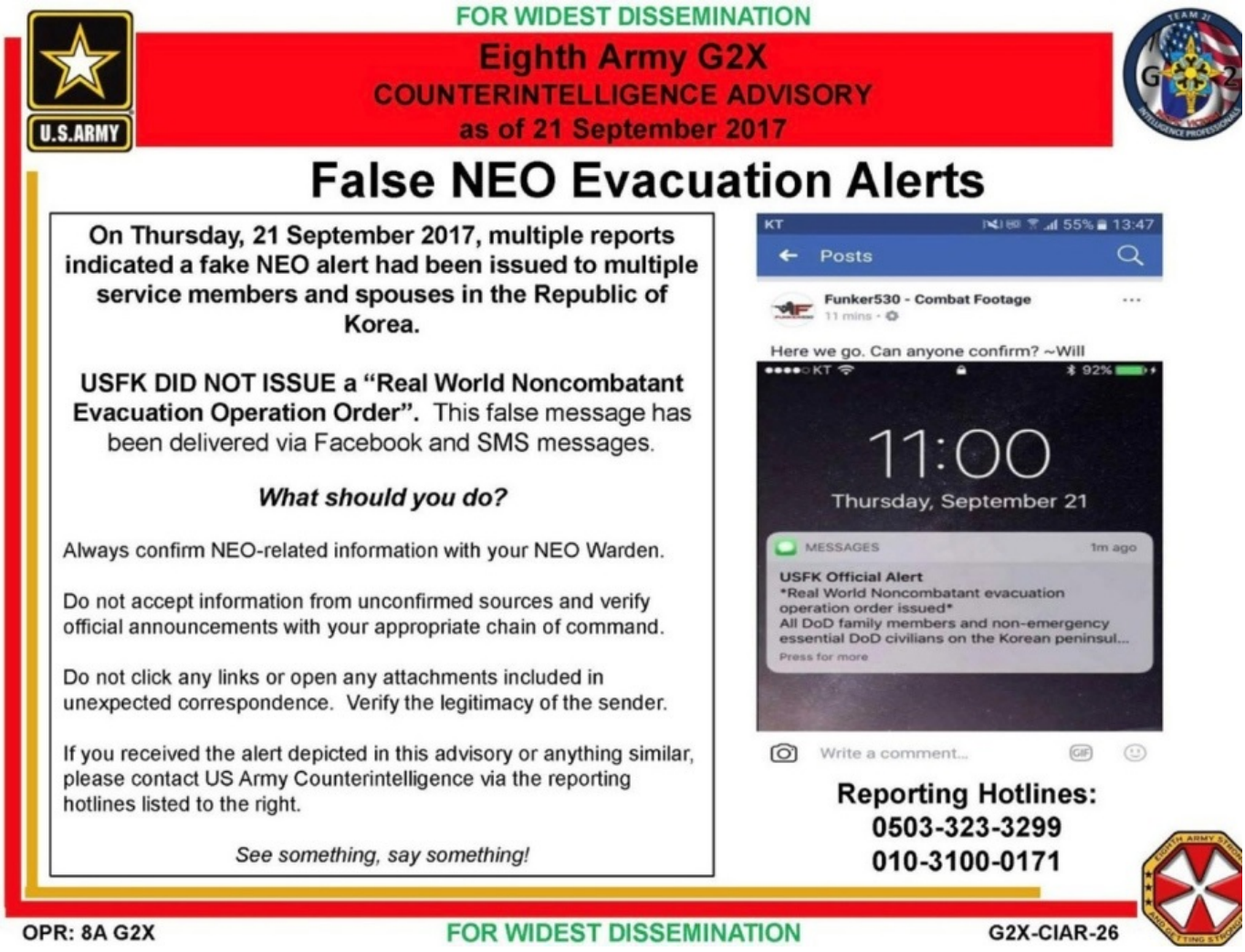
REDUCING THE RISK THAT SOCIAL MEDIA STORMS TRIGGER NUCLEAR WAR: ISSUES AND ANTIDOTES



Recommended Citation

Peter Hayes, "REDUCING THE RISK THAT SOCIAL MEDIA STORMS TRIGGER NUCLEAR WAR: ISSUES AND ANTIDOTES", NAPSNet Blue Peter, February 01, 2018, https://nautilus.org/napsnet/reducing-the-risk-that-social-media--torms-trigger-nuclear-war-issues-and-antidotes/

PETER HAYES

FEBRUARY 1, 2018

I. INTRODUCTION

In this essay, Peter Hayes concludes that: **"**Some form of independent, impartial, and pre-existing global early warning system may be useful to all nuclear armed states to provide a supplementary way to check not only the truth content of these alerts and social media storms, but the operation of their own "routine" early warning systems."

