provided in the annexes.

"Do no harm" principle

- ✓ Personal data of attackers or third parties is not published, even when they first published personal data.
- ✓ Threats are not forwarded or published without first removing personal information that could increase the exposure of targeted individuals.
- ✓ Public messages do not reveal internal security measures and operational details (e.g. settings and vulnerabilities) that would help attackers adapt.

Standard Operating Procedure (SOP) - Phased Response

This protocol organizes the response into three chronological phases. The goal is to keep the newsroom functional, protect targeted individuals, preserve evidence, and act in an appropriate order through relevant channels.

Phase 1 - the first 24 hours

In the first 24 hours, the priority is to stabilize access, preserve evidence, and act quickly through appropriate channels.

1. Confirm the incident type and set a priority

The incident coordinator determines whether the attack is targeting visibility, identity, access, or security. If there is doxxing, threats, or compromised access, the incident is treated as high risk.

2. Extend evidence collection

After the initial evidence package, the evidence officer expands the material so that it can be used for both platform and institutional action. Each piece of evidence is given context, a timestamp, and a clear link to the incident log.

- The evidence is organized in a structured folder with clear subfolders by platform and type of attack.
- Each file is given a name that includes the date, time, platform, and a brief description of the contents.
- When archiving, the personal data of the targeted individuals and third parties is minimized, and access to the folder is restricted.

3. Activate procedures on the platforms - reporting, complaints and restoring access

The platform administrator initiates procedures according to the type of attack. Each submission is supported by an evidence package, and the incident log records the report number or status.

When there is a risk of losing a profile or compromised access, the priority is to regain access control, followed by public communication.

4. Manage moderation and comments without evidence loss

Moderation follows one rule: evidence first, intervention later. Once evidence is preserved, the editorial team can take limited measures to mitigate the damage, assessing whether the measure will exacerbate the attack.

- *Comments* and messages are moderated according to a clear logic that distinguishes criticism from harassment and threats.
- Any element containing a threat or personal data is documented and treated with a higher priority.
- Mass deletion is avoided when it destroys the evidence trail and makes action difficult.

5. Controlled communication

The communications officer ensures that the editorial team speaks with one voice, with minimal risk of further inciting the attack. In the first 24 hours, the public message is brief and focused on the facts, not on arguing with the attackers.

When there are threats or doxxing, public communication is cautious and does not reveal personal details, locations, schedules, or internal protection measures. In such cases, communication with institutions and support networks takes priority.

6. Support from AJM and action before institutions when there are grounds

When an attack is coordinated and causes significant damage, the editorial team can contact AJM with a brief summary, incident log, and key evidence. This allows for more effective coordination and guidance on next steps.

If there are threats, doxxing, compromising of profiles or misuse of personal data, the editorial team can move on with proceedings before competent institutions, with a clear evidentiary package and minimization of unnecessary personal data.

Phase 2 - 48 - 72 hours

In the next phase, i.e. hours 48-72, the priority is to consolidate evidence, monitor responses from platforms, reduce exposure, and stabilize the work of the editorial team.

1. Monitor responses from platforms and keep track of status by subject

The platform administrator keeps a list of all reports and complaints, with date, status, and next steps. If the platform responds with automatic sanctions against the account or content, the editorial staff files a complaint with clear arguments and an organized evidence package.

2. Consolidate an "evidence package" for institutions and partners

The evidence officer prepares a consolidated package that contains a timeline, key evidence, a brief explanation of the damage, and clear requirements. The package is made readable and usable without additional interpretations.

3. Reduce exposure and protect the team

The editorial team assesses the exposure of individuals, especially journalists who are directly targeted. If there are risks, measures such as limiting public visibility of personal profiles, enhanced privacy, and additional support for psychological and physical safety are considered.

4. Decide on whether to continue publishing/posting and on an editorial line

The incident coordinator, together with the editorial team, makes a decision on whether the editorial team will continue publishing the story that caused the attack and how it will do so without unnecessarily increasing the risk. The decision is based on a risk assessment and the support that the editorial team actually has.

5. Prepare a brief "lessons learned" analysis and update the measures

After the strongest wave subsides, the editorial team conducts a brief internal analysis with three questions: what worked, what didn't work, and what changes are necessary in terms of roles, tools, and procedures. This analysis is brief, but it leads to concrete improvements.

