The Behavioral Intelligence Paradigm Shift in Fighting Cyber-Crime: Counter-Measures, Innovation, and Regulation Issues

Phillipe Baumard

This paper investigates the technological evolution of cyber-crimes from its emergence in the early 1980s to its latest developments in 2013. From this evolution, we draw implications for doctrines, policy, innovation, incentives, and roadmaps as we propose the emergence of a new "behavioral intelligence" paradigm, both in the attack and defense arenas.

Cyber-crime refers to the unlawful use of numeric, electronic, and software capabilities to misuse, temper, devoid, destruct, or influence public or private information systems. Cybernetic and informational components may not be the primary target or final outcomes of cyber-crime campaigns.

The origins of cyber-crime are concomitant with the pioneering efforts of technology enthusiasts in exploring the possibilities offered by technological innovation. Exploration and autonomous appropriation are still, to date, a core motivation in the creation of "hacks". John Draper was one of these computer enthusiasts who helped popularize the first "phreaking" hack, consisting of a multi-frequency tone generator, later known as the Blue Box to pitch the exact 2600 Hz frequency to hack into the long distance phone system of AT&T in the early 1970s.

Most of the early attacks were spontaneous, motivated by technology exploration, non-directed (without a specific target in mind), and immediate in their effects. With the rise of personal computers, these early pioneers of hacking started to group in spontaneous associations, espousing discourses of the times on individual freedom, resistance to authority, and amusement with detours of emerging technologies. Phreaking and hacking became both shared practices that cemented long friendships between developers, industry pioneers (Wozniak, Jobs, etc.), and politically motivated technology enthusiasts. The borders between an emerging underground culture (yippies and hackers) and a criminal sub-culture were blurry and unstable, with very little self-regulation, and comprising teenagers, advanced computer developers, and self-taught technology explorers. We call this era the "code breaking years", where talented individuals are mostly motivated by symbolic and small gains, a feeling of belonging to a new community and self-identity.

However, in the mid-1980s, technical bulletin boards from hackers' groups started to disclose attack guidelines for intrusions, sometimes both physical and code-based (such as the first issue of the Legion of Doom LOD/H Technical Journal, on Jan. 1, 1987). LOD and MOD (Masters of Deception) hence became influential in transforming these early movements into more organized "cracking" communities, moving a step away from the original hacking culture (see Figure 1).

[A] Professor, *Ecole Polytechnique*, Chair Innovation and Regulation of Numerical Services

