

3675, original, urquiola

Virtually Massive: Understanding Mass and Combat Power in Cyber War

Captain John "Strider" Cobb, USAF

“Quantity has a quality all its own”
-apocryphally attributed to Josef Stalin, discussing Russian weapons production in WWII

The US Air Force defines “mass” as “to concentrate the effects of combat power at the most advantageous place and time to achieve decisive results”¹; Air Force Doctrine Document (AFDD) 3-12 echoes this definition while noting that cyber forces “[m]ust integrate and synchronize with other forces”². But what does this mean for strategy in the cyber domain? Some have suggested that the concept of mass no longer applies in cyberspace, and that a handful of attackers could launch devastating attacks from anywhere in the world³. Col (Ret) Gregory Rattray in “Strategic Warfare in Cyberspace” discusses the support functions, such as network intelligence, targeting, and tool development, that can make cyber attacks more effective. This suggests another perspective on mass—one including not just the “tooth” of operators attacking targets, but also the “tail” required to put together a team capable of consistently launching successful attacks (since he suggests the analysts and programmers might need to significantly outnumber the people actually carrying out attacks)⁴. This is a significant difference, and an important question to resolve, but personnel issues are not the only way to understand mass in cyber war.

Mass can also be thought of in terms of the volume of attack traffic—at the simplest level, in terms of the number of bits or packets passing into the target. At other times a more meaningful definition of volume might be the number of viruses being released simultaneously, typically from the perspective of the firewalls and antivirus tools that will try to block the malware. What attackers or researchers consider one virus might use self-modifying code to appear as hundreds of different viruses to the filters it tries to sneak through. Alternatively, the number of nodes under attack could be a worthwhile definition of volume—at the lowest level, the number of devices being attacked; at a higher level, the number of network domains or physical sites (i.e. bases) under attack. This is an oversimplification, but these different aspects of “volume” can be thought of as a cyberspace version of firepower. A third way mass can be understood in cyber war is in terms of the robustness or survivability of networks being defended. A network with lots of spare devices, bandwidth, and redundant paths will be more survivable (or at least able to recover more quickly) against a variety of attacks. Likewise, the number and skill of the technicians maintaining and defending the networks under attack is an often-overlooked way to consider mass. Finally, most countries will be more constrained in peacetime cyber actions than when launching wartime cyber attacks, so mass may not take on as many aspects in peacetime covert actions as in open cyber war.

Since many readers may not be familiar with how firewalls and antivirus software protect networks, here's a quick explanation. Typically a firewall scans the traffic entering or leaving a local network, while antivirus software scans the hard drive of a specific computer. There are exceptions like Network-based Intrusion Detection Systems or router Access Control List filters, but the majority of the devices protecting a network are firewalls or traditional antivirus- and most alternative tools use similar rules for their scans and filters⁵. There are two primary ways they scan for viruses, signatures and heuristics. The most common way a network's defenses search out malware is by checking signatures, usually just matching part or all of a file (or network packet) against a list of known malware. This can be fairly effective against known attacks, but it obviously fails to detect most new attacks, and it often fails to catch known attacks that have received even minor changes⁶.

The other way network defenses find malware is by using "heuristics"- rather than looking for specific text or files that are known to be malicious, they look for suspicious patterns or behavior. This is more effective (though not perfect) at catching new ("zero-day") attacks and much better at catching new variations of old attacks; however, most heuristic algorithms have high false positive rates. Because they frequently flag or block legitimate traffic and applications more often than actual attacks, they can be extremely time-intensive and frustrating to operate. These filters face statistical problems similar to some cancer screenings⁷- since the vast majority of network traffic isn't malicious, flagging 95% of viruses and 1% of legitimate traffic will result in more false positives than actual viruses⁸. Many commercial products use a combination of both approaches, and security researchers are working on several alternative ways to detect attacks, but today all but the most important and secure networks tend to rely heavily on signature-based detection, since lack of resources makes other methods infeasible⁹. As a result, on many large military or industrial networks new attacks will not be detected until it's too late and they have already overrun the network.

