North Korea and the Sony Hack: Exporting Instability Through Cyberspace

Stephan Haggard
Jon R. Lindsay

SUMMARY

The North Korean cyber attack against Sony Pictures Entertainment in connection with the planned release of The Interview raised important questions about the feasibility of deterrence in cyberspace, the protection of First Amendment values, and the responsibility of the US government to safeguard private networks. It also resulted in the unprecedented attribution of responsibility for a cyber attack to a nation state by a US president, despite public controversy over the evidence. North Korea has long engaged in provocative behavior on the Korean peninsula, recently including cyber attacks, but the probability of general war with South Korea remains quite low. Strategists describe this problem as the stability-instability paradox. North Korean coercion targeting a corporation on US soil in effect extends this dynamic into global cyberspace. It is impossible to deter all forms of cyber harassment, but policymakers can manipulate the threshold of ambiguity that makes limited aggression more or less attractive.
In late 2014, Sony Pictures Entertainment (SPE) experienced a barrage of network intrusions and disruptions, accompanied by extortionate threats. The hack ultimately failed to prevent people from seeing *The Interview*, a satire about a plot to assassinate North Korea’s dictator, Kim Jong Un. On the contrary, it created an advertising bonanza for the poorly reviewed movie. The hack embarrassed executives, but it caused only minor financial losses for SPE.¹

Yet the reaction of the US government was unprecedented, resulting in the first-ever attribution of a cyber attack to a nation state by a US president as well as modest retaliatory measures. The hack also raised important questions about foreign obstruction of First Amendment freedoms and the responsibility of the government to protect corporate networks.

Most analysis of the Sony hack has focused on the technical dimensions of the attack including North Korea’s cyber capabilities and the evidence for—and against— attribution.² There has been relatively less focus on the North Korean motivation for the attack and its implications for other asymmetric conflicts. Far from idiosyncratic, the Sony hack exemplifies the problems the United States and its allies face in responding to provocations that fall below the usual threshold for effective deterrence. Strategists describe this problem as the stability-instability paradox.

The low probability of general war on the Korean peninsula has long encouraged provocative behavior by North Korea. The Sony hack in effect extends this paradox into global cyberspace. We should thus expect that online harassment like the Sony hack will be limited in its coercive potential but nonetheless both potentially costly and difficult to deter. Policymakers might limit the likelihood or magnitude of such provocations by reducing the ambiguity of deterrence, but such attacks cannot be eliminated altogether.

### Hacking Sony and Setting Precedent

On November 24, 2014, a menacing neon skeleton from the “Guardians of Peace” (GOP) appeared on SPE computer screens. The attack also wiped several hard drives clean.³ Some days earlier, executives had received a ransom note from “God’s Apostles,” albeit with a pecuniary rather than political demand and no apparent link to North Korea.

Several days later, the attackers started to release private Sony data onto the Internet (“doxing” in hacker argot). Using file-sharing hubs, the hackers dumped a number of Sony films, internal emails from senior executives, financial records, film contracts, personal information on celebrity actors, and employee health records. These releases did not come all at once, but in tranches that kept the story alive for weeks.

On December 5, an ominous note threatened physical harm to Sony employees if they did not sign a denunciation of the firm. The note appeared to come from the GOP, but was subsequently renounced by it. On December 8, the GOP made explicit demands that SPE not release *The Interview*. This note was the first link to the parodic film, which lampoons and ultimately kills North Korean leader Kim Jong Un, setting the stage for a democratic revolution in the country. On December 16, the GOP made veiled threats to attack theaters where *The Interview* was scheduled to open, with a warning to “Remember the 11th of September 2001.” These developments substantially widened the debate over the Sony hack into an issue of both counterterrorism and free speech.

The FBI claimed no credible evidence of terrorist threats linked to the release of the film. Nevertheless, major theater chains refused to show the film, citing concerns over the safety of patrons. Sony decided on December 17 to suspend the release. For the first time (aside from small-scale ransomware attacks against individual users) it appeared that an act of coercion through cyberspace had compelled a major firm to alter its business plans.

Investigators at Sony and in the US government almost immediately suspected North Korea. They identified striking similarities to malware used in previous suspected North Korean attacks on US and South Korean targets, as well as other compromising forensic evidence. North Korea’s motivation for punishing Sony or preempting the
release of *The Interview* also seemed obvious. In July, Pyongyang had asked Washington to block the release of the film and had even lodged a protest to the Secretary-General of the United Nations, Ban Ki-moon. Not surprisingly, the government of North Korea denied any involvement in the hack.