[1] http://www.textfiles.com/magazines/LOD/lod-1

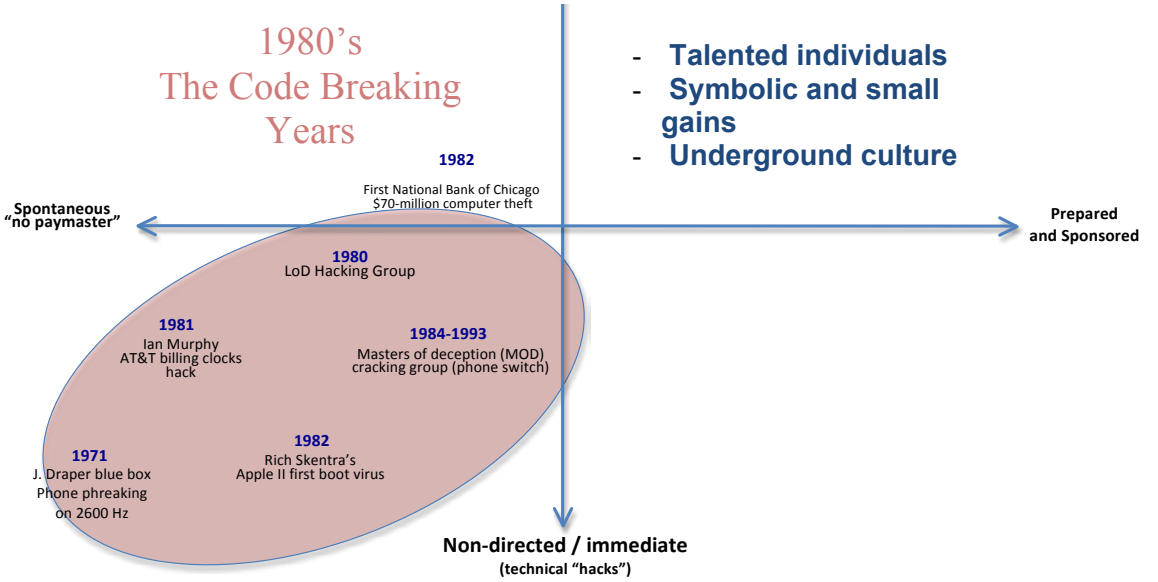

Figure 1. The early years: the code-breaking paradigm

The Cold War and the underground battle for a free Berlin played a determinant role in the evolution of the hacking culture of the late 1980s. The Clifford Stoll episode (an LBL astronomer who accidentally discovered a computer intrusion from West Germany in his laboratory) was the first case to raise the importance of agency coordination and the difficulties of attribution in international computer attacks (Stoll 1989). This case is also one of the early symptoms (1986) of yet to come advanced persistent threats, highlighting the complexity and sophistication of intrusion campaigns (for details see Stoll's article, 1988[2]).

The early 1990s are hence concomitant with the emergence of the criminal sub-culture of hacking. In the 1980s, cracking events that led to theft or large-scale attacks were rare. Two notable exceptions are the 1986 Pak Brain logic bomb, known as the first virus, and the 1982 First National Bank of Chicago computer theft ($70 M USD). The "Great Hacker War" (conflict between Masters of Deception and Legion of Doom, circa 1991–1992) is an example—today disputed as an exaggeration of trivial confrontations—of the interpersonal dynamics of the early 1990s. A blend of prestige seeking, bravados, and playfulness were the core incentives of these early confrontations[3]. The publication of exploits by hackers' groups triggered, however, the interest of Law enforcement. Operation Sundevil, in 1990, was hence the first large-scale cyber-enforcement operation, involving 15 U.S. cities and leading to three arrests[4]. Most cyber-crimes involved wire-tapping, calling card fraud, and credit card fraud. The relative failure of this operation led to an increased awareness of the central role of cyber-deterrence for federal agencies (Sterling 1994).

Publications such as 2600 and the rise of the cyber-space participate in a democratization of cracking, phreaking, and hacking techniques, which render them more versatile to their use "beyond technology". Focus on distant control, resident threats (democratization of Trojans) creates both a more organized criminal sub-culture, and the birth of a societal reach for the attacks (see Figure 2).

While attack preparation is targeted to single point of aggression, the early 2000s is adopting a whole new dynamic. The rise of electronic commerce means a better monetization of cyber-crime with an expectation of large-scale profits for organized crime. The digitalization of the cultural industry (MP3s) creates an appeal for the popular growth of cracking. Profiles of hackers accordingly change in two directions: on the one hand, amateur crackers (script kiddies and mass market consumers) start to use without advanced knowledge available tools (P2P file sharing and cracking "CDs"). On the other hand, malware production becomes a profitable black market. Corruption of DNS paths, denial-of-service attacks, defacing campaigns, and corporate thefts find a rapid monetization. The years 2000–2002 are among the most active in malware generation with viruses such as ILOVEYOU, Klez.h., Code Red, etc. The group Anonymous is created in 2003 as a loosely coupled and spontaneous coordination of various interests, ranging from militant activism, cracking

[2] http://pdf.textfiles.com/academics/wilyhacker.pdf

[3] http://www.textfiles.com/hacking/modbook4.txt

[4] Anthony Lawrence Clapes, *Softwars: The Legal Battles for Control of the Global Software Industry*. (Westport, CT: Quorum Books, 1993).

