Deterrence in Cyberspace

Ethan Bueno de Mesquita
University of Chicago

“Whereas a missile comes with a return address, a computer virus generally does not.”

William Lynn, Deputy Secretary of Defense

“Casually applying well-known concepts from physical space like deterrence, where attribution is assumed, to cyberspace, where attribution is frequently the problem, is a recipe for failure.”

General Michael Hayden

Building Blocks of a Game Theoretic Model

Multiple potential attackers

- Positive benefits of attacking
- Costs of being retaliated against
- Decide whether to attack based on comparing benefits to expected retaliation

Defender

- Sees lots of information, but may remain uncertain of attribution
- Decides whether to retaliate and against whom
- Correct retaliation is beneficial
- Mistaken retaliation is costly

Deterrence stronger = Attackers requires larger benefit to attack

Four Findings

1. Deterrence in cyber-space is global and interconnected, not bi-lateral.

2. Optimal cyber-deterrence blends aggressive retaliation when attacks are clearly attributable with forbearance when they aren't.

3. Retaliatory efforts should be focused on our most deterrable, rather than most aggressive, adversaries.

4. Technological improvements in attribution will not always improve deterrence.

ATTRIBUTION PROBLEMS MAKE CYBER-DETERRENCE GLOBAL, RATHER THAN BI-LATERAL

Interconnectedness is fundamental in cyber-deterrence

Inputs to attribution

- Specifics of attack
- Generalized strategic environment

Some adversary becomes more aggressive

- Now more suspect following *every* hard to attribute attack
- Other adversaries less suspect

Other adversaries less likely to face retaliation, so they become more aggressive too

If we become worse at deterring one adversary, we become worse at deterring them all

The Washington Post
Democracy Dies in Darkness

Russian spies hacked the Olympics and tried to make it look like North Korea did it, U.S. officials say

The PyeongChang 2018 Winter Olympics opened with a dazzling ceremony Feb. 9. (Pawel Kopczynski/Reuters)

By **Ellen Nakashima**
February 24

Russian military spies hacked several hundred computers used by authorities at the 2018 Winter

WIRED

KIM ZETTER SECURITY 12.23.14 01:52 PM

EXPERTS ARE STILL DIVIDED ON WHETHER NORTH KOREA IS BEHIND SONY ATTACK

Attributing GhostNet

"The most obvious explanation…would be that this set of high profile targets has been exploited by the Chinese state for military and strategic-intelligence purposes…

However, we must be cautious to rush to judgement…

[T]his network of infected computers could have been targeted by a state other than China, but operated physically within China for strategic purposes… perhaps in an effort to deliberately mislead observers as to the true operator(s) and purpose of the GhostNet system."