“Signal - Response - Next Steps” Matrix

This matrix serves as a practical guide at a glance and can be adapted to a specific case and platform (for example Meta - Instagram/Facebook/Threads, X, YouTube, TikTok).

Signal	Response (immediate)	Next steps (if it continues/if it is serious)
Mass false reporting + reach drop/limit	Preserve evidence (screenshots/video) + file a report/complaint on the platform	Contact AJM if the profile is seriously compromised or has recurring sanctions.
Impersonation (fake profile "as a media outlet/journalist")	Document the fake profile + establish/announce what the official channel is	Application for impersonation on the platform + AJM if there are multiple profiles / broader targeting
Compromised profile (suspicious announcements/changes)	Restore access + change passwords + 2FA + log out unknown sessions	Procedure for restoring access to the platform + institutions if there is abuse/threats
Doxxing or threats	Urgent safety assessment + preserve evidence + reduce exposure	Urgent action before institutions + contact AJM for coordination
Coordinated harassment (insults/organized attack)	Preserve evidence + careful moderation	Reporting violations on the platform + AJM if there is significant damage / multiple platforms
Manipulated "evidence" / cropped quotes	Archive source + prepare factual correction	Report if there is impersonation/abuse of privacy + inform partners as needed

RECORD KEEPING AND PRESERVATION OF EVIDENCE

Evidence is crucial for two purposes: to obtain a timely response from platforms and to provide a basis for action before competent institutions when there are threats, doxxing, compromised access or misuse of personal data. In a coordinated attack, evidence often disappears quickly - content is deleted, profiles are deactivated, and context is lost. Therefore, collection and preservation begin immediately and are done systematically. In some cases, such attacks are connected to broader manipulative campaigns in the information environment (including FIMI[3]), making disciplined evidence collection and treatment essential for identifying the overall pattern.

Basic principles

- Evidence is collected before interventions such as deletion, blocking, or changing settings, except when there is an immediate security threat.
- Evidence preserves context, because an isolated screenshot without surrounding elements is often not enough.
- Evidence is kept with limited access and with minimal processing of personal data, in accordance with the “do no harm” principle.
- Evidence is linked to the “incident log” to ensure a clear timeline and verifiability of activities.

Minimum evidence package

The minimum evidence package is a set that the editorial team provides in the first hours, regardless of the type of attack. This package is usually sufficient for initial reports to platforms and for a quick risk assessment.

- Each piece of evidence contains a date and time stamp and is associated with a specific link or profile.
- Each screenshot shows the entire screen with visible usernames, context, and at least some of the navigation.
- Each video screen recording shows how to open the link and how to access the content in question, to preserve context.

3 Foreign Information Manipulation and Interference

- Each threat is documented with full text and surrounding context, including a note on whether the threat is repeated or escalated.
- Every notification or email from the platform is saved in its original form, with subject, date, and content.

Expanded evidence package for platforms and institutions

When the attack is coordinated, the minimum package is expanded into an "evidence package" that can be delivered as a whole. This package should be readable and allow for quick action, without the need for additional explanations.

The evidence package includes:

- A brief summary of the incident with a description of the type of attack, duration, and damage.
- A timeline with the most important events, linked to the "incident log".
- A table of key links, profiles, and posts, with a short note indicating or confirming each element.
- Selected supporting evidence, organized by type of attack, such as mass false reporting, impersonation, compromised profile, doxxing, or threats.
- Clearly stated requests, for example restoring a profile, removing content, sanctioning a fake profile, or taking immediate action on threats.
- A contact point in the editorial office, with an official contact and a person in charge of monitoring the case.

Naming/labelling, organizing and access control

Systematic organization reduces the risk of losing materials and allows for faster delivery and use of evidence. One agreed format is sufficient, if applied consistently.

- The main folder has the date and short name of the incident.
- Subfolders are organized by platform and attack type to avoid mixing materials.
- Each file contains a date, time, platform, and a short description, so it can be found quickly and without searching by content.
- Access to evidence is limited to individuals who have a role in the incident, and sharing is recorded in the "incident log".