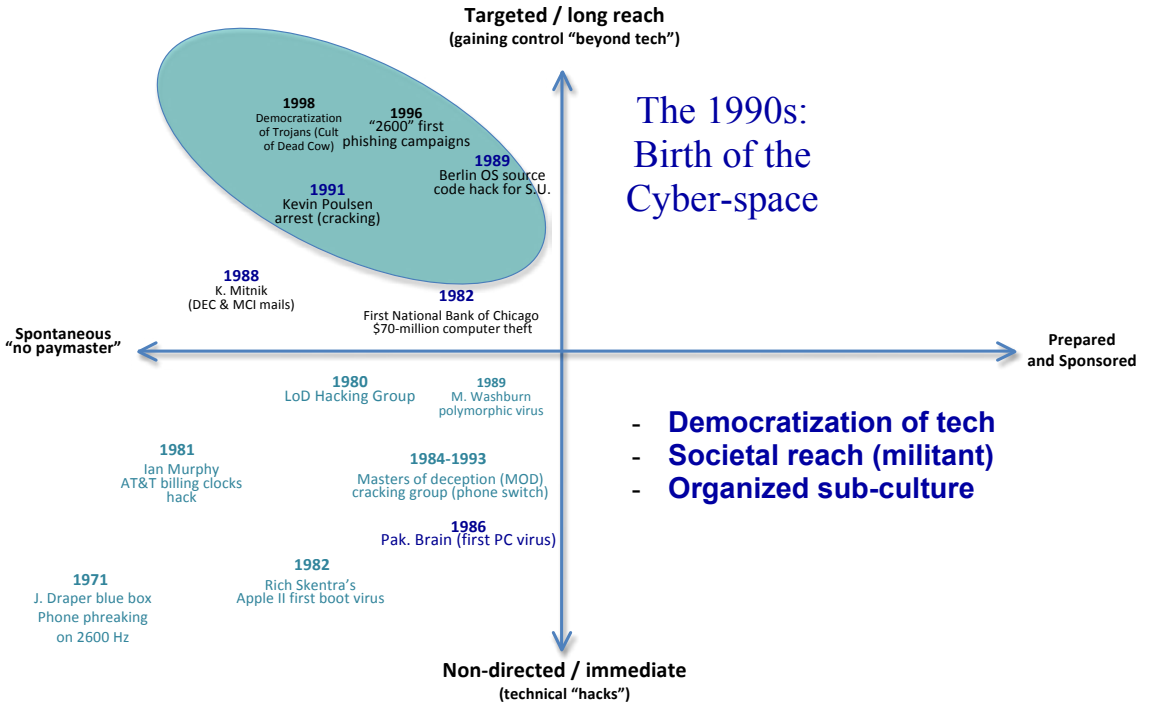

Figure 2. The 1990s: the democratization of cyber-crime

techniques sharing, and image sharing around the 4chan platform. Massive raids and pranks, known as "4chan raids", popularize a perspective of hacking as a blend of activism, bullying, and satirist information campaigns, although opting out of political campaigns in the early years (2003–2006).

Meanwhile, preparation and sponsorship of large-scale attacks also gain considerable traction as the core philosophy of hacking (based on freedom and activism values) is fading away with the diffusion of embedded cracking tools and libraries. Titan Rain (2003–2006) is an exemplar of these first explorations of cyber-warfare involving low-tech methodologies embedded into advanced campaigns (see Figure 3).

The years 2005–2013 are marked by a double shift, and to some extent a seizure, between "target and sponsored campaigns" led by States or organized crime, and more pervasive "spontaneous and long-reach campaigns" led by activist groups, hackers' collectives, and loosely coupled entities such as Anonymous and LulzSec. This period is characterized by a rapid growth of strategic and politically motivated attacks (Kerem125 against the United Nations, Chinese APT1 global campaign, Estonia DoS attacks, Stuxnet, and Operation Aurora) (Figure 4).

The technology used in these large-scale campaigns does not dramatically differ from the early days of hacking. One hundred and twenty-five lines of codes are still very efficient in 2013 to conduct the exploitation of vulnerabilities, even when the lines of defense have exponentially grown in the past 25 years. As most innovation disruptions in the early twenty-first century, the performance of these campaigns is rooted in the accessibility and diffusion of combinatory learning, i.e., the capacity of outpacing the defensive learning of targets by a better and faster behavioral intelligence.

The formation of two distinctive groups (large-scale spontaneous groups versus sponsored targeted large-scale campaigns) is typical of the two paths that can be used to attain a superior collective behavioral learning advantage. Large spontaneous groups benefit from distributed astute learning, i.e., the learning conducted by individual hackers who can coordinate on a very large scale, making their collective learning ubiquitous and efficient. Targeted sponsored campaigns (such as APTs) benefit from the advance of automated artificial intelligence embedded into technology (e.g., Stuxnet and FLAME).

Most defensive systems are based on the recognition of signatures ("embedded malicious codes") of malwares, or on the normative analysis of behaviors compared to "healthy behaviors" (knowledge-based detection systems). Both the collective learning of spontaneous groups and advanced machine learning currently outpace signature-based detection systems. The nature of the current paradigm shift is, in this sense, very similar to the evolution of information warfare in the early 1990s. We are witnessing a strategic disruption where defenders are consolidating their information infrastructures, while attackers are engaging in knowledge-warfare (Baumard 1994). Superior knowledge, through astute combination, can be gained from truncated and partial information. Superior information rarely defeats even poorly articulated knowledge.

