GFCE
Global Forum on Cyber Expertise

Global Good Practices

Practice: **Establish a clearinghouse for gathering systemic risk conditions data in global networks**

#Clearinghouse

> **We assess our personal health based on the trusted data we receive from doctors. Cybersecurity is like public health: if CERTs and operators have trusted data — regularly updated — about weaknesses in our networks, this helps them mitigate vulnerabilities, preserve cyber-health, and prevent incidents.**

Related thematic areas:

Research and development

Cooperation and community building

Incident management and infrastructure protection

Of particular interest to:

PRIVATE SECTOR

EXPERTS

Description

Internet networks are replete with systemic vulnerabilities. CERTs and other trusted operators require reliable information about their network's health over time. Various organisations have set up systems to scan networks for vulnerabilities and/or monitor cyber-attacks. Many of these sources are open, but their provenance and collection processes are often opaque. To acquire a truly satisfactory picture of the Internet's behaviour, a clearinghouse is needed that does not simply collect data, but leverages its collections to improve the process.

The clearinghouse collects raw data from multiple sources and processes it, in order to feed into Internet health metrics. Data is collected from carefully selected comprehensive data sources, and processed to ensure it is accurate and extensive, and its biases understood and addressed. It can then be analysed and contextualised to produce reliable metrics about how healthy the Internet is.

Actors (or who this is for)

The clearinghouse produces quality data sources that can be used by CERTs, top-level ISPs, and national infrastructure organisations, as well as skilled technical departments within companies or organisations, and regulators to track the health of the ecosystem and suggest improvements. It also allows them to use the clearinghouse's aggregated data along with local proprietary data to generate their own statistics to measure and track the ecosystem's health.

Researchers from multiple communities — academia, CERTs, and industry — are also involved. They can both benefit from the quality data sources for their research work, and contribute to improving transparent and open algorithms, encouraging scientists to work with and on them. Not least, they can offer additional sources of data to the clearinghouse.

The big picture

The tactics used in cyber-attacks change constantly, but the overall process and goals vary less. For instance, a DDoS attack can be conducted using a botnet, a collection of script kiddies, or through a reflection attack using misconfigured servers on the Internet. However, they all represent minor variations of the same basic approach — flooding connections with garbage data in order to stop a target from communicating. Most solutions are oriented either towards investigating each new variant of threat that emerges separately, or blacklisting some corrupted servers.

It is the root of the problem that should be addressed, however, because the root cause of attacks are vulnerabilities, both in the implementation and the specification of software. These vulnerabilities constantly change as new problems are discovered. It is impossible to simply patch ourselves into safety. Instead, CERTs, operators, and policymakers must address this as a problem of triage – what are the most impactful problems, and what are the most effective mechanisms to mitigate that impact? While there are numerous sources of information about vulnerabilities across the Internet, this data should be carefully selected, compiled, and synthesised in order to construct a reliable and trusted transparent image of the health of the Internet and its segments.

The clearinghouse establishes accurate and comprehensive data sources through continuously seeking to identify, recruit, and process the best possible globally available data sources to service the measurement of global cyber health and risk conditions. These sources should evolve as the understanding of cyber health improves and measurement advances. Being processed to raise transparency and comparability of data between various actors, they serve as a trusted base for mitigation.

The technical community, and particularly CERTs and operators of Autonomous Systems, acquires a tool that increases their capacity to identify risk conditions. If the clearinghouse 'engine' is open source, the knowledge of how to validate and process raw data to make it actionable and support effective decision-making can be improved.

Instructions

The clearinghouse develops and maintains quality data sets related to risk indicators. It should collect and process data.

Collecting data

Data should be collected from existing sources (academic and research projects, corporate initiatives) and from own network scanning.

Data collection methods must be transparent (which is not always the case with commercial sources). Elements for measuring the quality of the collected data should be developed and made public.

While network scanning can bring additional quality information, it is also a risky approach: if too many players are involved, it can make a network a very noisy environment. In addition, since scanning is done by the perpetrators as well, the scan targets may block the traffic if they start receiving too many scans. Several legitimate organisations involved in scanning networks already exist — though without much coordination; so, it may be better to use their results, or at least coordinate efforts with them.

Processing data

Processing the collected raw data may include

- Cleaning
- Matching
- De-biasing
- De-duplicating

Network security data, such as scan results, are highly time-sensitive. There is an enormous amount of transient activity on the Internet, and it is reasonable to believe that some fraction of true positives from scan results are invalid within hours of the scan's completion. Because of this, pure IP address information provides an illusion of precision.

Cleaning ensures a focus on relevant data. Matching ensures a more complete picture, as various sources report on different parts of the Internet at different times. De-biasing considers the fact that most commercial sources are biased in some way. De-duplicating removes duplicated data.

It is important that data processing and statistical engines be open source and available for free to security operations teams, so that others can replicate the platform, analyse, and develop statistics with their own sources, and contribute with possible improvements to the engine.

Timing

The timeline vastly depends on technical and operational capabilities and specific needs. A provisional timeline for developing a clearinghouse may be as follows:

Collecting data

Data, especially OSINT sharing is based on trust derived from security operations and collaboration. It might take some time to build trust, if the organisation is not in the trusted operations community already. This can take between six months and a year.

Processing data

The organisation needs to hire really good data scientists, who also have a cybersecurity background. These people are rare, and it is hard to recruit them, so establishing a team may take six months or more.

Example

The GFCE initiative CyberGreen makes the cyber-ecosystem healthier through measuring and visualising the state of the global cyber ecosystem, and producing materials for mitigating negative impacts. The clearinghouse is one of the main achievements of the initiative, along with metrics and visualisations, and support for mitigation.

CyberGreen works with data scientists and statisticians from multiple national CERTs as well as private industry. CyberGreen also works with Regional Internet Registries - RIRs (APNIC, LACNIC, RIPE, etc.), Regional CERTs (APCERT, TF-CSIRT, Africa-CERT, ITU-ARCC) for mitigation training and capacity building.

CyberGreen's current sponsors include JPCERT/CC, the Cyber Security Agency of Singapore, and the UK Foreign and Commonwealth Office. These and other policymakers benefit from having increased visibility of the risk levels that are present in their countries.

Source, support, and mentoring

CyberGreen statistics site: http://stats.cybergreen.net
Data sources catalogue by CyberGreen: http://www.cybergreen.net/data-inventory/
Bulk data (and API) of CyberGreen for download: http://stats.cybergreen.net/download/

Contact CyberGreen:
https://www.cybergreen.net/contact/

Contact point:
Yurie Ito (yito@cybergreen.net)