CyberGreen Metrics

near- to medium-term direction as of 21 October 2016

---

author	Dan Geer
team	Scott Guthery, Chris Horsley (Cosive), Aaron Kaplan (CERT.AT), Eireann Leverett (Univ. of Cambridge), Art Manion (CERT/CC), Manel Medina (UPC), Kayne Naughton (Cosive), Joe St. Sauver (Farsight Security), David Watson (ShadowServer Foundation)
executive director	Yurie Ito (CyberGreen)

Security metrics are desirable when they enable something, when they have a role to perform that has a receiver ready to make use of them. Otherwise they are stamp collecting.

The issue is one of purpose. The only purpose that makes security metrics worthy of pursuit is that of decision support, where the question being studied is one more of trajectory than exactly measured position. We at CyberGreen are not in this for reasons of science, though those that are in it for science (or philosophy) will also want measurement of some sort to backstop their theorizing. We are in this because the scale of the task compared to the scale of our tools demand force multiplication, and no game play improves without a way to keep score.

It stands repeating that the core Internet protocols were designed against a particular goal state: optimal resistance to random faults in the network fabric. There is no need to elaborate on that here except to remember that it is impossible to have a network design that is optimally resistant to random faults while at the same time optimally resistant to targeted faults. The CyberGreen effort is focused on targeted faults that have effect at something approximating global scale.

There are two genus and five species of cyber attacks:[1]

Passive, *i.e.*, pure listening attacks
  Traffic Analysis — who is talking to whom
  Release of Contents — who is saying what to whom
Active, *i.e.*, packet insertion attacks
  Message Stream Modification — change what they said
  Denial of Service — don't let them talk
  Spurious Association Initiation — false claim of identity

Because (attempts at) passive attacks cannot be detected, they must be prevented. Because (attempts at) active attacks cannot be prevented, they must be detected. Ergo, the possible goals for any communications security technology or strategy:

Prevention of traffic analysis attacks
Prevention of release of contents attacks
Detection of message stream modification attacks
Detection of denial of service attacks
Detection of spurious association initiation attacks

Of those, CyberGreen has nothing directly to do or say about (the prevention of) passive attacks. CyberGreen also has nothing directly to say about false claims of identity ("spurious association") nor about in-network attacks on data integrity ("message stream modification").

At the outset of Obama's first term, Hathaway led a "sixty day review"[2] of the U.S. cybersecurity stance. She concluded that the primary targets were members of the Defense Industrial Base and technology firms with global reach, that the secondary targets were the primary targets' counterparties, and that the tertiary targets were any endpoints that could be used as staging areas for attacks on the primary and secondary targets. It is the latter group — the endpoints that have value as staging areas — that CyberGreen can measure and, perhaps, influence. Nation states, like the U.S., will protect their primary targets and will not look to CyberGreen for assistance. It seems certain that risk-carrying interactions between primary and secondary targets will not be observable to CyberGreen in any useful way. Ergo, consistent with the previous paragraph, that leaves to CyberGreen the particular focus on entities that can be used to stage attacks, attacks that could, in

turn, reach global consciousness. Ergo, it is to CyberGreen to measure, and report on, the degree to which an entity, any entity, is or could be a risk to others.

At this time, Chief Information Security Officers (CISOs and like titles) face two especially difficult kinds of active attacks: denial of service, particularly distributed denial of service (DDoS), and the recruitment of unwitting end-users into inviting trouble in, particularly by way of phishing e-mails.

This short note is about the near- to medium-term direction for CyberGreen, and is therefore limited in its reach. The reader is asked to not conclude that our focus is as narrow as is described here. Over the long-term, CyberGreen will have a broad, effectively exhaustive set of measures of the healthiness of the Internet at large. With that in mind, we begin narrowly so as to do one good job at a time.[3]

It is CyberGreen's considered opinion that the most likely way in which a random entity poses a risk to others is that entity's potential for participation in a DDoS. Therefore, the first stage of CyberGreen's intervention is to report the potentiality of DDoS attacks — that is to say that we accept as a given that there will always be some motive to attack present in some part of the Internet, so our job is measure its opportunity. While it is true that exploited endpoints can be used for any bad activity, we cannot measure some kinds of bad activity so must stick to those we can measure and, to the point, measure in a way that is solid decision support. As such, building CyberGreen's metric suite will begin with an entity's risk to others as measured by its availability to participate in a DDoS.