- Evidence is kept with at least one backup copy, in a secure location that is also accessible to the incident coordinator.

Minimization of personal data

Coordinated attacks often involve personal data and content that can further expose the targeted person. The editorial team is careful to preserve evidence while not causing additional damage.

- If evidence is shared outside the immediate incident team, two versions are prepared - “internal” and “for sharing” - with unnecessary personal data removed.
- If obfuscation or redaction is done, an original version with limited access is also kept, so as not to lose evidentiary value.
- If the evidence contains an address, phone number, or sensitive data, it is not forwarded to wider groups or published, even when it is part of the attack.
- If the evidence involves minors or family information, it is treated as high-risk and is only shared when necessary for the proceedings.

Integrity of evidence and confidentiality

Even in a practical context, basic discipline of integrity and confidentiality reduces the possibility of disputing authenticity.

- Each piece of evidence is preserved as close to its “original state” as possible, without any editing that alters the content.
- Each piece of evidence has a note about who obtained it, when and how, and this is also recorded in the “incident log”.
- Each sharing of an evidence package is recorded, with the recipient and purpose of the sharing specified.

DOXXING AND THREATS PROTOCOL[4]

Doxxing and threats are treated as high-risk situations because they can quickly escalate from online harassment to a real safety risk. In these cases, the priority is to protect the targeted individuals and their loved ones, followed by protecting profiles, content, and reputation.

This protocol is activated when there is at least one of the following elements: publicly released personal information, threat of physical violence, targeting of home/family, call for a "visit", stalking, or repeated and escalating threats.

The first 15 minutes - stabilization and safety

The first step is a quick assessment of whether there is an immediate threat. If the threat is concrete, imminent, or contains information that suggests the attacker knows the address/route/schedule, it is treated as a security emergency.

- ✓ The targeted person is notified directly and discreetly, without sharing the threat in wide groups.
- ✓ A brief "security status" is agreed upon - whether the person is at home, on the road, alone or with others, and whether they feel they are being followed.
- ✓ It is helpful for the person not to move alone and, when possible, to change the route, place or time of movement to reduce predictability.
- ✓ A contact person from the editorial office is appointed who remains available for coordination until the situation calms down.
- ✓ Evidence collection begins, but without forwarding personal data to individuals not involved in the incident.

The first 60 minutes – limit exposure

The goal in the first hour is to reduce the visibility of sensitive data and prevent further damage.

4 Council of Europe, "Improvement of national legislation on doxing and source protection", 02.07.2024, Platform to promote the protection of journalism and safety of journalists, https://www.coe.int/en/web/freedom-expression/-/improvement-of-national-legislation-on-doxing-and-source-protection

- ✓ Check whether personal data appears in multiple places and platforms, and then create a list of links and screenshots with context.
- ✓ Report content containing personal data and/or threats through platform privacy and harassment mechanisms, clearly stating that the content contains personal data and poses a risk.
- ✓ Public information on the profiles of the targeted person and the editorial team is reviewed, and unnecessary details are removed or limited.
- ✓ Close contacts are only notified to an extent that increases safety, not to an extent that creates panic or prompts additional data sharing.

When is the threat considered serious

A threat is considered more serious when it is concrete, repetitive, contains personal details, calls for immediate action, or comes from multiple profiles in coordination. In such cases, there is a basis to request immediate institutional protection and to conduct a procedure with a full evidentiary package.

- If there is immediate danger, the competent emergency services are contacted.
- If there is a threat, doxxing, or stalking, a report is made to the competent institutions with a clear timeline and evidentiary material.
- If necessary, additional physical protection measures are considered, such as a companion, a temporary change in routine, or a temporary avoidance of public appearances.

Evidence of doxxing and threats - what is most important to capture

In the case of threats and doxxing, the evidence should show exactly what was said or posted, who posted it, where and when, and whether there was an escalation in intensity.

- Each piece of evidence displays full context, including username, date/time, and link.
- Notifications or messages indicating that the content was visible or shared are also saved.
- Repetitions and variations of the threat are preserved, because amplification is an important indicator.
- It is noted whether the threat is related to a specific post, investigation, or topic, to understand the motive and the pathway of dissemination.