A *behavioral intelligence paradigm* is synonymous with an inescapable rise of "zero days" threats. Pervasive and highly available combinatory learning allows the creation of many variants of an exploit (exploitation of a vulnerability) within 24 hours of its discovery. Re-encapsulating and re-combining the exploits of undiscovered flaws ("zero days") is made possible

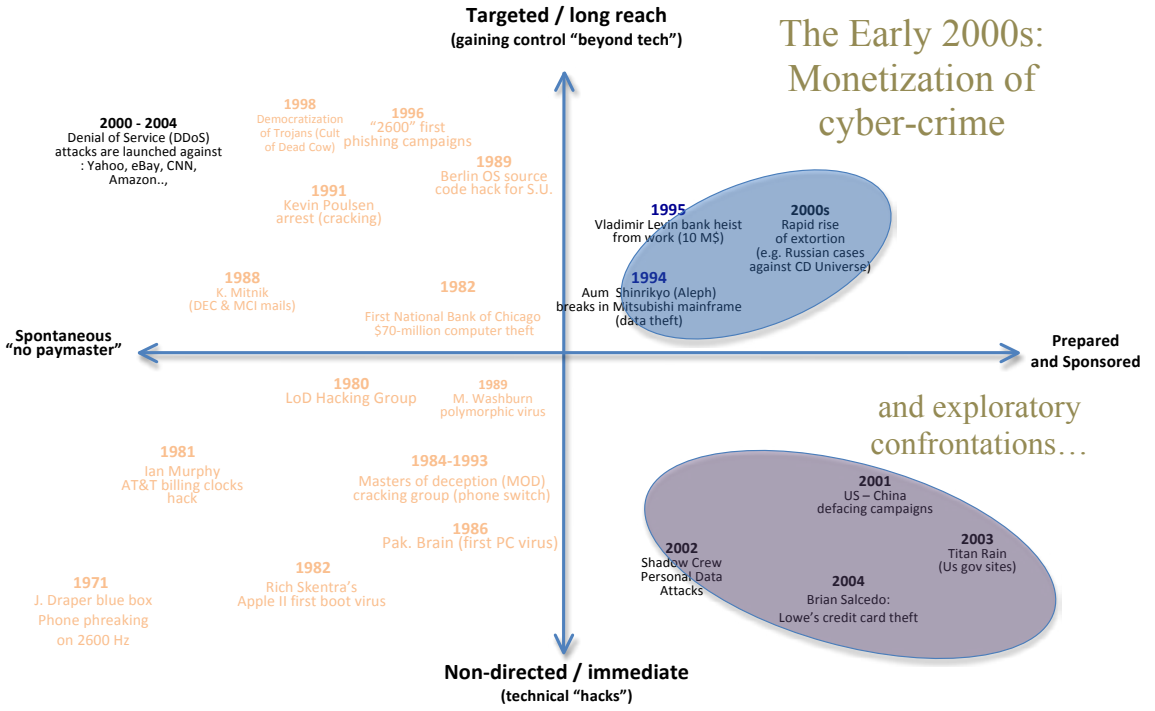

Figure 3. The monetization of cyber-crime and first State confrontations

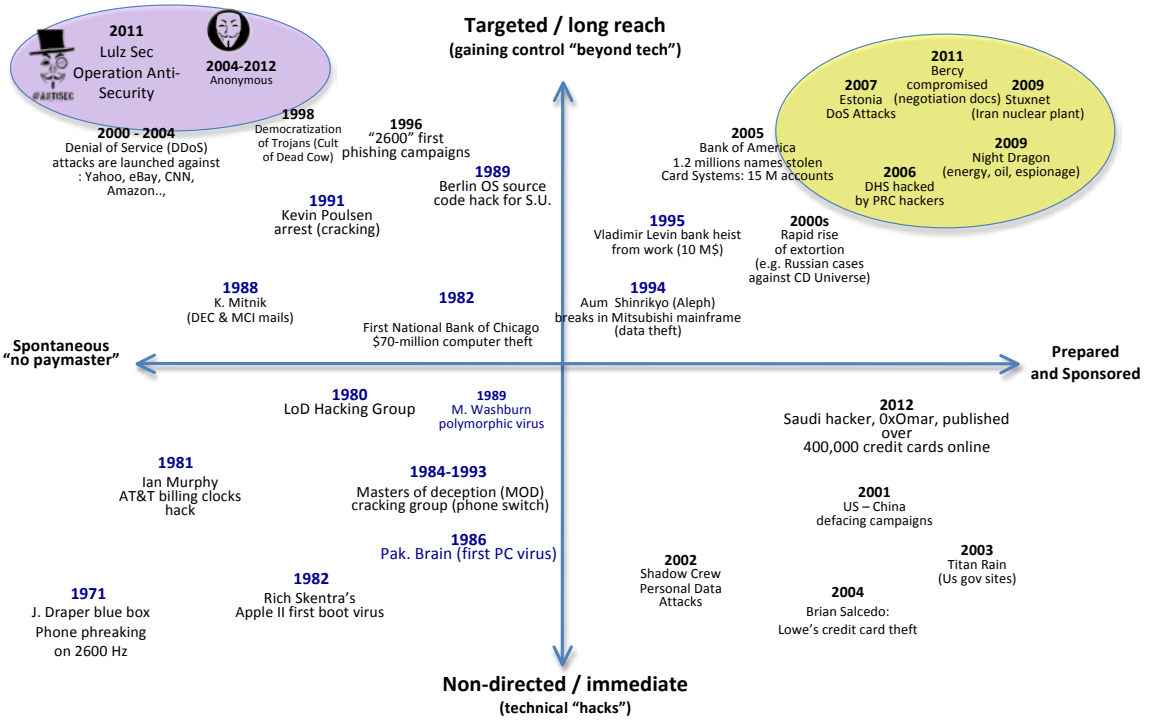

Figure 4. Beyond technology: the rise of large-scale targeted campaigns (2005–2013)

by the advancement of causative learning techniques, or when inaccessible, by the very large number of spontaneous hacking groups sharing their recombination experiments. In such a paradigm, focusing on ex-post defense strategy based on known and identified vulnerabilities is likely to fail.

Putting contemporary doctrines to the test of technological shifts

By gathering data from public sources on published Cyber-Defense doctrines, we try in the second part of this analysis to assess the soundness of Cyber-Doctrines for the deterrence of behavioral intelligence-driven threats. We analyzed 38 national strategies to fight cyber-crime, implement cyber-defense, and promote resilient information infrastructures and cyber-security.

We used the framework developed earlier on the history of cyber-criminality to categorize four categories of cyber-crimes, based on their destination ("targeted and long-reach" versus "immediate or non-directed") and their preparation ("spontaneous" versus "prepared and sponsored"). Hence, we identify four classes of cyber-crime: "code warriors" (I), "cyber free riders" (II), "autonomous collectives" (III), and "sponsored attackers" (IV).

Different classes of attacks require different responses. Immediate and spontaneous attacks (Class I) can be handled with robust information security, including causative learning that can deter sophisticated AI attacks. Most national doctrines have a sound understanding and appropriate range of responses for such attacks. Prepared and