Peter Hayes is Director, Nautilus Institute and Professor, Center for International Security Studies, Sydney University.

Acknowledgments: This report was funded by MacArthur Foundation.The author thanks Paul Bracken, Chris Diehl, Deborah Gordon, Philip Reiner, Eric Sears, and Steve Weber for insight on this topic. The author remains solely responsible for its content and any errors.

The views expressed in this report do not necessarily reflect the official policy or position of the Nautilus Institute. Readers should note that Nautilus seeks a diversity of views and opinions on significant topics in order to identify common ground.

Image: USFK message declaring false an message ordering evacuation of military noncombatants, from here.

Note: This text was updated on February 1,2018 at 8.30pm PST.

II. NAPSNET BLUE PETER BY PETER HAYES

REDUCING THE RISK THAT SOCIAL MEDIA STORMS TRIGGER NUCLEAR WAR: ISSUES AND ANTIDOTES

FEBRUARY 1, 2018

In the last year, false news and false alarms in Seoul, Guam, Hawaii, and Japan highlighted the risk that social media or what the Chinese call "Internet storms," hereafter termed social media storms, may trip the early warning systems of nuclear armed states, thereby compounding and increasing the risk of war and nuclear war (see Table 1).

Table 1: False Alarms and Social Media Storms about Nuclear and Missile Attacks

SEOUL: D. Lamothe, "U.S. families got fake orders to leave South Korea. Now counterintelligence is involved," *Washington Post*, September 22, 2017.

GUAM: E. Adamczyk, "Guam residents unnerved by accidental 'civil danger warning,' UPI, August 15, 2017.

HAWAII: A. Wang, B. Lyte, 'Ballistic Missile Threat Inbound To Hawaii,' the alert screamed. It was a false alarm," *Washington Post,* January 13, 2018 .

TOKYO: AP, "Japan public TV sends mistaken North Korean missile alert," January 16, 2018.

***

Such social media contagions and early warning system false alarms are not new. In fact, a similar event occurred after a US Trident missile launch in 2016;[1] and radio false early warning alerts (one insider instigated) in St. Louise was experienced not once but twice (1971, 1991).[2]

However, the coincidence of these events with cycles of high tension and the increasing risk of war between the United States and the DPRK, including constant references to nuclear war by their leaders, suggests that it is urgent to create a system to offset this risk, and to serve as an early warning antidote for those charged with providing early warning to leaders of nuclear armed states.

Important as they are, the key issue is not the curtailment of psychological and other harms arising from early warning false alarms and/or social media contagions, including fear (of dying in ten minutes), loathing (aliens are attacking), rejoicing (the second coming), anger (that states can screw up so massively), and other "fallout" for the alarmed populations. This is an important topic in itself, but bears only tangentially on how the content of social media contagious response might amplify the significance and meaning of an unintended early warning indicator to a nuclear adversary.

Rather, the primary concern if this essay is with how nuclear adversaries may read social media response to false early warnings and false positives to the effect that nuclear attack may be underway when it is not, and may in turn combine this almost instantaneous information with other indicators to advise nuclear commanders that they may be or are under attack. Thus, in the US-DPRK conflict context, of the four events listed in Table 1, the first was the most dangerous because such non-combatant evacuation is likely a critical indicator to North Korean analysts that the United States is preparing to attack it.

Making such determinations in a given nuclear early warning system, even when as endowed as the United States, is not a simple matter. A simple taxonomy in Table 2 shows at least six combinations of the type of report (the report is positive and an attack is underway, contested, or is negative, there is no attack) versus the possible truth content of such reports (false, mixed, true). Of course, each side will have its own reading of the probable status of a report—one side (for example, the DPRK) may view it genuinely as a TP—that a nuclear attack is underway, while its nuclear adversary in which the false alarm and social media event originates (for example, the United States) might read it as a FN—that no attack is underway, while a third but interested nuclear armed state (for example, China or Russia) might read it as a MC—a partly truthful indicator that some kind of nuclear attack is underway, but notes that it is contested. Adding to the complexity is that these false alarms and social media storms may take place at the same time in more than one party to a nuclear-prone conflict.