Because of these limitations, in the current state of cyberspace an attack's effects are not always proportional to its sophistication. The size, complexity, defenses, and interconnectedness of the target determine how sophisticated an attack must be to succeed. An attacker's political constraints, such as requiring stealth and non-attribution--or false attribution--often increase the level of sophistication required, as well as the workload for the attackers. However, because those factors vary between major military and infrastructure targets, some critical targets are nearly impossible to secure without crippling their functionality, while others--especially smaller targets--even if less critical, may require highly sophisticated attacks. DoS attacks can often afford to be less sophisticated than espionage attempts; large, geographically-distributed, multi-node networks offer more opportunities to slip an attack in and make the job of monitoring defenses much harder; Commercial Off-The-Shelf (COTS)-based hardware and software can be attacked with standard training and tools, whereas the specialized systems often found in intelligence, nuclear, or infrastructure settings may require extensive reconnaissance and custom-built attacks¹⁰.

Although his work makes some highly questionable assumptions, Thomas Rid¹¹ is correct in noting that stealthy, highly targeted, "strategic" attacks like Stuxnet¹² require significant resources to assemble and are not easily reused; this type of attack requires dozens of intelligence analysts, programmers, and

operators to design, assemble, and launch. While different groups may combine some or all of those functions in the training an individual cyber warrior receives, this type of attack against a highly defended target will still require months of work from dozens of highly-trained people. To the extent that a cyber campaign seeks to quickly corrupt or disable large numbers of hardened, stand-alone targets, it will clearly require significant investments, and those investments rise if stealth or non-attribution is needed- particularly since the teams working on the stand-alone targets will be unavailable for “routine” cyber intelligence work, which will presumably still be needed.

However, there are also many cyberspace targets that are less robustly defended, and some of them can be high-value targets. Although nuclear facilities like the ones targeted by Stuxnet are likely to remain “air-gapped” and carefully defended, many military and industrial systems require much broader access to be effective. Power grids, for example, cannot function efficiently unless the individual power stations are constantly communicating with each other. They may not be “online” in the sense of sending unencrypted traffic directly over the Internet, but that interconnection can still be used to control, degrade, or destroy the entire grid with a single attack- although it must be noted that in large countries like the US or Russia, “the power grid” is actually a set of loosely coupled regional grids; for example, the US has 3 regional grids¹³, and Russia, though more centralized than the US, has 7 regions with limited interconnections¹⁴. Many military networks face the same situation- they must link a large number of units and bases to provide the force multiplier effects all modern militaries rely on, which makes it very difficult to keep sophisticated attacks from spreading once they compromise one node on the network.

In this environment, the second definition of mass is probably a better way to understand it. Some targets are vulnerable to attacks that can be created by a single team and then launched against the entire network, and (if successful) can severely hamper an entire region. Linked utilities are one example, particularly power grids, which tend to be larger, more interdependent networks than other utility networks; another might be the military networks used for C2, logistics, and operational situational awareness. These functions are needed at the tactical or operational level, and therefore typically require a network shared across a large number of critical nodes. At the extremely simple end of the spectrum of possible cyber attacks, mass in a Distributed Denial of Service (DDoS) attack is simply the amount of bandwidth used to try to overwhelm the target network. As we progress up the spectrum, if an attack uses effective self-modifying code, the defender may need to block not one but hundreds or thousands of new virus signatures at dozens to thousands of network perimeters (each firewall protecting the local area or network must be updated with all the relevant virus signatures to block). In some cases, the number of “exploits” viruses are using might be more relevant than how a virus mutates to sneak through filters- an exploit is the malicious code that uses an error in a target’s software to take control of the system; when more vulnerabilities are exploited, the virus is more likely to succeed¹⁵. Most military networks have technicians who spend enormous amounts of time trying to patch known vulnerabilities before an attack takes advantage of them, knowing that firewalls and antivirus don’t stop all attacks. An attack that only uses one exploit will fail if the corresponding vulnerability is thoroughly patched on its target; while patching quickly and thoroughly is a very hard problem¹⁶, there are tools that enable large modern networks to patch most systems in a matter of days.¹⁷ If an attack uses more than one exploit, it increases the challenge defenders face; however,

exploits (particularly zero-day exploits) can be valuable intelligence assets, and many authorities advise against revealing them wantonly.¹⁸ On the other hand, these exploits are more numerous than many people realize; for example, between January and May of 2012, Microsoft announced--and released patches for--31 serious vulnerabilities in the Windows 7 operating system and its common applications (i.e. the Microsoft Office suite).¹⁹