COORDINATION AND ACTION

Contact AJM when there are grounds

In cases of doxxing and threats, contacting AJM can help provide guidance, coordination, and public support when it is safe and beneficial. When reporting, a brief summary, links to key evidence, and an assessment of whether there is an immediate risk are sufficient.

Proceedings before competent institutions

When a report is filed, institutions benefit from a clear structure. It is usually sufficient to submit:

- A brief description of the event and why it is considered a threat or doxxing.
- A timeline of key events, linked to the “incident log”.
- Selected evidence showing the threat or disclosed personal information.
- Information on whether there is repetition, coordination, or sharing on multiple platforms.
- Contact point for further communication and submission of additional evidence.

Communication in the context of doxxing and threats

Public communication in such situations requires extra caution. Sometimes silence and limited communication yield better results than a public debate.

- A public message, if sent out, remains brief and factual and does not repeat personal information or details of the threat.
- Screenshots containing addresses, phone numbers, or other sensitive data are not published, even when the goal is to “show what’s going on.”
- No next steps are announced that would help attackers adapt, such as reporting details, security measures, or internal organization.

What to avoid in such situations

Certain reactions seem “logical” under stress, but they often worsen the condition or increase exposure.

- Do not engage in direct discussions with profiles that make threats or publish personal information.
- Do not encourage the public to “find” the attacker, as this may cause additional abuse and legal risks.
- Do not share evidence with broad groups, especially when it contains sensitive personal data.

WAYS TO FILE A REPORT, COMPLAINT AND TO REGAIN ACCESS TO PLATFORMS

Platforms often respond first through automated systems. Therefore, the outcome depends on whether the report is filed under the correct category, whether there is enough context, and whether the evidence is clear. In coordinated attacks (including cases that carry elements of FIMI - synchronized targeting and imposition of false narratives), discipline in facts and evidence is especially important.

This section provides practical pathways for reporting, filing complaints and regaining access. Links are provided in the footnotes, in order to keep the text readable and usable under pressure.

General rules for reporting and complaints

- Evidence is always preserved first, and then intervention is done by reporting, deleting, or blocking.
- The report is submitted under the most appropriate category, as the wrong category often results in an automatic rejection.
- One report is more effective when it is complete and consistent, than multiple reports with insufficient context.
- Each submission is recorded in an “incident log”, with date, time, link, and status.
- Reports do not repeat unnecessary personal information, especially in the case of doxxing and threats, but instead use a description and attached evidence.

Checklist - the minimum that the report should contain

- The report begins with a clear sentence that names the violation and exactly what action is requested.
- The report contains a direct link to the content/profile and at least one other example that shows the context.
- The report provides a brief timeline: when it started, whether it is ongoing, and whether there has been an escalation in scope/intensity.
- The report states why the attack is considered coordinated (repeated profiles, synchronization, mass reporting, cross-platform activity).
- The report ends with a specific request: removal of content, sanctioning of a fake profile, restoration of access, or immediate action in response to a threat.

Meta - Instagram, Facebook and Threads

1. Mass reporting and reach limitation

When there is a suspicion that a profile is being targeted through mass false reports, it is practical to react in two ways: to preserve the notifications/sanctions and to file an appeal through the available mechanisms for “review” or “appeal”. If a restriction, removal or suspension occurs, it is useful to preserve the entire notification from the platform and attach it to the evidence package. This type of case often requires persistence and consistency in appeals.

2. Impersonation

When there is a profile posing as a journalist or media outlet, the quickest way is to report impersonation and submit a link to the official profile being impersonated.

3. Doxxing and privacy

When personal data is published, the privacy and harassment categories are used, with a clear statement that it is a misuse of personal information. The text of the report itself avoids repeating addresses, telephone numbers, and other sensitive details. Instead, screenshots and links are attached.

4. Compromised profile and regaining access

When an account is compromised or you suspect unauthorized access, it is most important to follow official recovery paths and avoid “advice” from unverified profiles offering “help.” In such cases, it is useful to start from a device that has previously been used to log in, as platforms often use this as a signal of trust.