sponsored immediate attacks (computer theft by organized crime, free-riding, phishing, and cracking—Class II) require a coordinated range of technical and jurisdictional responses. Signature-based detection systems and knowledge-based defenses are usually sufficient to deter most threats, as far as regulation is judicially enforced. Socially and society-rooted attacks (hactivist groups, temporary or goal-driven groups with political, societal, or economic motives—Class III) involve perception warfare, information warfare, and sense-making capabilities so as to respond to rapid and emergent distributed deployment. Finally, offensive campaigns with embedded behavioral intelligence (Class IV) require transversal responses that encompass proactive deterrence "beyond tech" and "beyond claim". Class III and Class IV threats call for real-time sense-making on unprecedented scales, involving large-scale human cognitive learning on one side (III) and large-scale behavioral learning on the other side (IV).

Our analysis of the evolution of national cyber-crime doctrines over the period 1994–2013 brings mixed findings. *"Power-sovereign" doctrines* (P-S, Class IV) emphasize the development of large specialized units, are often obsessed with critical infrastructures protection, and develop more or less publicly, offensive capabilities. While they deliver sustainable deterrence policies on State-sponsored cyber attacks, they usually develop a threat-rigidity dominant logic, which impedes their involvement in emergent societal change. The risk for P-S doctrines is therefore disconnecting with emergent hacking movements, and a lack of reactivity to distributed cognitive warfare. *"Societal Resilience"* doctrines (Class III), on the other hand, are more sensitive to opinion movements, try to leverage the public space, and focus their offensive capabilities on information warfare. Motivation for such doctrines is not always rooted in a democratic and progressive view of the Internet. Yet, the digitalization of society is clearly identified as both the core threat and core opportunity for cyber-defense and cyber-development.