Unwitting participation in a DDoS does not require pre-existing malware compromise; it requires (only) that proven corrections to protocols and configurations have yet not been made. Those configuration and protocol corrections are well known, and it is possible to remotely test whether the corrections have been done. As such, CyberGreen can analytically determine the prevalence of uncorrectedness and by that test derive a measure of whether an entity is a DDoS risk to others.

Protocols that can be used for a DDoS are those by which an attacker can send disabling amounts of traffic to a target. Some protocols in their natural state return a much bigger reply to a requestor than the size of the requestor's initial query. An attacker will thus make queries appear to come from his actual target so that his actual target gets the response to the initial queries. Because the reponse is larger than the query, the particular protocol involved is said to "amplify" the quantity of an attacker's input into larger quantities of output — for example, if 50 bytes "in" produce 1,000 bytes "out", then we would say that there is an amplification factor of 20. From the attacker's point of view, the higher the amplification factor the better, all else being equal.

There are four protocols that can be particularly useful to an attacker attempting a DDoS,

DNS	Domain Name Service
NTP	Network Time Protocol
SNMP	Simple Network Management Protocol
SSDP	Simple Service Discovery Protocol

in that they are widely present in the Internet and can provide significant amplification. (The technical details of exactly how these protocols can be abused are outside the scope of this note.)

Through the kindness of our partners, CyberGreen has a good and steady stream of observational data on the prevalence of those four protocols in the configuration that makes them available to an DDoS amplification attack. We may well, over time, seek data on other protocols, but with these four we capture a significant share of the recruitable DDoS firepower residing across the Internet and, hence, a measure of risk to others (that these four represent in the aggregate).

For CyberGreen's v2.0 deliverable, we will report a crude measure of DDoS risk to others by country, by AS, and by such alternate entities (*e.g.*, enterprises) as seem relevant. That crude measure is the count of nodes within the scope of control of the country, the AS, or the entity otherwise defined that have the configuration that allows them to participate in a DDoS. The count will be reported by protocol and in sum across all four protocols. Countries, ASs, and alternate entities will

be ranked by the count of nodes available to the operator of a DDoS amplification attack, *i.e.*, a rank of 1 is that of the highest risk. It is that rank that is the v2.0 CyberGreen Index value.

We say "crude" as the Index value is from a straight numeric count, but in point of fact the simple count is a direct measure of the size of the mitigation task that the country, the AS, or the alternate entity needs to undertake if it is to reduce the potential risk to others due to nodes within its scope of control. In short, the v2.0 CyberGreen Index equates risk to others to the size of unmet mitigation tasks required to zero the country's, the AS's, or the alternate entity's risk to others.

Where CyberGreen's 2.0 metrics estimate risk to others by way of the country's, the AS's, or the alternate entity's unmet responsibility to mitigate the end nodes they control, CyberGreen's v2.1 metrics report risk to others in terms of "How bad could it be?" This means that CyberGreen v2.1 metrics factor in the scale potential for amplification by protocol by node. Whereas the v2.0 Index is a rank order by the size of the unmet mitigation need, the v2.1 Index is a rank order by the size of the DDoS that could be mounted from the country, the AS, or the alternate entity should all of their nodes currenly available to attackers were to be used in a single attack. In short, the v2.1 Index measures "offensive potential" — with the obvious caveat that we do not mean intentional offense but rather the degree to which the country, the AS, or the alternate entity can be made to engage in offense whether it wanted to or not.

