SOCIAL MEDIA STORMS AND NUCLEAR EARLY WARNING SYSTEMS: A DEEP DIVE AND SPEED SCENARIOS WORKSHOP REPORT

Recommended Citation

Nautilus Institute, Technology for Global Security, Preventive Defense Project, "SOCIAL MEDIA STORMS AND NUCLEAR EARLY WARNING SYSTEMS: A DEEP DIVE AND SPEED SCENARIOS WORKSHOP REPORT", NAPSNet Special Reports, January 08, 2019, https://nautilus.org/napsnet/napsnet-special-reports/social-media-storms-and-nuclear-early-warning-systems-a-deep-dive-and-speed-scenarios-workshop-report/

NAUTILUS INSTITUTE, TECHNOLOGY FOR GLOBAL SECURITY, PREVENTIVE DEFENSE PROJECT

JANUARY 8, 2019

I. INTRODUCTION

This workshop concluded that: "[I]ndividuals, organizations, and even states may start to use social media to try to provoke nuclear attacks against their adversaries; or for other political-ideological or religious reasons; that they will be effective in terms of reaching some highly influential people as well as large numbers of people; and included in these two types of readers are likely to be some people making nuclear early warning assessments, and nuclear command decisions" and: "The workshop participants were convinced that social media platforms and social media users can shift the center of gravity away from the current, celebrity-driven and conflict-amplifying social media conflict amplifying dynamic that degrades the quality of much information and towards more reliable, authenticated information while preserving the ability of users to free speech and near-instantaneous networking.

This report was prepared by staff of Nautilus Institute, Preventive Defense Project Stanford University, and Technology for Global Security with assistance from N Square Collaborative. It is the report of the October 20, 2018 workshop on *Social Media Storms and Nuclear Early Warning Systems.* The workshop was co-sponsored by the Nautilus Institute, the Preventive Defense Project—Stanford University, and Technology for Global Security, and was funded by the MacArthur Foundation. This paper is published simultaneously here by Technology for Global Security.

The Executive Summary is published below. **The full PDF report is downloadable here.**

Banner image: by Grace Farley, Technology for Global Security.

II. NAPSNET SPECIAL REPORT BY NAUTILUS INSTITUTE, TECHNOLOGY FOR GLOBAL SECURITY, PREVENTIVE DEFENSE PROJECT

SOCIAL MEDIA STORMS AND NUCLEAR EARLY WARNING SYSTEMS: A DEEP DIVE AND SPEED SCENARIOS WORKSHOP REPORT

JANUARY 8, 2019

EXECUTIVE SUMMARY

Social Media Emerges as a Possible Trigger of Nuclear Early Warning Systems

This workshop was convened because social media burst onto the stage of nuclear warfare in 2018.

In the Asia-Pacific region alone, six instances of social media playing a role in nuclear-prone conflicts occurred between August 2017 and January 2018. For decades, strategists have worried about the possibility that states armed with nuclear weapons might mistakenly launch a nuclear strike due to a false alarm originating in its early warning system or due to degraded decision-making.

Why does this matter? The primary reason is the terrifying combination of speed and unique scale of violence when it comes to nuclear weapons that continues to set them apart from all other means of coercion. The simple problem is that nuclear commanders must make decisions to use nuclear weapons for mass destruction in time measured in minutes and seconds, not hours and days.

How should a nuclear weapons state treat the vast amounts of social media, including content that may be factually accurate that is transmitted almost instantly, but may also be created by malevolent parties that aim to deceive, manipulate, and mislead its early warning system and pollute its nuclear command decision-making process with fake information?

In addition to "traditional" sources of NC3 error such as accidents, hardware failure, and human error, new technologies such as including artificial intelligence, autonomous vehicles, quantum computing and sensing superimposed on the legacy US NC3 system may create new types of coincident error involving social media.

In the past, the insulation of nuclear commanders from unclassified data and their near total dependence on official, classified information systems in the midst of crisis might have served to reduce the influence of erroneous or deceptive information. In today's world where the President of the United States makes public declaratory policy on Twitter, one cannot be sure that social media will not be influential.

Lessons from Social Media in Non-Nuclear Domains

To investigate how social media might play out in the world of nuclear early warning the workshop considered the use of social media to promote extremist views and behavior in promoting anti-vaccination, anti-Semitism, gang, ethnic, and terrorist violence in cities. From this evidence, two sets of lessons were drawn. The first focused on the issue of false data and false alarms leading to conflict as well as conflict escalation. The second lesson primarily focused on what antidotes exist for false alarms on social media or via other ways of creating an authoritative and credible reference knowledge.

Citing detailed case studies and data, presenters showed that social media is proving capable of quickly mobilizing fear and channeling aggregated animosity against real-world targets. Therefore, in these domains there are precursors of how activists might attempt to mobilize mass sentiment online to generate fear and alarm in the face of a nuclear threat. Additionally, the targeting of nuclear commanders or key individuals in a nuclear command and control system may occur in an effort to launch a nuclear attack against a third party presented as worthy only of nuclear annihilation.

The conclusion is unavoidable that individuals, organizations, and even states may start to use social media to try to provoke nuclear attacks against their adversaries; or for other political-ideological or religious reasons; that they will be effective in terms of reaching some highly influential people as well as large numbers of people; and included in these two types of readers are likely to be some people making nuclear early warning assessments, and nuclear command decisions.