Alternatively, there are attacks where mass is better understood as the number of networks targeted by an attack or series of attacks. The author's previous work discusses attacks that can cut local networks off from centralized control or response²⁰; in these types of attacks the number of geographic locations or local networks attacked may be more meaningful than the manpower used to create the attacks or the variety of attacks used. This applies primarily to Denial of Service (DoS) and DDoS attacks²¹, or to highly centralized network defenses- if there are skilled network defenders at each location, they should be able to clear up simple attacks relatively easily (particularly DDoS attacks), whereas a more centralized approach to network defense may lead to defenders being cut off from the networks underneath them, or to defenders being spread too thin to react in all locations simultaneously.

Conversely, in these types of attack manpower can be the critical element of mass for the defenders; unlike highly targeted attacks, these types of DoS attacks can be highly asymmetric. Although they are subject to the arms race between attackers and defenders, as filters, scans, and network tools race to catch up to hacking techniques, these DoS attacks can be launched by fairly small teams, and can require large numbers of defenders to patch and clean each local network under attack.

The relevant measure of combat power in cyberspace varies depending on the types of attacks in play and networks needing defense; while knowledgeable technicians and operators are always important²², the appropriate balance between quality and quantity depends on the attacks and tools a military anticipates using and facing. Given the budgetary limits all militaries face, and the wide range of possible cyber attacks, these are trade-offs that all militaries must make. If the primary threat is from stealthy attacks similar to Stuxnet, intended to covertly corrupt data or invisibly damage specific facilities, then network defenders need to be very highly trained, especially in forensic techniques. However, they may not need to be much more numerous than their attackers, and network defenses can often afford to catch attacks after the attack has (at least partially) succeeded, since the attacks will often spread slowly and inflict damage slowly in order to avoid detection. If the primary threat is instead attacks designed to quickly take down or isolate critical systems, network defenders will need to be much more numerous, and their skills should focus on countering overt attacks (and rebuilding the systems an attacker does successfully take down). To acquire enough knowledge about potential adversaries to know which case applies typically requires a significant investment in cyber-focused intelligence personnel; more such personnel will be needed to enable successful attacks in response.²³ It is possible for both types of attacks to be launched simultaneously, but in most cases if these attacks are hitting the same networks and systems they will interfere with each other (often resulting in the attacker losing control of the more precise highly-targeted attacks); what is more likely is for overt DoS or DDoS attacks to target one area while a different system or network is attacked more carefully and covertly, away from the obvious

attack. Of course, in any conflict today coordinating and deconflicting cyber attacks--both from each other and from nearby kinetic attacks--is a critical planning effort.²⁴

A look at airpower history may help clarify these different approaches- in some ways, cyber attackers face a similar problem to the one faced by the USAAF and RAF in their bombing campaigns against Germany's industrial system in WWII. Highly targeted (Stuxnet-style) attacks--taking down a handful of isolated high-value targets with one-time uniquely-crafted attacks, often slowly and covertly--can be thought of as similar to the WWII US approach to strategic bombing. In this approach, the attacker finds a handful of critical nodes and spends large amounts of manpower, time, and combat power to bring those systems down²⁵, confident that the enemy will be unable to function without those critical nodes. In contrast, the various forms of DoS attacks can be thought of as similar to the British approach to targeting- rather than directly attacking specific (often highly defended) critical systems, the attacker attempts to take down some aspect of the local or regional network the critical systems rely on. The WWII RAF approach to targeting was initially based on targeting key factories, but after initial failures (and high losses), the RAF switched to simply bombing major industrial cities, targeting the entire industrial ecosystem rather than a series of discrete "key nodes". The USAAF, despite a similar disappointing start, targeted what it considered vital nodes--primarily aircraft production, fuel, transportation, and ball bearings--from 1943-1945, attempting to crash the German industrial system by destroying or disabling a handful of key nodes (such as the ball-bearing factory at Schweinfurt and the refinery at Ploesti)²⁶. Of course, in execution, the USAAF approach and the RAF approach were not always quite so different (USAAF B-17s and B-24s were often incapable of placing their bombs on target with weather and German defenses interfering),²⁷ but the two targeting approaches bear similarities to the cyber equivalents discussed above. Of course, PCs and switches on a military network are less controversial targets than the populace of industrial cities, but many DoS attacks are less precise and more likely to spill over and have unintended consequences in civilian cyberspace. It is important to note that, depending on the situation, both types of attacks can be highly effective- Stuxnet is an obvious example of a highly targeted attack; the 2007 DDoS attacks on Estonia provide an example of one DDoS attack that had significant impacts²⁸, while DoS attacks alleged to have been coordinated by Russia damaged Georgia's ability to respond to Russian armor advances in 2008²⁹.