Official links - Meta

- Instagram/Threads - impersonation report (form): https://help.instagram.com/contact/636276399721841.
- Instagram - instructions for reporting impersonation: https://help.instagram.com/370054663112398.
- Instagram/Threads - reporting a privacy violation: https://help.instagram.com/contact/512241091300432.

- Instagram - reporting harassment/bullying: https://help.instagram.com/547601325292351.
- Meta Account Recovery Hub (Facebook/Instagram/Threads): https://www.meta.com/account-recovery-support/.
- Instagram - „hacked account“ support: https://www.instagram.com/hacked/.
- Facebook - „hacked account“ tool: https://www.facebook.com/hacked.

X

1. Harassment and abuse

Reporting on X is most often done via the “Report” option directly from a post or profile, by selecting a category and providing additional context.

2. Private information and doxxing

The social platform “X” has a separate policy for private information and reporting abuse. In the case of doxxing, it is useful for the report to be clear, with a link to the post and a brief explanation of why the posted information is private and how it is being used for abuse.

3. Compromised profile

For a compromised profile, X offers a form to restore access. If the profile is “locked” or “limited”, it is practical to follow the steps for verification and restoration of functions.

Official links - X

- Reporting abuse: https://help.x.com/en/safety-and-security/report-abusive-behavior.
- Personal information policy: https://help.x.com/en/rules-and-policies/personal-information.
- Hacked or compromised form: https://help.x.com/en/forms/account-access/regain-access/hacked-or-compromised.

YouTube

YouTube explicitly treats threats and doxxing as prohibited conduct under its harassment and cyberbullying policies. When reporting, it is helpful to include a specific link, a timestamp on the video, and a brief explanation of how the content targets a person or media outlet.

Official links - YouTube

- Harassment and cyberbullying policy: https://support.google.com/youtube/answer/2802268.

TikTok

TikTok has a centralized “Report a problem” menu and specific steps for reporting a profile, content, or abuse. In coordinated attacks, it is helpful to collect several examples and report them with a clear indication of repetition and coordination. If an account is compromised, TikTok provides basic steps to restore security.

Official links - TikTok

- „Report a problem“: https://support.tiktok.com/en/safety-hc/report-a-problem.
- How to report a user: https://support.tiktok.com/en/safety-hc/report-a-problem/report-a-user.
- If the account has been hacked: https://support.tiktok.com/en/log-in-troubleshoot/log-in/my-account-has-been-hacked.

When reports don't yield results

Sometimes platforms don't respond quickly, or decisions are inconsistent. In such cases, it's helpful for the editorial team to quickly "refresh" the approach:

- ✓ Check whether the application category is correct and whether there is sufficient context.
- ✓ An appeal or second report is filed only when there are new facts or additional evidence.
- ✓ The evidence package is consolidated into one readable whole, instead of sending fragments.
- ✓ Contacting AJM is considered when there is high risk, recurring sanctions, or significant damage to operations.

Quick reporting and support request package

When an attack is coordinated, most of the time is spent on "assembling the story" from multiple messages and files. That's why it's useful to have a standard "quick package" that is sent as a **single message**, with a clear structure and **a few key pieces of evidence**. Documenting and logging (date, link, what happened) is important because content and profiles often disappear or change.

When to contact the AJM, and when to contact institutions directly

Contact AJM when there is a clear need for coordination and faster support, for example:

- A profile/page has been restricted, taken down or is at real risk of suspension due to mass false reports
- impersonation of a journalist/media outlet (especially if it involves collecting personal data or spreading harmful disinformation)
- compromised access or serious takeover attempts (including repeated phishing towards the team)
- publicly posted personal information for the purpose of harassment, doxxing, or threats of violence
- significant operational damage (disruption of publication, loss of access to audience, serious blow to trust)

Institutional action is considered when there is:

- a personal safety risk
- published personal data or doxxing, especially address/phone/family details
- threats of violence or calls for “visitation” / stalking
- unauthorized access to profiles and indications of misuse of personal data
- organized coordination that creates measurable damage

What is the “fast package” - 8 elements you send in one message

- **Subject/title:** Coordinated attack - [media outlet/journalist] - [platform] - [date]
- **Short summary** (4 - 6 sentences): what is happening, what points to coordination, status (is it ongoing)
- **Timeline:** start + key moments + current status
- **Attack type:** mass false reporting / impersonation / access takeover / doxxing / threats / other
- **Who is targeted:** at the role level (journalist, editor, page admin), without unnecessary personal details
- **What steps were taken:** the most important steps (report, complaint/appeal, changed passwords, 2FA, restricted access)
- **What are you asking for:** guidance, coordination, legal support, public reaction, contact with institutions (only what is really needed)
- **Evidence:** 3 - 5 key links + attachment/folder with “mini package”

A practical rule: **one complete package** is better than multiple “I’ll send more later on” messages.