Information Warfare Monitor

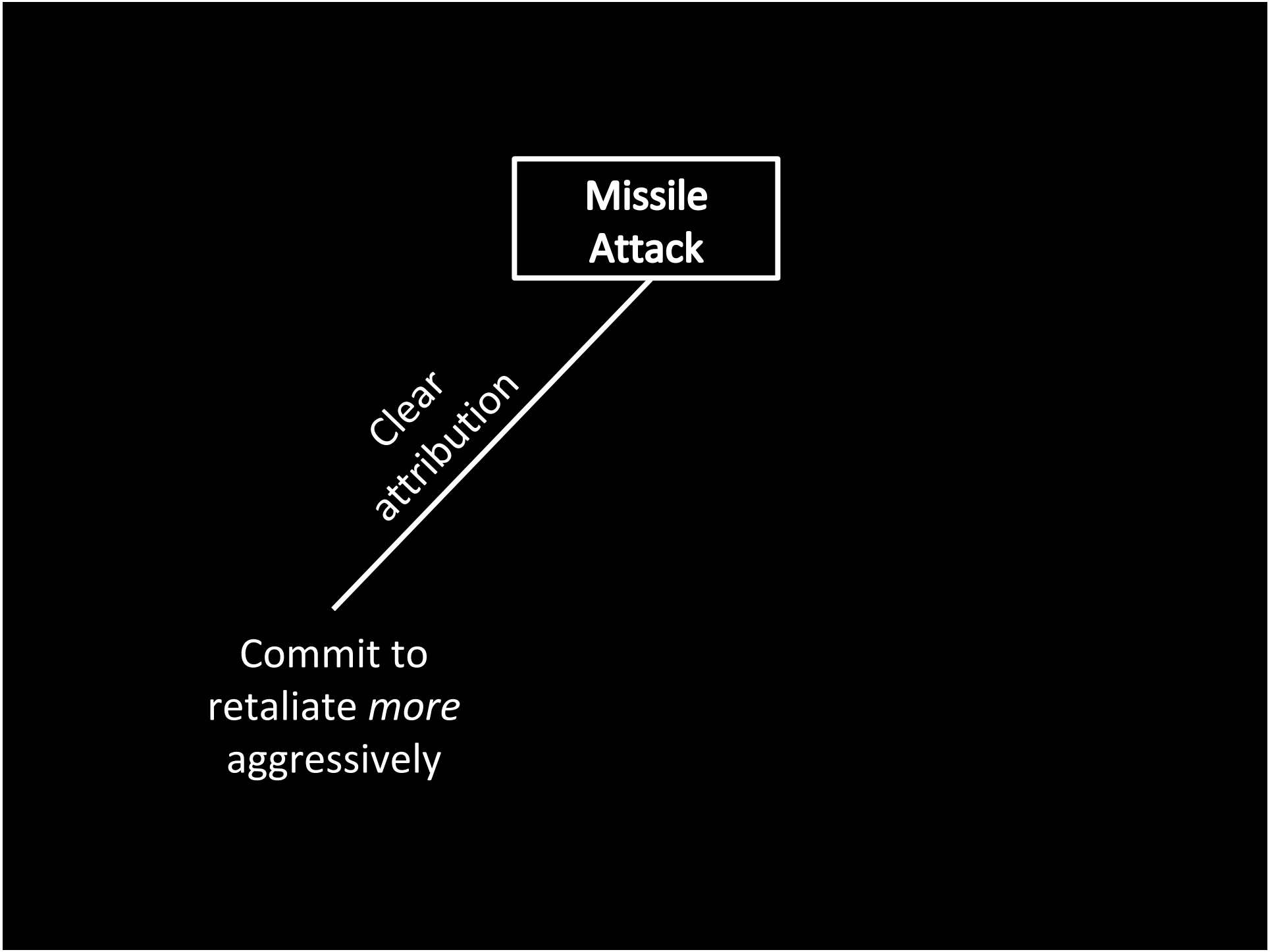
TOWARDS A CYBER-DETERRENCE DOCTRINE

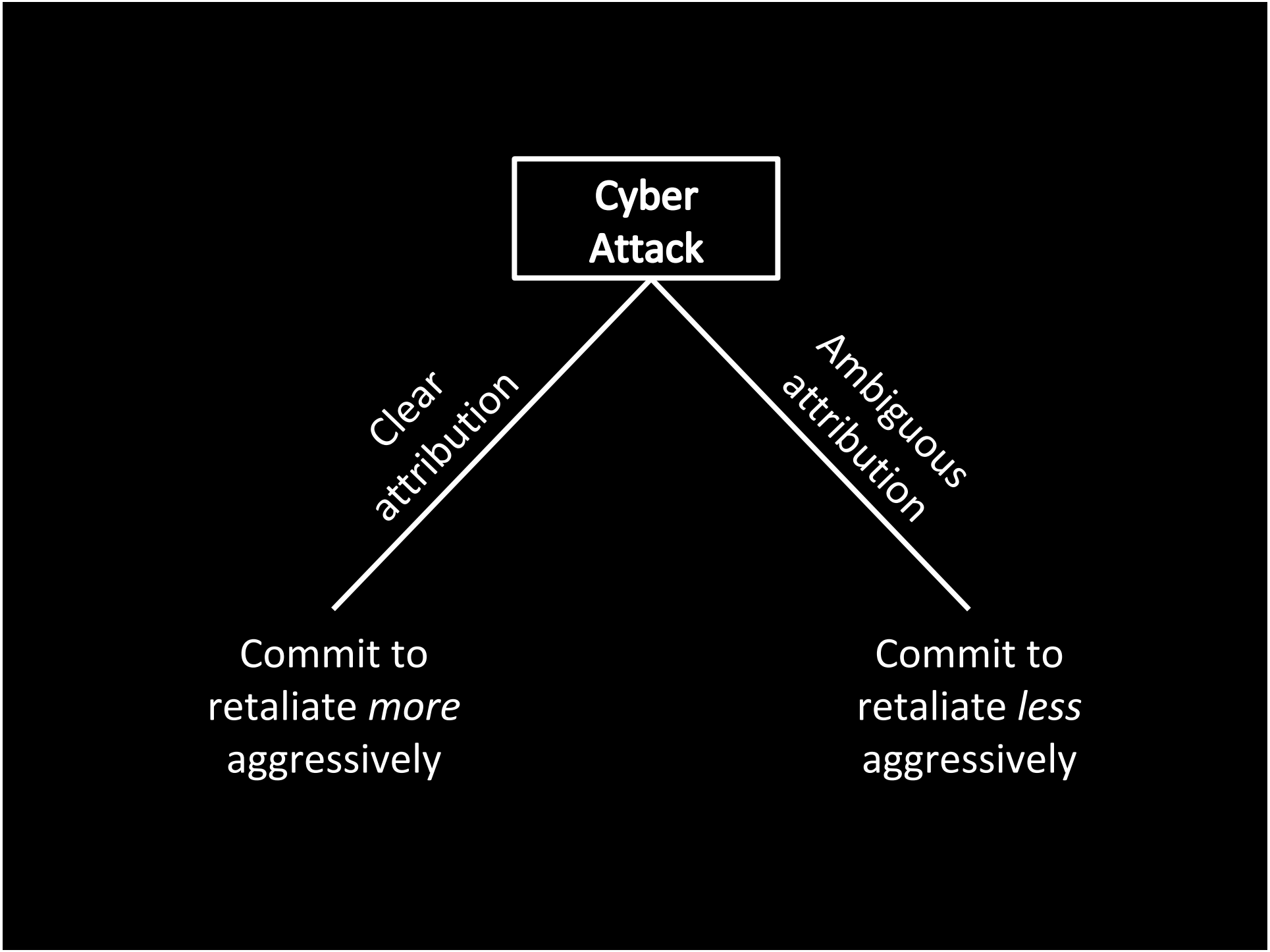
Commitment and Deterrence

Commitment problems make doctrine particularly important in deterrence

Standard deterrence theory says commit to increased aggressiveness across-the-board

Recent policy suggestions call for applying same idea to cyber, but our analysis disagrees


Missile Attack
Clear attribution
Commit to retaliate *more* aggressively


Cyber Attack
Clear attribution
Ambiguous attribution
Commit to retaliate *more* aggressively
Commit to retaliate *less* aggressively

Who should be the focus on deterrence?

"Our focus will be on the States that can pose strategic threats to U.S. prosperity and security, particularly China and Russia."

Department of Defense Cyber Strategy 2018

Focus on the most *deterrable* adversaries

Cyber-deterrence is not bi-lateral

An adversary is deterrable if:

- Detectable
- Responsive to incentives

We maximize the efficacy of deterrence by focusing on most deterrable, not most aggressive adversaries

IMPROVING ATTRIBUTION CAN HAVE UNINTENDED CONSEQUENCES

DOD 2015 Cyber Strategy

"Attribution is a fundamental part of an effective cyber deterrence strategy…DoD and the intelligence community have invested significantly in all source collection, analysis, and dissemination capabilities, all of which reduce the anonymity of state and non-state actor activity in cyberspace…"

3 components of the attribution problem

False alarms

- Buckshot Yankee
- 2018 DNC hack

Detection Failure

- Stuxnet

Mis-identification

- Solar Sunrise
- Guccifer 2.0

Improve information in ways that increase confidence about retaliation

Reduce false alarms