Consequently, while taking a single reactor down is extremely difficult and typically requires carefully crafting a brand-new attack, disrupting C2 or logistics across a military is often practical using "off-the-shelf" hacking tools to put together a relatively crude attack. Unlike in most other domains, in cyberspace a larger, more-distributed target is much easier to attack and cripple. Recent discussions of cyber war have often mistaken espionage for war, and as a result some commentators have assumed that exploits are more valuable than they would be in wartime, that stealth is more necessary than it would be in wartime, and that large military or infrastructure networks possess the same defenses that small intelligence or nuclear systems often have³⁰. These misunderstandings distort the nature of mass in cyber war, and they can lead to major mistakes in organizing cyber war units or creating network defenses. While cyber espionage--to include covert cyber attacks in peacetime--will tend to involve stealthy, highly targeted attacks, cyber attacks in open war can span all the different forms of mass and combat power discussed above. Diplomatic and political constraints may vary from a "cyber-only"

conflict to a traditional (“kinetic”) war that includes cyber attacks, but both are likely to include more overt attacks that sacrifice stealth to strike harder and faster, enabling widespread attacks with limited resources.

Mass matters in cyber war, but its meaning varies depending on the nature of the attack and the target. While the tactical level may not always be impacted too severely (particularly in non-US militaries and/or ground forces, where networks may be less heavily used), at the operational level modern militaries rely heavily on their networks for logistics and situational awareness. These networks are “centers of gravity”, and if they can be disrupted it can significantly reduce the effectiveness of the forces that depend on them. While highly targeted attacks can be devastatingly effective, there are also DoS alternatives that can be crippling when targeted and executed correctly. The correct definitions of mass and combat power in cyber war are fluid, like cyberspace itself, and militaries that restrict themselves to one facet of it risk defeat when an adversary attacks in ways that do not match their doctrine and organization.

¹ Air Force Doctrine Document (AFDD) 1, p. 32

² AFDD 3-12, Cyberspace Operations, p. 16

³ Martin Libicki, *Cyberdeterrence and Cyberwar* (Santa Monica, CA:RAND Corp, 2009), p. 59. Available at <http://www.rand.org/pubs/monographs/MG877.html>

⁴ Gregory J. Rattray, *Strategic Warfare in Cyberspace* (Boston: MIT Press, 2001), pp. 100, 464

⁵ See Edward Skoudis, ed. Franklin D. Kramer, Stuart H. Starr, and Larry K. Wentz, “Information Security Issues in Cyberspace”, *Cyberpower and National Security* (Dulles, VA: Potomac Books, 2009)

⁶ “Antivirus Software”, Wikipedia, http://en.wikipedia.org/wiki/Antivirus_software. Accessed 30 May 2012.

⁷ Maia Szalavitz, “Why People Stick with Cancer Screening, Even When it Causes Harm”, Healthland, Time.com, <http://healthland.time.com/2012/05/25/why-people-cling-to-cancerscreening-and-other-questionable-medical-interventions-even-when-they-cause-harm/>

⁸ “Antivirus Software”, Wikipedia

⁹ For more information, see Shon Harris, *CISSP All in One Exam Guide, 4th ed.* (New York:McGraw Hill, 2008), pp. 250-257.