“Mini package” of evidence - enough to get started

- Several **screenshots with context** (full screen, username, visible link/title, part of the environment)
- A **short video screen recording** when the context is critical (how to open the link and where to view the content in question)
- The most important **notifications/emails** from the platform (restriction, warning, suspension)
- Short **table of links** (3 - 10 rows) with a “what does this prove” column
- If there are threatening emails - save them in their original form (do not “forward” them as the only evidence unnecessarily), so that technical data/context is not lost.

Template 1 - Quick message to AJM

Subject: Coordinated attack - [media outlet/journalist] - [platform] - [DD.MM.YYYY]

Summary (4 - 6 sentences):

- ✓ Describe what is happening:
- ✓ How is coordination evident:
- ✓ Status: (in progress / has settled down / resurfacing)
- ✓ Main risk: (loss of profile / security / other)

Attack Type: [select]
Target: [role - no personal data]
What steps have been taken: [3 - 6 short points]
What we are asking from AJM: [1 - 2 sentences]

Key links (3 - 5):

- ✓ [link] - what does it show?
- ✓ [link] - what does it show?
- ✓ [link] - what does it show?

Attachment/folder: [restricted access link or “mini package” attachment]
Contact: [name, position, work phone, work email]
Backup contact: [name + contact]

Template 2 - “Incident Log” (minimal)

This is an internal record that creates a timeline and saves you from “where was that link?” (It is recommended that it be a table that is updated.)

Incident Name: [short]
Date: [DD.MM.YYYY]
Incident Coordinator: [name and role]
Status: [ongoing / stabilized / closed]

Time	Platform	Link/location	What happened?	What did we do?	Proof
[HH:MM]	[IG/FB/X...]	[link]	[1 sentence]	[1 sentence]	[file name/ link]

Template 3 - Table of evidence (index)

Tag	Date/time	Platform	Link	What does it show?	Why is it important?
D1	[DD.MM.YYYY HH:MM]	[ ]	[ ]	[ ]	[threat/doxxing/impersonation/access hijacking...]

Template 4 - Mini format for reporting to institutions (if applicable)

Subject: Notification/report of [threats / doxxing / unauthorized access / organized harassment] - [DD.MM.YYYY]

1. **Short description:** (5 - 8 sentences: what, when, where, who is targeted - at the role level)
2. **Why is it a risk:** (1 - 2 sentences: security / privacy / access / harm)
3. **Key evidence:** (D1, D2, D3 + one sentence for each, to briefly describe them)
4. **Damage so far:** (specifically: restricted profile, personal data published, threats, lost access)
5. **Request for action:** (what do you expect: admission, referral, urgent action, protection of personal data)
6. **Contact and confidentiality:** (contact point + sentence that the evidence contains sensitive data and is for official use)

Practical rules for privacy when sharing

- Do not repeat addresses/phone numbers in the text - refer to the evidence where they appear.
- If you are sending to multiple recipients - send a “sharing” version (with unnecessary personal information removed), and keep the original with limited access.
- Do not send evidence to large groups/chats if it contains personal information.

CASE STUDY - “SLOBODEN PECAT”

This case shows how a coordinated attack can start as “unusual comments” and then shift to other forms that are harder to control, such as mass likes and follows, with consequences to regular operations and the visibility of content.

Context

The official Instagram accounts of “Sloboden Pecat” are @slobodenpecat and @sloboden-pecat.vesti. The accounts have been maintained continuously for about eight years, with organic growth and without the use of paid promotions for followers or reach.