On December 19, an FBI press release asserted that “North Korea’s attack on SPE reaffirms that cyber threats pose one of the gravest national security dangers to the United States.” The same day, at his year-end press conference, President Obama doubled down on the FBI assessment by arguing that SPE made a mistake in caving in to the attack and that the United States would respond proportionally at “a place and time and manner that we choose.” The willingness of senior US officials to confidently blame a nation state for a particular cyber attack was unprecedented. The May 2014 Department of Justice indictment of five Chinese military officers for economic espionage had focused on individual culpability and stopped short of blaming the Chinese government. Official statements were met with widespread skepticism in the cybersecurity community. Experts challenged the reliability of technical evidence cited by the FBI, questioned North Korean technical competence, and described alternative theories such as a “false flag” operation impersonating North Korea or even cooperation from disgruntled employees.

Nonetheless, President Obama signed an executive order on January 2 explicitly citing the Sony hack as one motive for new sanctions against three North Korean organizations and ten government and party officials. This action was largely symbolic given the blanket of sanctions the United States already had thrown over North Korea, but was nonetheless unprecedented. There has also been speculation that the United States may have been responsible for a large-scale Internet outage in North Korea in late December, although it might well have been the work of private hacktivists.

Whether coincidentally or not, the end of the hack corresponded closely with the FBI’s public statement and the president’s press conference. On December 18, the GOP effectively stood down. Despite the fact that the Sony decision was subsequently reversed and *The Interview* moved into release both in theaters and online venues, neither the threats from the GOP nor the release of sensitive information resumed. It is possible that the GOP accomplished all that they intended at that point, but the specter of US government retaliation may also have spooked the attackers.

The Sony hack is one of only a few instances in history of an attempt by a nation state to use cyberspace for explicitly coercive purposes. The Sony hack was also notable because the US government vigorously and publicly rallied to the defense of a private firm targeted for such coercion. Previous corporate intrusions by foreign actors have typically been handled through criminal investigations rather than state retaliation.

However, in one important respect, the hack was familiar: it resembled a style of asymmetric conflict that North Korea has undertaken vis-à-vis South Korea and the United States for decades, and particularly since important political developments on the peninsula in 2008.

**The Stability-Instability Paradox on the Korean Peninsula**

Despite the appearance of persistent instability on the Korean peninsula, the risks of large-scale conflict are relatively slight. The United States and the Republic of Korea enjoy overwhelming conventional advantages that have only been partly offset by North Korea’s acquisition of nuclear weapons. At the same time, the proximity of Seoul to North Korean artillery effectively rules out any attack on North Korea except in the unlikely case of extraordinary provocations or the onset of war.

As theorists have long noted, stable deterrence generates what is known as the stability-instability paradox. During the Cold War, the risk of nuclear Armageddon—however slight—deterred both nuclear war and large-scale conventional conflict in Europe. However, threats of mutual suicide were not credible for deterring minor conventional aggression or

---

*The Sony hack is one of the few instances in history of an attempt by a nation state to use cyberspace for coercion*
covert assistance to guerrillas in peripheral countries. The use of threats to attenuate the risks of general war, ironically, can incentivize lower-level aggression.

Such provocations have long been a component of the North Korean arsenal. In Jang Jin-sung’s fascinating book *Dear Leader*, the defector describes his time working in North Korea’s United Front Department during the Kim Dae Jung presidency. The country was still recovering from the famine, and South Korea’s Sunshine Policy held out hope of material support. However, Pyongyang did not want to make any concessions to Seoul. According to Jang, its solution was the so-called “Northern Limit Line (NLL) strategy.” Instead of promising anything tangible, North Korea would launch provocations—such as the naval skirmishes of 1999 and 2002—and then promise to stop them in exchange for talks and aid.

Although North Korea has continually tested the limits of US and South Korean deterrence, these provocations increased in frequency and intensity following the inauguration of conservative president Lee Myung Bak in 2008 and the collapse of the Six Party Talks later in that year. The North sought to demonstrate that South Korea’s reversal of the Sunshine Policy (and the aid that went with it) was doomed to failure. With respect to the United States, North Korea also sought to show that sanctions and other diplomatic efforts to pressure the country were unlikely to work.

In the six years since the collapse of the talks, North Korea undertook its second and third nuclear tests (May 2009 and February 2013), two “satellite” launches that were in effect tests of a long-range missile, and countless short- and medium-range missile tests. In addition, it sank a South Korean corvette, the *Cheonan*, with the loss of 46 lives in March 2010 and shelled the island of Yeonpyeong in November of the same year. These provocations were carefully calculated to inflict small costs and generate some risk of escalation, but ultimately avoided crossing thresholds that would trigger retaliation. In response, South Korea and the United States were repeatedly challenged to adjust response thresholds and to maintain and toughen the credibility of deterrence.