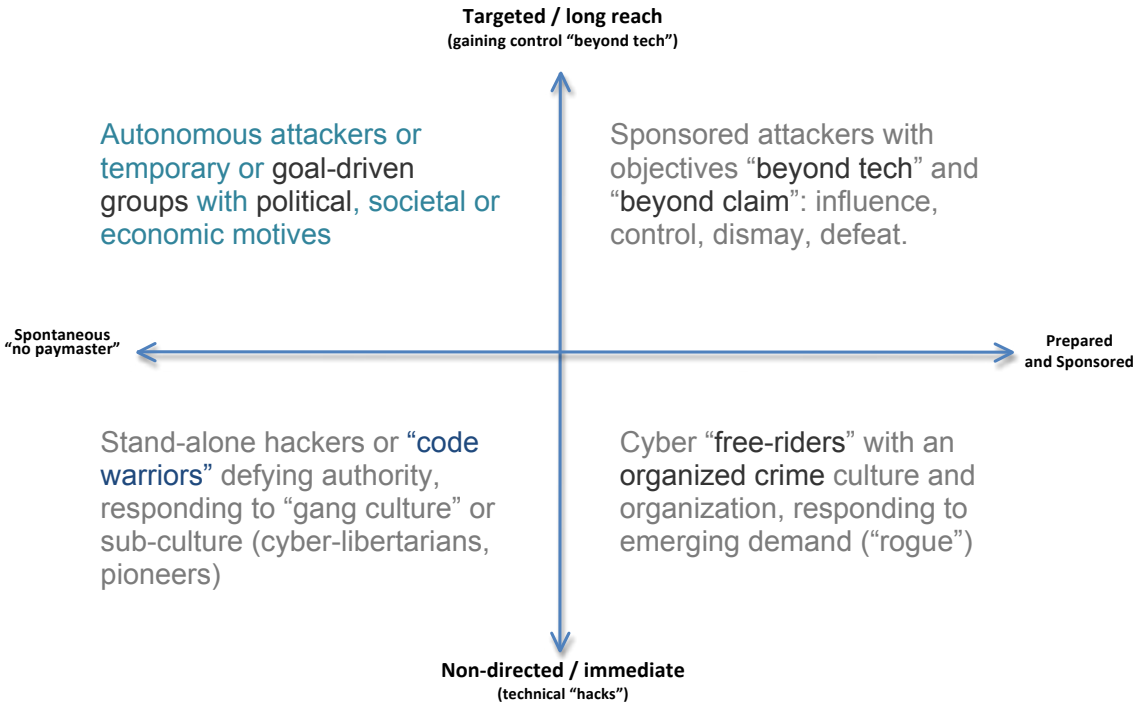
Targeted / long reach
(gaining control "beyond tech")
Autonomous attackers or temporary or goal-driven groups with political, societal or economic motives
Sponsored attackers with objectives "beyond tech" and "beyond claim": influence, control, dismay, defeat.
Spontaneous "no paymaster"
Prepared and Sponsored
Stand-alone hackers or "code warriors" defying authority, responding to "gang culture" or sub-culture (cyber-libertarians, pioneers)
Cyber "free-riders" with an organized crime culture and organization, responding to emerging demand ("rogue")
Non-directed / immediate
(technical "hacks")

Figure 5

Finally, "Social order" doctrines (Class I) and "Technocratic" doctrines (Class II) only differ in their perception of control. The main difference lies in a control at the source (I) versus a control by a normalization of the outputs (II). Technocratic perspectives often suffer from a delayed perception of technological change, mainly inspired by an incident-response philosophy or a late entry to the field. Doctrines that favor social order generally suffer from a lack of national vision or national strategy, or have built their policies by borrowing (or aligning to) external national visions.

The following graph presents the positioning of different national cyber-crime deterrence and cyber-defense strategies (year indicates date of first document analyzed). The findings illustrate the trade-off between national policies that focused on organized cyber-crime and policies driven by the surveillance (or the support) of the societal rooting of cyber-developments. Interestingly, the Russian cyber-doctrine is closer to emergent societal developments than its Chinese or U.S. counterparts.

Measuring the robustness of national strategies: what to expect?

Most of the studied national strategies derive their national cyber criminality deterrence with an average delay of 10–15 years with the advancement of technology. Accordingly, society-wide disruptions have been systematically overlooked. Typically, cyber-policies grow in the fxourth class, while the most disruptive change is taking place in the third.

Core hacking technologies have been steadily stable in the 1990–2012 period. Advanced Persistent Threats (APTs) are not *per se* the result of a disruption in core exploits, but rather a paradigmatic change coming from peripheral technologies (mainly machine learning, automation, and combinatory reconfiguration). Such a paradigmatic change thrives on the obsolescence of an aging infrastructure. Combinations are made possible when flaws can be exploited cross-systems. The growing interoperability of vulnerable systems increases the probability of the on-the-fly exploitation of cross-vulnerabilities. In such a context, vendors, by pushing cyber-criminality deterrence to focus on "points of access" vulnerability assessment impede investment in behavioral learning technologies (by maintaining a poorly performing, but highly profitable, signature-based defense paradigm).