Brief overview of the incident

The incident took place between 06.11.2025 and 10.11.2025 on Instagram. The attack began with mass identical comments, and then shifted to mass likes and follow requests from suspicious profiles. The main consequences were disruption of the work of the newsroom and a drastic drop in reach, with suspicion of algorithmic limitation of visibility.

How did the attack unfold?

On 06.11.2025, a first wave was observed - a large number of identical comments with the word “like” submitted in the same minute, from profiles with no prior interaction with the media outlet and with characteristics typical of falsely generated or bot profiles. The same phenomenon happened again in a short period of time, under subsequent posts.

During that same day, a second wave took place - massive comments with politically charged content and repetitive formulations, in which “Sloboden Pecat” was accused of censorship and deletion of comments, with narratives related to international conflicts. The comments came from profiles with no history of interaction with the outlet, which further pointed to coordination.

After activating the platform’s security features and commenting restrictions, the mass comments decreased, but the attack was redirected. On 09.11.2025, a sharp spike in likes from the same type of suspicious profiles was observed on a post with high reach. Since there is no practical mechanism for selectively “cleaning” mass likes, the profile was made private in order to protect the continuity of work and the organically built audience. On 10.11.2025, more than 999 follow requests were registered, most of which were from obviously fake generated profiles, which required manual cleaning.

Evidence trail and indicators

The case is an example of how evidence should reflect the timeline and show the "evolution" of the attack.

- Links to posts where the first signals occur, along with screenshots showing context and repeatability.
- Visualizations and comparisons of statistics ("views over time") that indicate a sharp deviation in reach after the attack, as a basis for suspicion of algorithmic throttling.
- A screen recording showing the atypical dynamics of likes and follows.
- Platform notifications, when available, as additional proof of status and measures.

Practical lessons that can be applied

1	The attack rarely stays in one form, so limiting comments can shift the attack to likes, follows, or other platform mechanisms.
2	Manual "cleaning" quickly consumes the newsroom's capacity, so it is useful to activate an incident mode with roles and a timeline early on.
3	Commenting and moderation restrictions are effective when combined with systematic evidence collection and consistent reporting to platforms.
4	Temporary lockdown ("privating") can be a tactical measure when an attack creates a massive technical overload, but such a measure is more effective if it is foreseen in advance as an option in the response plan.
5	The evidence package is stronger when it shows the phases of the attack and the actual damage, not just individual screenshots without context.
6	Early coordination with the AJM and recording the case with competent institutions creates a basis for further support and a better institutional response in future cases.

NEWSROOM READINESS CHECKLIST

A. Roles and coordination	✓ Assign an incident manager and 2 backups (minimum: decisions, profile management, evidence collection). ✓ Agree on one internal channel for quick coordination and a rule: decisions go through the manager. ✓ Prepare a list of key contacts (internal, IT/admin, legal support, AJM).
B. Account Access and Security	✓ Enable 2FA (two-factor authentication) for key accounts and work emails; keep backup codes safe and with limited access. ✓ Regularly review administrator roles and remove "forgotten" accesses. ✓ Use strong and unique passwords (password manager, where possible). ✓ Avoid logging into official profiles from public networks and unsecured devices, especially during periods of risk.
C. Early detection and moderation	✓ Agree on who monitors notifications from platforms and who responds to alerts about suspicious activity/restrictions. ✓ Be alert to "red flags": sudden spikes in comments/follows, synchronized phrases, repetitive profiles, transferring the attack to other platforms. ✓ Have short moderation rules (what gets deleted immediately, what gets documented and reported, what gets left as legitimate criticism). ✓ Do not mass delete without first securing evidence (screenshots, links, timeline).
D. Psychological and team support	✓ Introduce a brief "check-in" with the targeted person within the first 24 – 48 hours. ✓ Consider temporarily reassigning tasks if the pressure is intense or there are threats/doxxing. ✓ When available, encourage professional support (especially for recurrent attacks).
E. Drills and updates	✓ Run a short simulation for 30 – 45 minutes (mass sign-up or impersonation) 1 – 2 times a year. ✓ After each incident, write down the "lessons learned" and update your contacts, accesses, and templates.