To this taxonomy might be added a sudden "no report at all" or silence (in contrast to routine silence, no alarm, no related social media storm) in that a report system suddenly blacks out—as when a forward-based early warning radar suddenly goes offline which might be due to first nuclear attack to blind a nuclear armed state, or to an accidental loss of the site or connectivity back to the early warning fusion center. The parallel between the classic "silences" listened for in the American early warning system in this era might be the sudden "silencing" of a negative or contested report, that suggests that an attacking state has somehow severed social media connectivity in a launch area, creating a sudden and suspicious social media black-out.

False Alarms, Social Media Storms and Early Warning A-Symmetry

Although this false alarm and social media storm issue has arisen primarily in the context of the United States and the DPRK, this issue is generic and global, and pertains to all the nuclear weapons states, and to the possible impact of social media on the early warning systems and decision-making in all of these states.

Table 2: Combinations of Truth Content and Report

	TRUTH CONTENT		
REPORT	**F: FALSE**	**M: MIXED**	**T: True**
P: POSITIVE attack is under way	FP	MP	TP
C: CONTESTED attack is, is not underay	FC	MC	TC
N: NEGATIVE no attack	FN	MN	TN

As follows from Table 2 and discussion above, the impact of a false alarm and/or related social media event that may trigger an adversary's early warning system may vary from one nuclear armed state to another, and is always relative to a specific set of nuclear conflict relationships. The United States, for example, has global, omnipresent, and almost instantaneous surveillance capacities drawing on redundant sensor types, monitoring platforms, and communication channels. These redundant systems enable it to cross-check an alert and/or social media storm that suggest a nuclear attack is underway for validity (to determine whether the "argument" of the alarm and/or social media storm is in principle logically possible) and soundness (is there evidence that the argument advanced by the alert and on social media is correct).

Thus, the United States can determine almost immediately if a social media storm in China or North Korea is matched by other data streams that an attack is underway, place in the context of level of tension and other indicators that war might be about to break out, and if the alert and social media storm is judged inconsistent and false, dismiss it as some local weirdness that may prefigure something is in process and bears further investigation, but not justify hitting the panic button. And, start to issue correctives to the alarmed locals.

Conversely, if the nuclear armed state has weak early warning systems of its own that barely reach outside of its own territorial boundaries using radar, or from floating or flying airborne platforms or local human intelligence reporting via telecommunications or Internet channels, then it may have almost no means of cross-checking the validity and soundness of an alert or social media storm that an attack is underway on someone, and possibly itself.

Obviously, a state of high tension in which nuclear threats already figure large, and in which many other elements of military force are already in position to prosecute a war, including nuclear delivery platforms such as strategic bombers or submarines, may incline a leadership to lend credence to reports of a social media storm announcing that bombers have taken off or that missiles are lifting off from the US mid-west. This inclination might be reinforced if they were simultaneously under heavy cyber-attack. As a result, they might not only put their forces onto alert, but even consider launching a damage-limiting first strike of nuclear weapons.

It is equally obvious that social media events that report that the United States is under attack by North Korea may be perplexing to the DPRK's leadership, and suggest that a deception campaign of

some kind is underway. Even a sarcastic comment by someone like Elon Musk saying that the contrail from a space launch rocket was a "nuclear alien UFO from North Korea"[3]may be of great concern. After all, in the DPRK all major figures in the rocket business work for Kim Jong Un. It possible, even likely, that the DPRK's leadership will mirror-image their system onto the United States and assume that Musk is working for the President, and that his comment is intended by the US government to send them a message. They might be forgiven for such an inference given that his company regularly launches secret US intelligence satellites. Now that President Trump has effectively weaponized his tweets in symbolic combat with the DPRK's Kim Jong Un, North Korean early warning analysts must have intuited that Twitter is a symbolic force field of extraordinary significance. No-one knows what the North Koreans make of this form of psychological warfare or how they report to Kim Jong Un.