### Same Logic, Different Domain

Cyberspace provides yet another domain for the stability-instability paradox to play out. South Korea is one of the most wired countries in the world and its highly open society complicates cyber defense. South Korean cyberspace is a target-rich environment for an asymmetric attacker. Because no one gets hurt or killed through distributed denial of service (DDoS) attacks or espionage intrusions, and because digital damage can often be mitigated, North Korean cyber harassment takes place well below the threshold for retaliation.

North Korea, by contrast, is not only closed and poor, but has relatively limited exposure to cyber risks. The country has only 1,024 IP addresses and extraordinarily limited broadband access. In addition, North Korea has—until recently at least—benefited from the attribution problem. Hackers use indirection and deception within large-scale computing infrastructure to hide their identity. In the case of North Korea they may also draw—as they do with respect to other illicit activities—on networks of sympathetic non-state actors. Plausible deniability makes it difficult to generate support for retaliatory actions, as the Sony hack amply illustrates.

We cannot attribute North Korean cyber attacks solely by the fact that the state has strong motives. But “means, motive, and opportunity” are important considerations for narrowing down the pool of suspects in any mystery.

What we know about North Korean cyber operations—at least in the public domain—is gleaned largely from the accounts of two North Korean defectors who had been assigned to cyber units and public statements by South Korean and US defense and intelligence officials. Although North Korea did not make its first known direct connection to the Internet until 2010, the country began investing in cyber capabilities in the late 1980s, almost certainly with Chinese support. Talented students were identified as early as middle school and channeled through newly created training programs. Entire universities were devoted to cyber training. Graduates were subsequently...
deployed into a number of operational units under several main chains of command, each of which ended in Kim Jong Il and Kim Jong Un: the Korean People’s Army General Staff, the Korean Workers Party, and the General Reconnaissance Bureau.

Not coincidentally, the General Reconnaissance Bureau is a unit formed in 2009 and believed responsible for other provocations associated with the NLL strategy. Its cyber units (Units 91 and 121) have responsibility for psychological warfare, propaganda, and exploitation and appear to have engaged in DDoS and exploitation attacks against US and South Korean targets. Unit 121 is strongly suspected of having a presence in China. North Korea may employ as many as 1,400 cyber operators, more than double the number in South Korea. Total overseas staff devoted to cyber operations probably number in the hundreds.

The observed pattern of attacks on South Korea and US interests on the peninsula suggests that they are tied to broader developments on the peninsula. Moreover, the costs inflicted by these actions are rising over time and attackers are clearly testing limits and gauging responses.

In 2004 and again in 2006, North Korea appeared to breach a number of South Korean military wireless communication networks. Following the collapse of the Six Party Talks in 2008 and the inauguration of Barack Obama, however, cyber activities appear increasingly linked to broader political dynamics on the peninsula. In a pattern seen in 2006, the country undertook missile tests in early 2009 that were condemned at the UN Security Council, after which the country responded with another nuclear test on May 25. The nuclear test was followed by the passage of additional multilateral sanctions. North Korea responded with threats of a “hi-tech” war, and DDoS attacks occurred and disk-wiping malware was distributed on July 4. The incident targeted South Korean and US government entities, media outlets, and financial websites.

In March 2011, coinciding with annual joint US-ROK exercises, South Korean media, financial, and critical infrastructure targets again suffered a DDoS and disk-wiping malware attack dubbed “10 Days of Rain” by McAfee. The hackers also targeted US and South Korean military entities and jammed the GPS systems of hundreds of civilian aircraft and ships.

In May 2013, “DarkSeoul” hackers attacked several South Korean financial institutions and DDoS attacks were launched against the South Korean government’s Domain Name System (DNS) registry in June. According to Symantec the 2009, 2011, and 2013 attacks were technically similar. In December 2014, following the revelation of the Sony hack, a South Korean power and nuclear operator was also attacked. Although less damaging than prior attacks, the target choice suggested an interest in critical infrastructure.

In each of these cases, the capacity of the South and the United States to respond was hampered by the attribution problem. As in the SPE case, attacks were undertaken by a succession of changing front groups, such as “WhoIs Team” and “New Romantic Cyber Army Team.” However, forensic analysis of attacker missteps, traffic patterns, and code artifacts suggests convincingly that these nominally disparate attacks probably emanated from North Korean state organs. Alternative theories of attribution are far less credible.