KEY SOURCES

Platform policies and tools (reporting, complaints/appeals, account recovery)

Meta. „Making it Easier to Access Account Support on Facebook and Instagram“. Meta, 04.12.2025. Accessed on 20.01.2026. https://about.fb.com/news/2025/12/making-it-easier-to-access-account-support-on-facebook-and-instagram/

Meta. „Account Recovery Hub - Facebook, Instagram, Threads“. Meta, undated. Accessed on 20.01.2026. https://www.meta.com/account-recovery-support/

Meta. „Inauthentic Behavior (Community Standards)“. Meta Transparency Center, undated. Accessed on 20.1.2026. https://transparency.meta.com/policies/community-standards/inauthentic-behavior/

Instagram. „Hacked Instagram Account“. Instagram Help Center, undated. Accessed on 20.1.2026. https://help.instagram.com/368191326593075/

Instagram. „Report an Impersonation Account on Instagram or Threads“. Instagram Help Center, undated. Accessed on 20.01.2026. https://help.instagram.com/contact/636276399721841

X. „X's Private Information Policy and Doxxing“. X Help Center, undated. Accessed on 20.1.2026. https://help.x.com/en/rules-and-policies/personal-information

X. „How to Report Abusive Behavior on X“. X Help Center, undated. Accessed on 20.1.2026. https://help.x.com/en/safety-and-security/report-abusive-behavior

YouTube. „Harassment & Cyberbullying Policies“. YouTube Help, undated. Accessed on 20.1.2026. https://support.google.com/youtube/answer/2802268

YouTube. „Report Inappropriate Videos, Channels & Other Content on YouTube“. YouTube Help, undated. Accessed on 20.01.2026. https://support.google.com/youtube/answer/2802027

TikTok. „My Account Has Been Hacked“. TikTok Support, undated. Accessed on 20.01.2026. https://support.tiktok.com/en/log-in-troubleshoot/log-in/my-account-has-been-hacked

Practical resources for digital security and incident response

Access Now. „Digital Security Helpline“. Access Now, undated. Accessed on 20.1.2026. https://www.accessnow.org/help/

Committee to Protect Journalists (CPJ). „Digital Safety: DIY Guides for Better Protecting Against Online Harassment“. CPJ, 04.09.2019. Accessed on 20.01.2026. https://cpj.org/2019/09/digital-safety-diy-guides/

Committee to Protect Journalists (CPJ). „Digital Safety: Remove Personal Data from the Internet“. CPJ, 04.09.2019. Accessed on 20.01.2026. https://cpj.org/2019/09/digital-safety-remove-personal-data-internet/

Committee to Protect Journalists (CPJ). „Editors’ Checklist: Protecting Staff and Freelancers Against Online Abuse“. CPJ, 07.07.2022. Accessed on 20.01.2026. https://cpj.org/2022/07/editors-checklist-protecting-staff-and-freelancers-against-online-abuse/

Coalition Against Online Violence. „Online Violence Response Hub“. Undated. Accessed on 20.1.2026. https://onlineviolenceresponsehub.org/

European and international standards relevant for practical protection

European Union. “Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19.10.2022 on a single market for digital services (Digital Services Act)”. EUR-Lex, 19.10.2022. Accessed on 20.1.2026. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

European Commission. “Recommendation (EU) 2021/… - on the Protection, Safety and Empowerment of Journalists (C(2021) 6650 final)”. EUR-Lex, 16.09.2021. Accessed on 20.1.2026. https://eur-lex.europa.eu/eli/reco/2021/1534/oj

Council of Europe. „Recommendation CM/Rec(2016)4 on the Protection of Journalism and Safety of Journalists and Other Media Actors“. Council of Europe, 13.04.2016. Accessed on 20.1.2026. https://rm.coe.int/16806415d9

Council of Europe. „Freedom of Expression and Journalist Safety - A Pocket Guide for Users“. Council of Europe, 09.2023. Accessed on 20.1.2026. https://rm.coe.int/pocket-guide2023-final/1680ac9318

OSCE - Representative on Freedom of the Media. “Guidelines for Monitoring Online Violence Against Female Journalists”. OSCE, commissioned 2022 (PDF). Accessed on 20.1.2026. https://www.osce.org/files/f/documents/b/0/554098_1.pdf.

www.znm.org.mk