If all nuclear armed states had redundant global sensor and communication systems like the United States, one might worry less about the potential that "false positives" from alert systems and social media might trip nuclear early warning systems. Unfortunately, in a world of enormous a-symmetry in situational awareness, nuclear armed states with short-range, weak, and non-redundant early warning systems may rely on social media for their first sign that they are under attack.

Can Social Media Solve this Problem?

Simply setting in motion a requirement for US-based social media that they issue denials that a nuclear attack is underway or act in other ways to shut down a false positive social media storm that portrays an attack is in motion will not suffice to overcome the potential destabilizing effects of social media. States and non-state groups are adept at creating contagious news events, and could create false negative social media storms even as they launch such an attack. It would be natural for a nuclear adversary confronted with early warning advice based on a social media storm to ask the question: why is there a social media storm that no attack is underway? What could possibly prompt such a contagion if it were not an orchestrated campaign by a state or non-state actor using all the means that are already well established such as bots, etc.?

Even worse in terms of perception might be a *contested* social media storm which contains true positives and false positives as to a nuclear attack being underway. Such mixed messages might suggest that some "real" users on the ground are in fact reporting accurately that an attack is underway while the state launching the first strike is conducting a deception campaign to report negatively that there no such attack exists.

Also, "standard" social media strategies to curtail contagious messaging that propagates hate speech, genocide, terrorist recruitment etc. by identifying and shutting down accounts or by supporting potent "counter speech" takes days and weeks to work, whereas nuclear early warning reporting and nuclear decision making takes seconds, minutes, and hours.

In short, social media platforms can't fix this problem by themselves, even if they want to, although they will certainly be a necessary but not sufficient part of the solution to the problem, with specific obligations that need to be spelled out and integrated into their organizational DNA.

Antidotes to False Early Warnings and Social Media

Aside from these difficulties, there are profound issues at stake for nuclear armed states to validate the truth or falsity of nuclear war alerts and/or social media storms. It is early days, but it seems likely that some form of independent, impartial, and pre-existing global early warning system may be useful to all nuclear armed states to provide a supplementary way to check not only the truth content of these alerts and social media storms, but the operation of their own "routine" early

warning systems.

Such a system could fuse a variety of data streams including distributed seismic monitors[4] and streaming video from pre-emplaced systems or “swarming” flash mob teams armed with smart phones to monitor launch sites in real time.

A global fusion center might draw on a variety of indexes mined in real time from big data such as the beta-tested GDELT that tracked the emotional state of world leaders every fifteen minutes.[5] It could include observational data of nuclear forces from real-time sensors fielded by non-nuclear state national technical means,[6] or from sensors on a constellation of micro-satellites launched by non-state actors.

Over time, such a system could be grounded on city-based reporters, trained to observe and report the status of nuclear weapons in their vicinity (they may host installations, be targets, be downwind), prefiguring the increasing role of cities in global governance and security issues.

Who knows, such a “global risk reduction early warning center” might also induce nuclear armed states to increase the operational transparency of their nuclear forces out of their own interest in creating and sustaining strategic stability, instead of nuclear postures based on secrecy, information denial, ambiguity, and opacity.

III. ENDNOTES

[1] Peter Hayes, "Nuclear War and Daily Life", Blue Peter NAPSNet, November 17, 2015,

[2] R. Schulueter, “Twice in St. Louis, radio listeners were told a nuke was on its way’ Answer Man blog, January 23, 2018.

[3] C. Wootson, “A ‘UFO sighting’ briefly freaked out the West Coast. There was an earthly explanation,” *Washington Post*, December 23, 2017

[4] As anticipated by the Citizen Array project, see Technology for Global Security, https://www.tech4gs.org/preventing-nuclear-catastrophe.html

[5] See “New GDELT World Leaders Index,” February 9, 2014, at: https://blog.gdeltproject.org/new-gdelt-world-leaders-index/

[6] A suggestion made by Bruce Blair, in essence, to multilaterarize the US-Russian nuclear risk reduction center concept.

IV. 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/reducing-the-risk-that-social-media-storms--rigger-nuclear-war-issues-and-antidotes/

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

nautilus@nautilus.org