JPCERT CC®

Introduction to CyberGreen

January 26th, 2016

JPCERT Coordination Center

Global Coordination Division

Disclaimer

What will be described is merely an approach towards accomplishing our goal. We are not stating that this is the absolute correct way or that other similar approaches are wrong.

Opinions that will help the cause are more than welcome. In fact we want more ideas to think about.

High-level Concepts

- The internet as a whole is a globally shared ecosystem that needs to be protected by everybody

- ‘Cyber Health’
  - —Provides a conceptual basis for achieving cyber security
  - —(or the health status of the internet) is defined as its ability to perform as expected
  - —Has a direct impact on the security of the internet

A Healthy Internet

- Stable

- Meets our expectations of security, reliability
  —free of malware / botnet infections, etc.

- Can we measure this?

Key idea:
Internet works as we all expect

An Unhealthy Internet

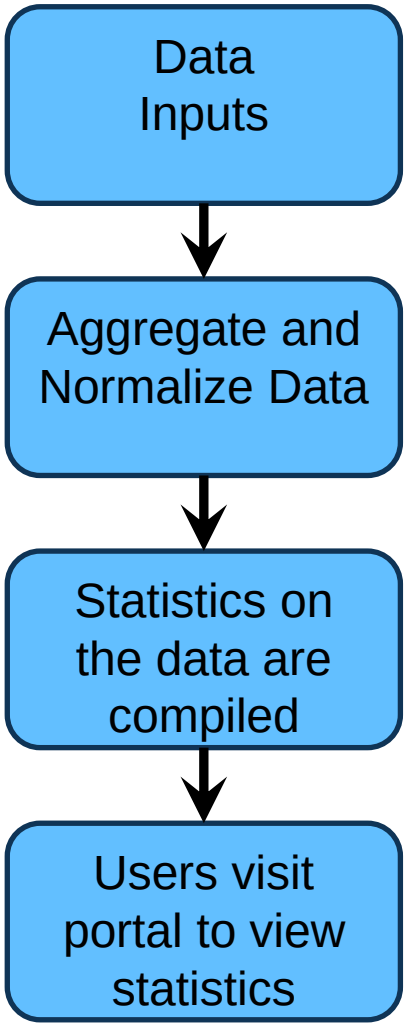
- Loss of data confidentiality, reliability, integrity
- Connections that do not meet expectations
- “People are afraid to use it” for various reasons

Motivations for a New Initiative

- Increasing dominance of cyber in communications, business
- Increase in cyber risks and awareness
- Instability is a likely result with far reaching impacts

Overview of CyberGreen

- Overall (simplified) flow of CyberGreen



- Data inputs from various sources
- Green Index and other statistics are calculated
- Users can visit the portal to view the compiled statistics

** More information on portal to follow **

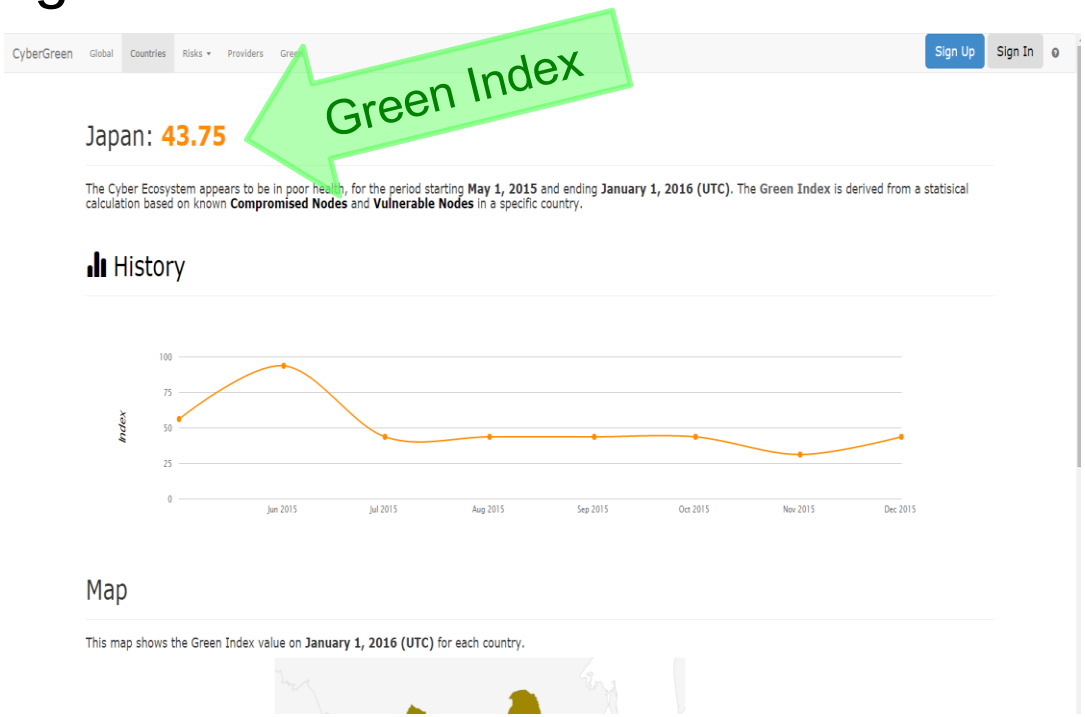
So..what is CyberGreen?