Rossow's 2014 data on amplification attacks[4] quantifies the opportunity for amplification and does cover the four protocols CyberGreen is using for the Index. Rossow's Table III is reproduced here verbatim:

		BAF		PAF	
Protocol	all	50%	10%	*all*	Scenario
SNMPv2	6.3	8.6	11.3	1.00	*GetBulk* request
NTP	556.9	1083.2	4670.0	10.61	Request "monlist" statistics
DNS-NS	54.6	76.7	98.3	2.08	ANY lookup at author, NS
DNS-OR	28.7	41.2	64.1	1.32	ANY lookup at open resolv.
NetBios	3.8	4.5	4.9	1.00	Name resolution
SSDP	30.8	40.4	75.9	9.92	*SEARCH* request
CharGen	358.8	n/a	n/a	1.00	Character generation request
QOTD	140.3	n/a	n/a	1.00	Quote request
BitTorrent	3.8	5.3	10.3	1.58	File search
Kad	16.3	21.5	22.7	1.00	Peer list exchange
Quake 3	63.9	74.9	82.8	1.01	Server info exchange
Steam	5.5	6.9	14.7	1.12	Server info exchange
ZAv2	36.0	36.6	41.1	1.02	Peer list and cmd exchange
Sality	37.3	37.9	38.4	1.00	URL list exchange
Gameover	45.4	45.9	46.2	5.39	Peer and proxy exchange

TABLE III: Bandwidth amplifier factors per protocols; *all* shows the average BAF of all amplifiers, 50% and 10% show the average BAF when using the worst 50% or 10% of the amplifiers, respectively. The Packet Amplifier Factor is a function of the protocol, *per se*.

While still conjectural, it is expected that with v2.2, CyberGreen Indices will begin to include densities, that is to say ranking will be by percentage of a country's, an AS's, or an alternate entity's nodes that the v2.0 count represents. Summing up, for v2.0, v2.1, v2.2, and beyond, the CyberGreen Index is a (set of) number(s) per observed entity, be that entity a country, an AS, or some other entity we can and do identify. It is the rank of a weighted sum computed in the ordinary way:

$$CG_i = rank(\sum_{j=1}^{j=N} count_{i,j} * weight_j)$$

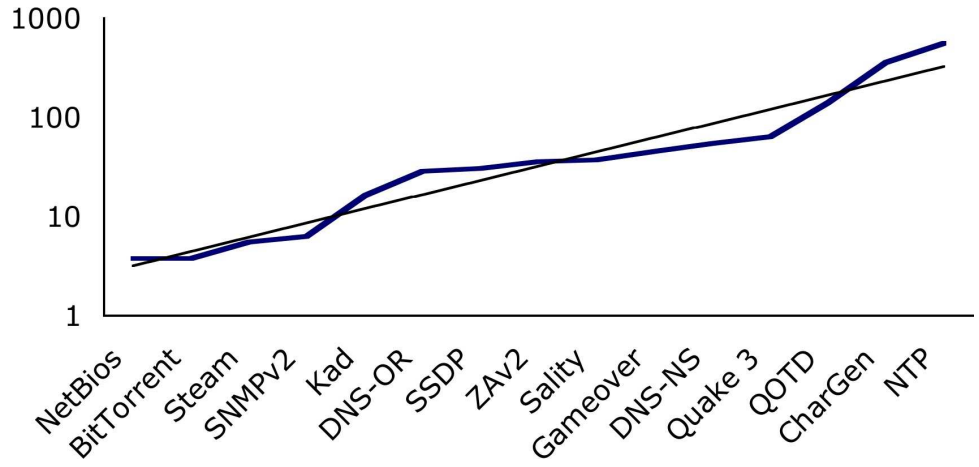
where (for now) $N = 4$ corresponding to the list given before of DNS, NTP, SNMP, and SSDP. The output of the weighted sums by entity form a rank order by risk, that is to say that they are sorted by numeric value. The rank is a declining value such that if entity i has $CG_i = 1$, then entity i has highest Index value and therefore poses the greatest risk to others. Putting v2.0, v2.1, v2.2, and subsequent developments in context:

Choice of weights:	v2.0	simple counts	$weight_j = 1 \; \forall \; j$
	v2.1	+ amplification factor	$weight_j = AF_j$
	v2.2	+ densities	$weight_j = (1/count(nodes\ in\ country_i, etc.))$
	v2.3	+ TBD	$weight_j = TBD$

which is to say that a given country, AS, or alternate entity will have one CyberGreen Index value—for absolute mitigation need — under v2.0, will have two — one for mitigation need and one for offensive potential — under v2.1, and will have three — one for absolute mitigation need, one for offensive potential, and one for prevalence of unmet need — under v2.2. Further development will follow this pattern until we are satisfied that we have captured DDoS reality in a sufficiently actionable form. We will then proceed to a second form of risk to others (yet to be named) and begin work on v3.0.