Exporting Cyber Instability to the United States

The Sony hack thus appears to be the culmination of long involvement in cyber actions rather than a maiden voyage. The difference in the Sony hack is that North Korea targeted an entity on North American soil—albeit of Japanese parentage—and enjoyed a moment of anonymity and a modicum of success in compelling a policy change by the target (SPE). On the surface, this willingness to directly challenge the United States by hacking the homeland appears to be a provocative escalation from a history of disruptive cyber attacks confined mainly to South Korean targets (although espionage targets have been more wide ranging). However, many of the constraints that operate on the Korean peninsula are also visible in an attack on the United States as well.

As the skeptical reaction of the cybersecurity community to early FBI statements demonstrated,
North Korea initially benefited from the attribution problem. North Korea exploited ambiguity in the willingness and ability of Sony and the US government to respond to cyber threats. Private firms like Target, JP Morgan Chase, Anthem, and even Sony itself had experienced major intrusions before, losing millions of customer records and valuable financial data. The US government has not mobilized a major response in any of these cases beyond standard criminal investigations. North Korea must also have noted that China has been able to infiltrate hundreds of American corporate and government targets without paying much of a price.

Nonetheless, attacking a US firm was not without risk. Why might North Korea be willing to run the risks of inflicting harm on the US homeland even within the relative buffer zone of a private corporate network? It is worth speculating on what some of North Korea’s motives may have been as they touch on some of the reasons other authoritarian regimes may engage in such behavior.

The North Korean political system is extremely personalist in form, and rests on the virtual deification of the leader. Slights to the Kim dynasty—let alone attacks on it—are treated as among the most heinous of crimes, even when inadvertent, and have spawned strong diplomatic actions when taken internationally. Most recently, the North Korean regime launched a particularly focused and wide-ranging diplomatic campaign to deflect calls for personal accountability against Kim Jong Un for the country’s likely commission of crimes against humanity. The build-up to the release of The Interview coincided with the run up to crucial UN votes on a widely read Commission of Inquiry report on human rights abuses in the country. The UN report implicated the top leadership, including Kim Jong Un. Pressure on human rights and the release of the film appear to have been seen by Pyongyang as part of a broader pattern.

The case also fits within a variety of related concerns that North Korea and other authoritarian regimes might have about information flows. Despite the fact that the regime invests heavily in propaganda and has managed to control the flow of information, it is not immune from transborder leakage. This is particularly the case as the country becomes more integrated with and dependent on China and as trade expands in sensitive cultural products, such as movies and music from South Korea. There is plenty of evidence that authoritarian governments use cyber tools to silence critics and dissidents abroad. North Korean cyber operators training with the Chinese would surely have been exposed to these techniques.

The stability-instability paradox makes ambiguity-exploiting harassment like the Sony hack possible, and perhaps inevitable. But North Korea crossed three lines that made some sort of US response all but inevitable.

First, in threatening consequences against a private firm if it released a film, North Korea elevated a technical intrusion into a First Amendment issue. Prior corporate cyber attacks imposed financial costs and embarrassed executives, but the normative assault on American liberal values by a foreign dictatorship raised the stakes and created public pressure for the US government to act.

Second, the attackers threatened not only harm in cyberspace but also physical harm to the families of Sony employees and, however improbably, violence against theaters daring to show The Interview. The US government clearly takes terrorist threats very seriously. Note that the US response to the hack—financial sanctions—also occurred outside of the cyber domain. The case suggests that cyber threats become most salient when coupled to actions in other domains on the part of the perpetrator.

Finally, North Korea appears to have underestimated the capacity of the United States to successfully attribute the attack to Pyongyang, just as it miscalculated with respect to South Korea in the sinking of the Cheonan. In this case, the attackers lost their anonymity through poor operational security. Moreover, some of the Snowden leaks provide evidence of US technical capabilities for attribution and exploitation (e.g., the NSA’s Office of Tailored Access Operations). One memo leaked by Snowden
details how the NSA gained information on North Korean computer network operations initially by monitoring South Korean espionage efforts, but subsequently by tracking North Korea directly.12

Successful attribution does not in and of itself constitute a deterrent. But it raises the stakes—including the political stakes or so-called audience costs—of any attack. In a public talk in February 2015, NSA Director Admiral Michael Rogers discussed some of the reasoning behind the US response.13 Because the attack had become a matter of such wide public knowledge, doing nothing would encourage others to go down the same road. Rogers said, “we’ve got to publicly acknowledge it, we’ve got to publicly attribute it, and then we’ve got to talk about what we’re going to do to impose cost.”