- Not an entirely new approach, but a combination of approaches from previous projects
- It gathers and aggregates various risk condition data
- **It provides an indication of whether we are getting better, or worse over time**
- Assist CSIRTs in obtaining and leveraging datasets for various CSIRT activities

What is the Green Index – Overview

- Represents if a region is becoming “more healthy” or “less healthy”
  - A number between 0 (getting worse) and 100 (getting better)

- Calculated using 2 categories of data
  - # of Compromised Nodes, such as zombie-PCs
  - # of Vulnerable Nodes, such as open resolvers



What can I get out of it?

- A holistic view of the overall health of the internet based on the gathered data

- Directly compare performance based off of the Green Index
  - —More on Green Index a little later
  - —For now, just think of it as a number between 0 and 100 (closer to 100 being ‘better’)

- **A way to “identify spots where we’re doing good work”, and to “identify spots where work needs to be done”**
  - —Looking to emphasize this going forward
  - —Details on this a little later…

CyberGreen Indicators (Data Categories)

- Compromised Nodes
  - —Knowing or unknowingly hosts harmful contents
  - —Intent to send unwanted traffic for probing or further compromise other Internet infrastructure
  - —E.g. – Phishing sites, Bot infected systems, etc

- Vulnerable Nodes
  - —Misconfigured or un-patched network infrastructure
  - —Currently focus on outward facing infrastructure (services accessible from the Internet – ‘scannable stuff’)
  - —This data is more expensive to obtain
  - —E.g. – Misconfigured DNS, UPnP, etc.

Green Index in a little more detail

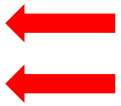
- Duplicates across data sources are taken out when ingesting the data
  - —Next slide shows an example of de-depulication

- Percentile Rank is used as the basis for calculating the Green Index
  - —It is the percentage of scores that are equal to or less than a given score
  - —The number of nodes over a given time period is used as the “score” here

Data Merge Example

- 3 unique phishing URLs from 2 sources

```
source, URL, server IP address, timestamp (GMT)
Src1, http://paypal.fakesite.com/, 1.2.3.4, 17:28:34
Src2, http://paypal.fakesite.com/, 1.2.3.4, 19:32:18
Src1, http://amazon.fakesite.co.jp/, 5.6.7.7, 21:23:45
Src1, http://amazon.fakesite.co.jp/, 5.6.7.8, 21:23:45
```



Green Index in a little more detail

- Green Index is a number representation of ‘current health status’ in comparison to its ‘past health status’ over the given time period
  - —A high Green Index shows that the node counts are smaller than the past
  - —A lower Green Index shows that the node counts are larger than the past

- When aggregating, “distinct IP addresses” are counted over a time period (day, week, month, etc.)
  - —Counts for each block of time are taken and compared with each other to calculate the percentile rank

What does the Green Index mean?