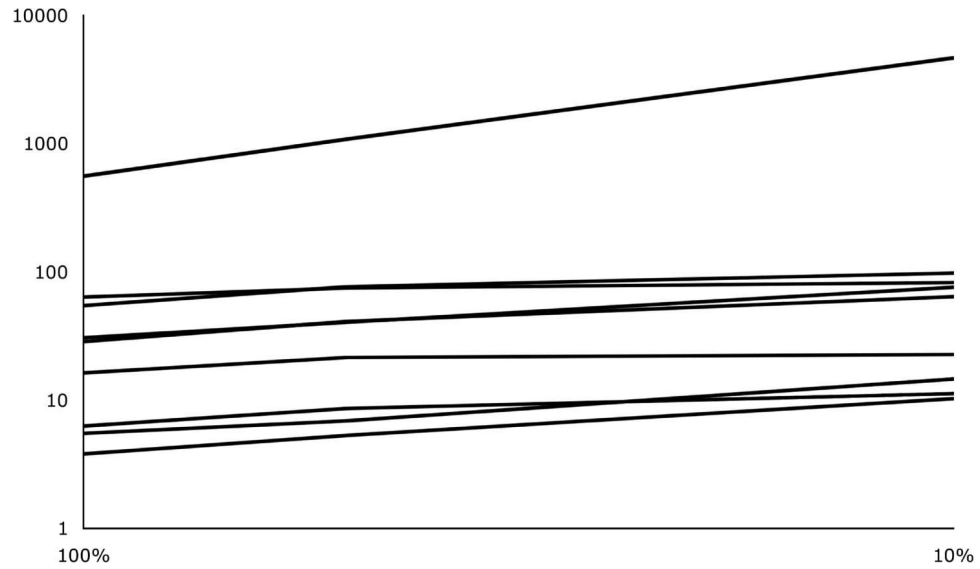
Aside

Plotting Rossow's Bandwidth Amplification Factors (BAFs) from above in sorted order produces a "power law" curve. While detailed discussion of power laws is outside the scope of this note, the term is used here to indicate that the magnitude of a measured variable is proportional to its rank order raised to some exponent. Specifically, this is $f(x) = ax^{-k}$ where x is the rank order, k is the power, and a is some constant.



Straight lines on a log-log graph => power law

In addition to that, each of Rossow's top 10 themselves read like power law curves as you move from 10% to 50% to 100%:



That these appear to be power law type curves may be directly relevant insofar as power laws have a well-defined mean only if their exponent exceeds 2 and have a finite variance only when their exponent exceeds 3. As most identified power laws in nature have exponents such that the mean is well-defined but the variance is not, they are capable of black swan behavior.

In a difficult paper,[5] Nassim Taleb trenchantly concluded that "[We are] undergoing a switch between [continuous low grade volatility] to ... the process moving by jumps, with less and less variations outside of jumps". Put differently, CyberGreen must be continuously on guard against implying that things are getting better and better if our argument for that is based on an assumption of Gaussian error and/or the Law of Large Numbers: When a distribution is fat-tailed, estimations of parameters based on historical experience will inevitably mislead. CyberGreen expects to find and explore power law relationships in this space, but that is for a later time.

Endnotes

[1] Voydock V & Kent S, "Security Mechanisms in High-Level Network Protocols," *Computing Surveys,* 15:2, June 1983

[2] Hathaway M, White House, "Cyberspace Policy Review," 8 May 2009, www.whitehouse.gov/assets/documents/Cyberspace_Policy_Review_final.pdf

[3] As in the secular by Albert Einstein — "Whoever is careless with the truth in small matters cannot be trusted with important matters." or as in the sacred by Luke 16:10 — "He that is faithful in that which is least is faithful also in much."

[4] Rossow C, "Amplification Hell: Revisiting Network Protocols for DDoS Abuse", Symposium on Network and Distributed System Security, 2014; Legend edited slightly for contextual clarity. Retrieved from www.internetsociety.org/sites/default/files/01_5.pdf

[5] Taleb N, "On the Super-Additivity and Estimation Biases of Quantile Contributions", www.fooledbyrandomness.com/longpeace.pdf