The only way to counteract and deter intelligent behaviors is by outpacing and outsmarting its behavioral intelligence. Very few studied doctrines have acknowledged this core systemic vulnerability. Confidence building and security measures (CBSMs) are hence rooted in a technological and societal understanding that may foster vulnerabilities, and suffer from a critical blind spot on the nature of future technological threats.

Technocratic (Class II) and social order (Class I) national doctrines are dependent on vertical and jurisdictional knowledge, while the evolution of threats is horizontal and a-jurisdictional. Most recent large-scale campaigns (APT1, Blaster-worm, etc.) have shown the limits of inter-jurisdictional coordination in responding to attacks with unpredictable attribution, unknown or undiscovered signatures, and using causative learning to adapt to common technical responses.

Most of the analyzed doctrines presented an outdated perception of authorship and attribution. Attribution is assimilated in most doctrines with a geographical point of emission (or several), a central intent, and a legalist perspective on tracking back attacks.

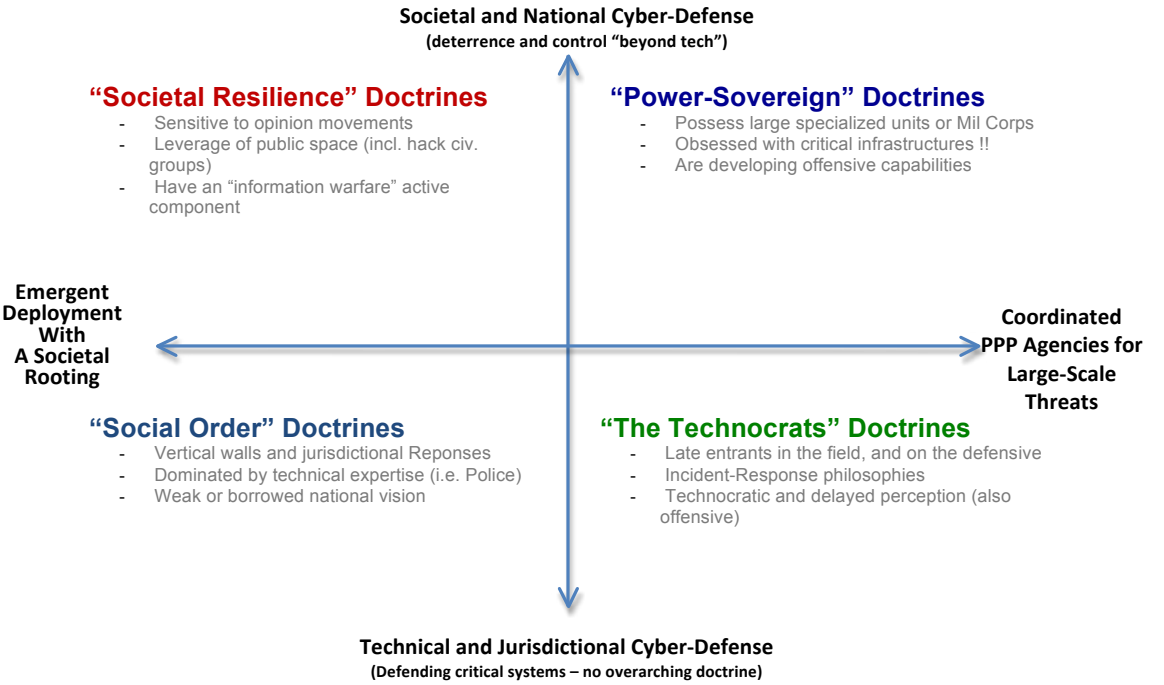

Figure 6

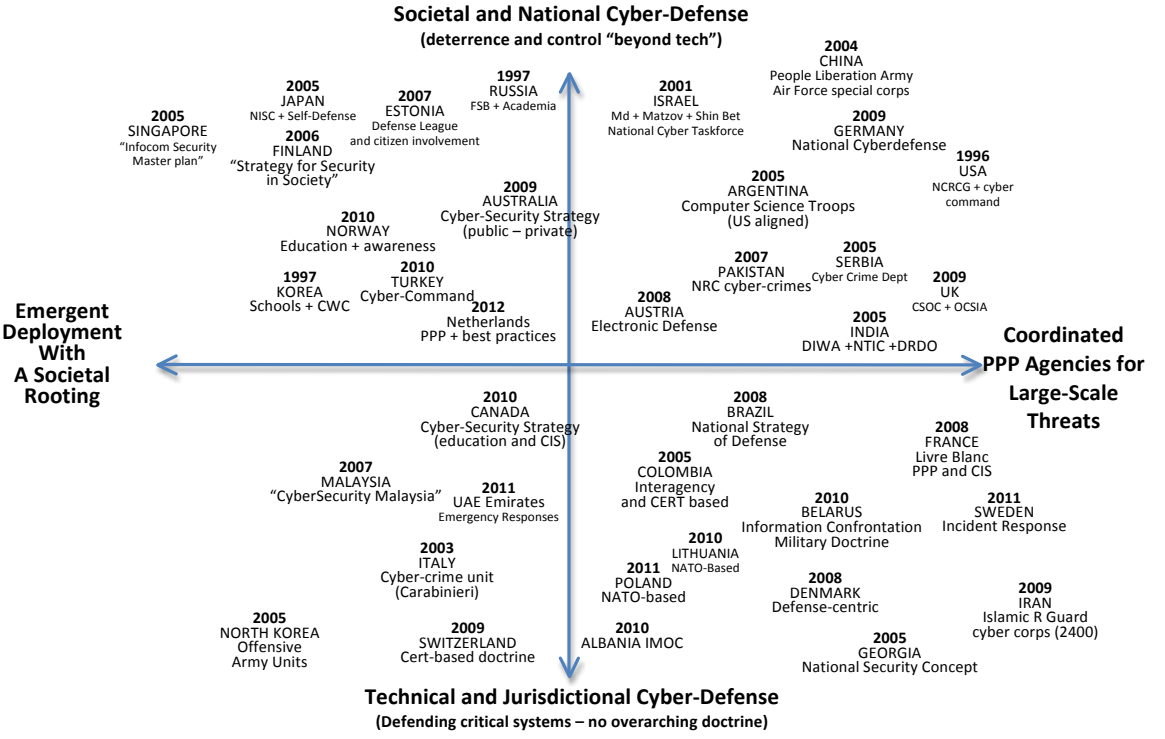

Figure 7

Erasing traces of presence, or traces of intrusion, has been long mastered by the hacking community, leading to the conclusion that diplomatic efforts are geared toward resolving an issue that has lost its technological pertinence before 2007.

Understanding the social psychology of threats development is becoming critical, as we are entering a pioneering period that strangely resembles the "phreaking" years of hacking (1972–1987). The improved portability of machine learning (embarked, distributed, or fully autonomous) is curiously absent from most national