- A simple analogy:
  - —Green Index is to Internet Health as Body Temperature is to the Human Body
  - —We do not compare our body temperatures with other people
  - —It is a barometer of “our” health

- So, we are comparing current performance (in node counts) against past performance
  - —A Green Index calculation for a region does not care about the number of nodes in a different region
  - —“My temperature is a little high / low today”
  - —The same as saying, “there were more (less) compromised / vulnerable” nodes observed over a given time period

What does the Green Index mean?

- One difference…
  - —Normally, you compare your body temperature to what your average body temperature is to determine if you have a fever or not
  - —Right now, the “average” Green Index is around 50.
  - —However, this 50 is not a good representation of being “average” in performance
  - —Looking into a better benchmark for average performance

- Underlying factors that may adversely affect our calculation:
  - —Adding more nodes or data sets
  - —A strategic change aimed to improve cyber health down the road

Green Index Calculation

- Details on the methodology behind the Green Index Calculation is provided on the portal https://stats.cybergreen.net/green

The Math

Green Index *G*

$$G = 1 - \left(\frac{(C + V)}{2}\right)$$

Where
C = PR of 'Compromised Nodes' daily count
V = PR of 'Vulnerable Nodes' daily count

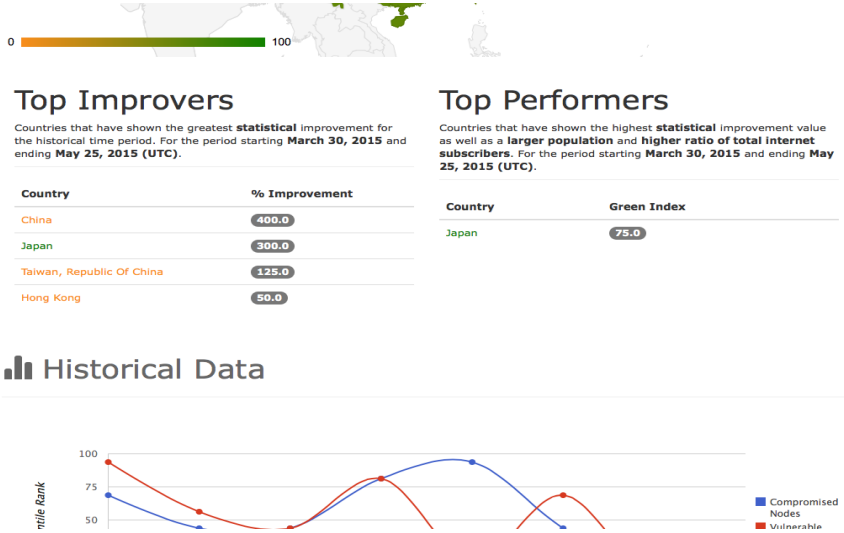
Percent Rank *PR*

$$\mathrm{PR} = \frac{L + (0.5 \times S)}{N}$$

Where
L = Number of below rank
S = Number of same rank
N = Total numbers.

Where can I find the Index?

- Located at https://stats.cybergreen.net/
  - —You'll be redirected to the “region" page based on where you are located
  - —You can read Green Index of your region
- Map of your region and neighbors
  - —the more green, the better Green Index
- Top Improvers
- Top Performers
- Historical Data Graphs



What does the Green Index mean?

Recap:
Green Index is a number that indicates whether a region is becoming “more healthy” or “less healthy” over time

- It helps represent the cyclical nature of cybersecurity volatility (turbulence). Shows “where you are in the cycle, are you getting better or worse?”
- If a region is continuously becoming less healthy over time, something needs to change
- If a region is continuously getting better over time, we need to learn from what they’re doing “right”

What does this “statistic” tell us

- If Continuously Getting Better…
  - — Continue what you’re doing
  - — Increase resourcing to shorten the volatility cycle
  - — Increase efficiency
  - — Etc.

- If Continuously Getting Worse…ask yourself..
  - — Do you need more policy?
  - — Do you need more resources?
  - — Do you need tools / training?
  - — Etc.

Deeper dive into the Index (currently in the works)

- Green Index “lumped” collected data into 2 buckets
  - Good start, but need more break down to better see “what needs to be remediated”

- Focus on collecting / normalizing aggregated ASN statistics per threat classification, examples:
  - OpenNTP servers per ASN
  - Open Resolvers per ASN
  - Nodes that send spam email per ASN
  - Nodes seen performing brute force SSH attacks per ASN
  - Etc.