One lesson for cybersecurity is that ambiguity both enables and constrains coercion. Complex sociotechnical infrastructure and public-private coordination problems undermine reliable cyber defense and thus deterrence by denial. It follows that deterrence by detection—improved capacity for attribution coupled with the threat of punishment—is a central aspect of cyber defense that requires substantial public investment.

Contrary to popular belief, cyber deterrence appears to be getting easier not harder. The scope and sophistication of Internet surveillance and logging in cloud-based data stores is improving. The operating environment for cyber attackers is thus becoming less permissive as government and corporate defenders employ more sophisticated counterintelligence practices. The 2015 Department of Defense Cyber Strategy is notably more sanguine about the feasibility of deterrence than the 2011 document it replaced.

Unfortunately, the deterrent benefit of improved attribution is lost if potential attackers do not appreciate it. The United States can improve the capacity to attribute through technical innovation, information sharing with private firms (including anonymized access to cloud data), and intelligence partnership programs with allies. The United States must also do a better job at publicizing these efforts and regularly highlighting attribution successes.

In the end, deterrence requires a credible retaliatory response. There is little reason to believe that additional sanctions and symbolic denunciation impose meaningful costs on a regime like North Korea, especially when it perceives near-sacred values under assault. Carefully targeted counterintelligence retaliation against the North Korean cyber attack infrastructure has some promise, although actions taken for deterrence must often compromise future collection opportunities. As in so many other areas, Chinese pressure might persuade North Korea to moderate its activity. But China must first be persuaded itself, a highly unlikely outcome given the current state of US-China cyber relations. There is little evidence to date that China is willing to pressure North Korea to adopt a more forthcoming foreign policy.

Ultimately the stability-instability paradox is real. The likelihood and magnitude of aggression depends on the probability of attribution, the costliness of punishment, and the balance of resolve on both sides, and these will vary for different forms of provocation. It is impossible to prevent all forms of cyber harassment, but the threshold of ambiguity that makes it possible can be manipulated through:

- Investment in attribution, counterintelligence, and information-sharing capabilities and advertising their effectiveness;
- Clarifying the protections the government extends to private firms and the level of risk firms should have to assume, in order to avoid moral hazard problems;
- International dialogue to establish norms of acceptable and unacceptable behavior in cyberspace;
- Considering retaliatory measures beyond the cyber domain that might be more credible or effective;
- Not exaggerating the harm of harassment that fails to meet the threshold for response.

Contrary to popular belief, cyber deterrence appears to be getting easier not harder.
References

3 There are numerous published timelines of the Sony hack, but one of the most comprehensive in its range of linked sources is maintained by Risk Based Security, “A Breakdown and Analysis of the December 2014 Sony Hack” at https://www.riskbasedsecurity.com/2014/12/a-breakdown-and-analysis-of-the-december-2014-sony-hack/.
8 Estonia was hit by service denial attacks in 2007 protesting the removal of a Soviet-era statue; Russian state involvement is suspected. China uses service denial attacks as part of its information control strategy, but usually only targets dissidents and minorities.

About this Publication

The AsiaPacific Issues series reports on topics of regional concern.

Series Editor: Elisa W. Johnston

Copies of this paper may be downloaded from the Center’s website. For information about the series, please see the Center’s website or contact:

Publication Sales Office
East-West Center
1601 East-West Road
Honolulu, Hawai‘i 96848-1601
Tel: 808.944.7145
Fax: 808.944.7376
EWCBooks@EastWestCenter.org
EastWestCenter.org/AsiaPacificIssues
ISSN: 1522-0966

© 2015 East-West Center

Recent AsiaPacific Issues


No. 115 "Broadcasting Justice: Media Outreach at the Khmer Rouge Trials" by Christoph Sperfeldt. July 2014.


No. 113 “See No Evil: South Korean Labor Practices in North Korea” by Marcus Noland. April 2014.

No. 112 “Democratic Change and Forest Governance in the Asia Pacific: Implications for Myanmar” by Stephen McCarthy. February 2014.

About the Authors

Stephan Haggard is the Lawrence and Sallye Krause Distinguished Professor of Korea-Pacific Studies at the School of Global Policy and Strategy, University of California, San Diego. He was a POSCO Fellow at the East-West Center in early 2015. He can be reached at: shaggard@ucsd.edu

Jon R. Lindsay is Assistant Professor of Digital Media and Global Affairs at the University of Toronto, Munk School of Global Affairs (as of July ’1)

He can be reached at: jlindsay@ucsd.edu