A variety of strategies were described that partly ameliorated the problem, although generally the

response by social media platforms themselves was slow, often only at the behest of outsiders, and inhibited by many problems of attribution and tracking of online aggressors.

Where social media platforms have attempted to curtail manipulative and dangerous use of their services, they have found that one of the best ways to do so is to simply slow down the ability of users to run their campaigns. This can be achieved by automated systems that shut down sites found to be fake sites and by identifying core behaviors used by social media activists that become markers that can be used to create traction and slow down their ability to game the rules of social media by using bots, anonymity, etc. One problem in such regulation of online behavior is that context is critical to determining what is and is not dangerous, and a human must be in the loop. However, once deep fakes become widespread, even having a human involved may not suffice to determine the truth content of a specific post or site.

Given the speed of nuclear early warning and nuclear command decisions, following these pathways is likely too little, too late. Indeed, as one social media practitioner said, "If we find something [really bad] in the nuclear community, then who do we really tell is the question!"

One possibility is that first responders, especially at the city level, might use their information systems to provide credible information to reporters and mass media; and to fold carefully evaluated social media reports into their information after first validating the early reports with a variety of real-time sensors. First responders noted that it is essential to ensure that the messages across cities and counties are aligned and consistent, to avoid public confusion and disenchantment with official sources of early warning. As one first responder noted, "You need to have a battery of circuit breakers to preempt and steer false information away - one message is not going to solve it."

Thus, even in the case of threats with extremely short timelines—minutes in the case of bombing, seconds in the case of earthquakes, it is possible to use social media to inform and guide humans to respond in ways that reduce risk, without the risk of the alert system being hijacked by malevolent actors. The critical element is that the users trust the reliability and authenticity of the information sent out.

Circuit Breakers & Short Circuits Speed Scenarios

Participants developed four "short circuit" scenarios that explored how and what circuit breakers might be created that avoid or overcome the destabilizing effect of social media on nuclear early warning systems and nuclear command decisions.

The four scenarios and circuit breakers include:

Korea A Sweltering Crisis—The United States and North: Multiple circuit breakers were prefigured including launching fact-finding missions, improving official communications with mass media, and leveraging trusted third parties willing to put themselves on the line as exchange-hostages to push the international community to a peaceful solution, reestablish trust, and mitigate panic.

Fast and Furious—India, Pakistan, and a Non-State Actor: The circuit breaker in this firestorm scenario aimed to create the necessary political space to deescalate the situation, starting with mitigation strategies set in motion by social media luminaries and companies to slow down communication combined with inter-state backchannel conversations and more traditional mutual hostage taking to stop conflict from spiraling out of control in conditions of extreme nuclear provocation.

Mutual Miscalculations: NATO, United States, and the Russian Federation: The circuit breaker set

out to answer the question: Are there ICBMs in the air and if so from where? While dealing with increasing panic. The circuit breaker is to create a new international organization with only one mission--to provide sensor-based information on international incidents that may related to nuclear war, to validate data events; and impartially to send data to all parties to a nuclear prone conflict.

Embrace Tiger, Retreat to Mainland—China, Taiwan, and the United States: Similar to the last scenario, the key question in this scenario is: What's the nature of the Chinese missiles, nuclear or conventional, that are in the air; and will they hit Guam and Okinawa or splash down nearby? In this case, China initiates a last minute, last second concerted, all-out *diplomatic* effort to stop a US retaliatory strike by calling on trusted persons to act as intermediaries. They activate personal backchannels including businessmen, politicians, and even religious leaders. Their first act (because it is the fastest) is to turn off aggressive social media in China itself.

Conclusions

Some themes recurred across all the scenarios, suggesting that these elements might lend robustness to many de-escalatory strategies. For example, whatever the role of social media in in creating or amplifying nuclear-prone conflicts, high-level communication between trusted individuals was an ingredient in resolving conflict. Other elements included high and low-level hostage exchange; doing whatever it took to slow escalatory spirals; and anticipating the loss of control induced by social media and other drivers by establishing hot lines, market and civil society-based communication channels, and trusted, third party, and impartial sources of authoritative information on the status of forces.

The workshop participants were convinced that social media platforms and social media users can shift the center of gravity away from the current, celebrity-driven and conflict-amplifying social media conflict amplifying dynamic that degrades the quality of much information and towards more reliable, authenticated information while preserving the ability of users to free speech and near-instantaneous networking. In this regard, cities and civil society emerged as a set of actors and networks that may be positioned to create new forms of governance and public information goods that restrains the aggressive use of social media that may contribute to false alarms and poor decision-making at the national level, while contributing to independent, impartial and validated information that is useful to nuclear early warning systems and nuclear commanders who may be relatively poorly served by traditional sensors, early warning systems, and conflict resolution mechanisms at the level of inter-state conflict.

III. NAUTILUS INVITES YOUR RESPONSE

The Nautilus Asia Peace and Security Network invites your responses to this report. Please send responses to: nautilus@nautilus.org. Responses will be considered for redistribution to the network only if they include the author's name, affiliation, and explicit consent.

View this online at: https://nautilus.org/napsnet/napsnet-special-reports/social-media-storm--and-nuclear-early-warning-systems-a-deep-dive-and-speed-scenarios-workshop-report/

Nautilus Institute
608 San Miguel Ave., Berkeley, CA 94707-1535 | Phone: (510) 423-0372 | Email: nautilus@nautilus.org