- Any detected attack is more likely to be real

Improve detection and identification, simultaneously

- Detect easy to attribute attacks that were previously not detected

Better detection without better identification can backfire

Suppose we start detecting more attacks that are hard to attribute to specific adversaries

Can increase reticence to retaliate following attack detection due to concerns about misidentification

This can make our adversaries more, rather than less, aggressive

Chasing too much certainty

Discover a marker always pointing to one adversary

- Perfect attribution when this marker is present

If marker absent, *less* certain of attribution to that adversary than before knowing about the marker

Can lead us not to retaliate following attacks we previously would have retaliated against

Weakens deterrence

Four Implications

1. Deterrence in cyber-space is global and interconnected, not bi-lateral.

2. Optimal cyber-deterrence blends aggressive retaliation when attacks are clearly attributable with forbearance when they aren't.

3. Retaliatory efforts should be focused on our most deterrable, rather than most aggressive, adversaries.

4. Technological improvements in attribution will not always improve deterrence.

Future Questions?

How does strategy change when weapons can only be used once?

How does the presence of cyber weapons interact with or disrupt traditional national security strategy?

What are the implications of offensive still being the domain of government, but defense being diffused across the private sector?

Technical paper

"Deterrence with Imperfect Attribution"

Sandeep Baliga (Northwestern)

Ethan Bueno de Mesquita (University of Chicago)

Alexander Wolitzky (MIT)

https://home.uchicago.edu/~bdm/PDF/deterrence.pdf