¹⁰ However, those specialized systems may be weaker and more vulnerable than their standard COTS counterparts once the attackers understand them- especially infrastructure systems, which were often designed without any thought to security

¹¹ Thomas Rid, “Think Again: Cyberwar” Foreign Policy, 27 February 2012, <http://www.foreignpolicy.com/articles/2012/02/27/cyberwar?page=0,2>

¹² Stuxnet was a highly sophisticated Trojan propagated over thumb drives that infiltrated and sabotaged Iranian nuclear facilities in 2010 for several months before being detected. See “The Stuxnet Outbreak”, The Economist, Sept 30, 2010. <http://www.economist.com/node/17147818>

¹³ Richard A. Clarke and Robert K. Knake, *Cyberwar: The Next Threat to National Security and What to Do about It* (New York: HarperCollins, 2010), 167.

¹⁴ Rinat Abdurafikov, "Russian Electricity Market: Current State and Perspectives", VTT Technical Research Centre of Finland, 2009. <http://www.vtt.fi/inf/pdf/workingpapers/2009/W121.pdf>

¹⁵ "Exploit", Wikipedia, [http://en.wikipedia.org/wiki/Exploit_\(computer_security\)](http://en.wikipedia.org/wiki/Exploit_(computer_security)), accessed 27 May 2012. See also Harris, p. 62.

¹⁶ See Stephen Northcutt, et al, *Inside Network Perimeter Security*, 2nd ed. (Indianapolis: Sams Publishing, 2005), pp. 257-258 for a discussion of factors that make patching a large network difficult.

¹⁷ Microsoft Security Update Guide, 2nd ed. Microsoft, 2011 <http://www.microsoft.com/en-us/download/details.aspx?id=559> For a brief summary of one patching system, see "WSUS Overview", Microsoft Technet, <http://technet.microsoft.com/en-us/library/cc539281.aspx>

¹⁸ Libicki, p. 57

¹⁹ "Patch Tuesday (January-May 2012)" Microsoft Technet, <http://technet.microsoft.com/en-us/security/bulletin/ms12-jan>, <http://technet.microsoft.com/en-us/security/bulletin/ms12-feb>, <http://technet.microsoft.com/en-us/security/bulletin/ms12-mar>, <http://technet.microsoft.com/en-us/security/bulletin/ms12-apr>, <http://technet.microsoft.com/en-us/security/bulletin/ms12-may>

²⁰ John Cobb, "Centralized Execution, Decentralized Chaos: How the Air Force is Poised to Lose a Cyber War", Air and Space Power Journal, Summer 2011, http://www.au.af.mil/au/cadre/aspj/airchronicles/apj/2011/2011-2/2011_2_16_cobb.pdf

²¹ DoS refers to any attack focused on taking a system or network down by overloading it. DDoS refers to a specific type of DoS where massive amounts of network traffic from computers around the world are sent to overwhelm the target server or network, causing it to freeze and stop responding to legitimate traffic. See Harris, pp. 1010-1013.

²² For a deeper discussion of this issue, see Kamal Jabbour, "Cyber Vision and Cyber Force Development" Strategic Studies Quarterly, Spring 2010, www.au.af.mil/au/ssq/2010/spring/jabbour.pdf

²³ Rattray, p. 142

²⁴ See AFDD 3-12, pp. 25-26 for further discussion of deconflicting cyber attacks

²⁵ Elinor Mills, "Expert: Stuxnet was Built to Sabotage Nuclear Plant". Insecurity Complex, CNET.com, http://news.cnet.com/8301-27080_3-20017201-245.html?tag=mncol;1n

²⁶ Geoffery Perret, *Winged Victory* (New York: Random House, 1993), pp. 240-244

²⁷ Rattray, p. 280.

²⁸ DDoS attacks alleged to have been coordinated by Russia caused significant disruption to Estonia's banks and government (Russia denied any official involvement). See Clarke and Knake, pp. 11-16.

²⁹ David Hollis, "Cyberwar Case Study: Georgia 2008". Small Wars Journal, January 6, 2011 <http://smallwarsjournal.com/jrnl/art/cyberwar-case-study-georgia-2008>, p. 5

³⁰ Examples of these mistaken assumptions include Rid (especially p. 5) and Libicki (see pp. 56-59, and compare to industry guides on patching referenced above, particularly Northcutt, et al- patching a large network requires significant effort over the course of days, often weeks, and when done too quickly the patch itself may break critical systems).