strategies' assumptions. This may be driven by the transposition of the principles of military capabilities escalation (weapons race, concentration, and decisive capacities) to the tackling of cyber-criminality. Cybernetic offensive capabilities do not respond to traditional escalation and reinforcement models. They derive their malevolent capabilities from their transformational nature, their distributed deployment, and their superior and autonomous learning.

References

Barreno, M., P.L. Bartlett, F.J. Chi, A.D. Joseph, B. Nelson, B.I.P. Rubinstein, U. Saini, and J.D. Tygar. 2008. "Open Problems in the Security of Learning", *First ACM Workshop on Security and Artificial Intelligence* (AISec), , Alexandria, Virginia, 19-26

Baumard, P. 1994. "From Information Warfare to Knowledge Warfare." In *Information Warfare*, ed. W. Schwartau. New York: Thunder's Mouth Press, 611-626

Bodmer, Kilger, Carpenter, and Jones. 2012. *Reverse Deception: Organized Cyber Threat Counter-Exploitation*. New York: McGraw-Hill Osborne Media.

Gaycken, S. 2012. "Die sieben Plagen des Cyberwar." In *Automatisierung und Digitalisierung des Krieges*, eds. R. Schmidt-Radefeldt, and C. Meissler. Berlin: Forum Innere Führung.

Rubinstein, B. I.P., B.Nelson, L. Huang, A.D. Joseph, S.-H. Lau, S. Rao, N. Taft, and J.D. Tygar. 2009. *"ANTIDOTE: Understanding and Defending against Poisoning of Anomaly Detectors", IMC '09: Proceedings of the Ninth ACM SIGCOMM on Internet Measurement Conference*, Chicago, IL, 1-14.

Sterling, B. 1994. *"Part Three: Law and Order". The Hacker Crackdown: Law And Disorder On The Electronic Frontier*. New York: Bantam Books.

Stoll, C. 1988. "Stalking the Wily Hacker." *Communications of the ACM* 31 (5): 484-500.

Stoll, C. 1989. *The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage*. New-York: DoubleBay.