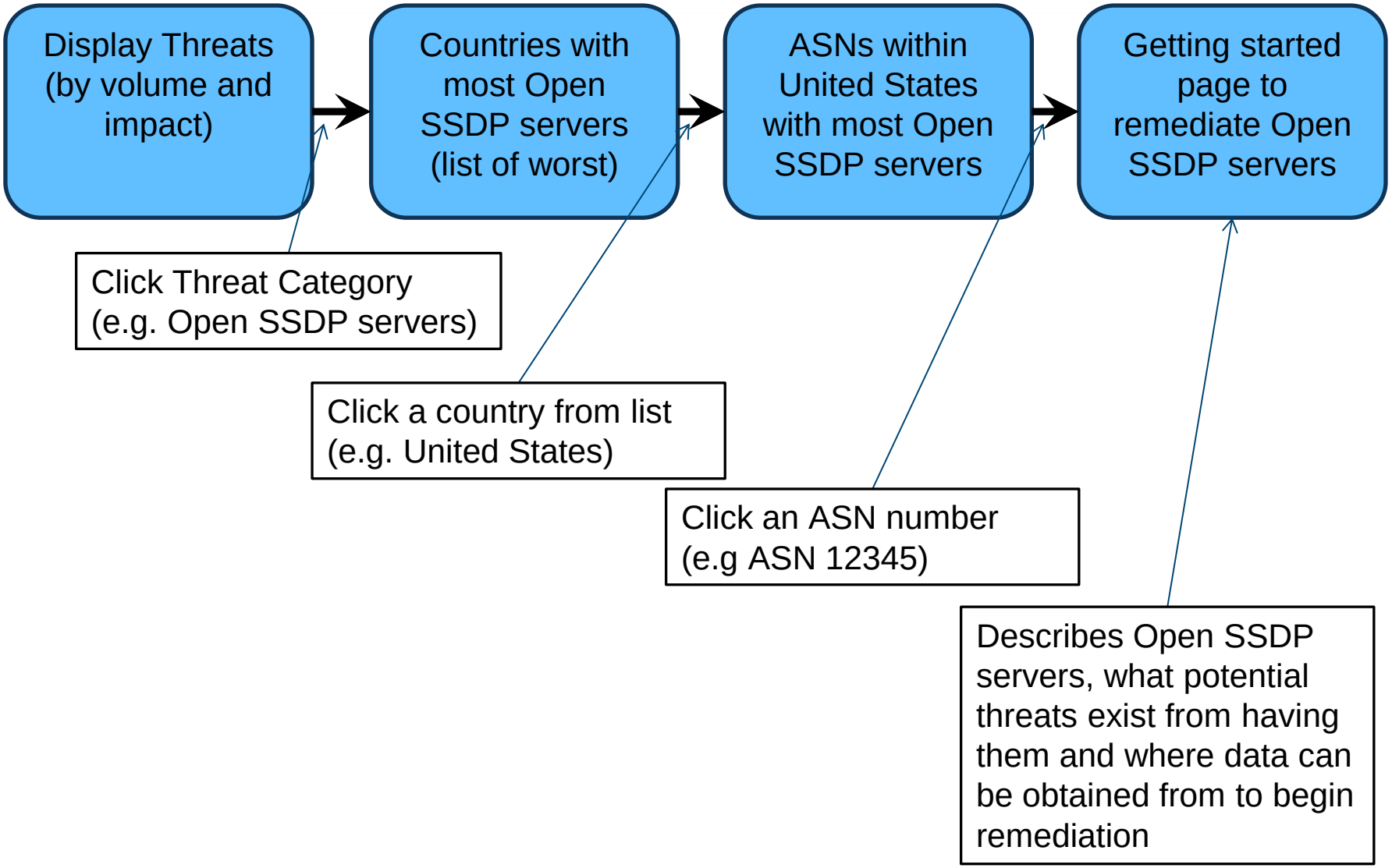
- With lots of data, it can be easy to overwhelm users with lots of information

Deeper dive into the Index (currently in the works)

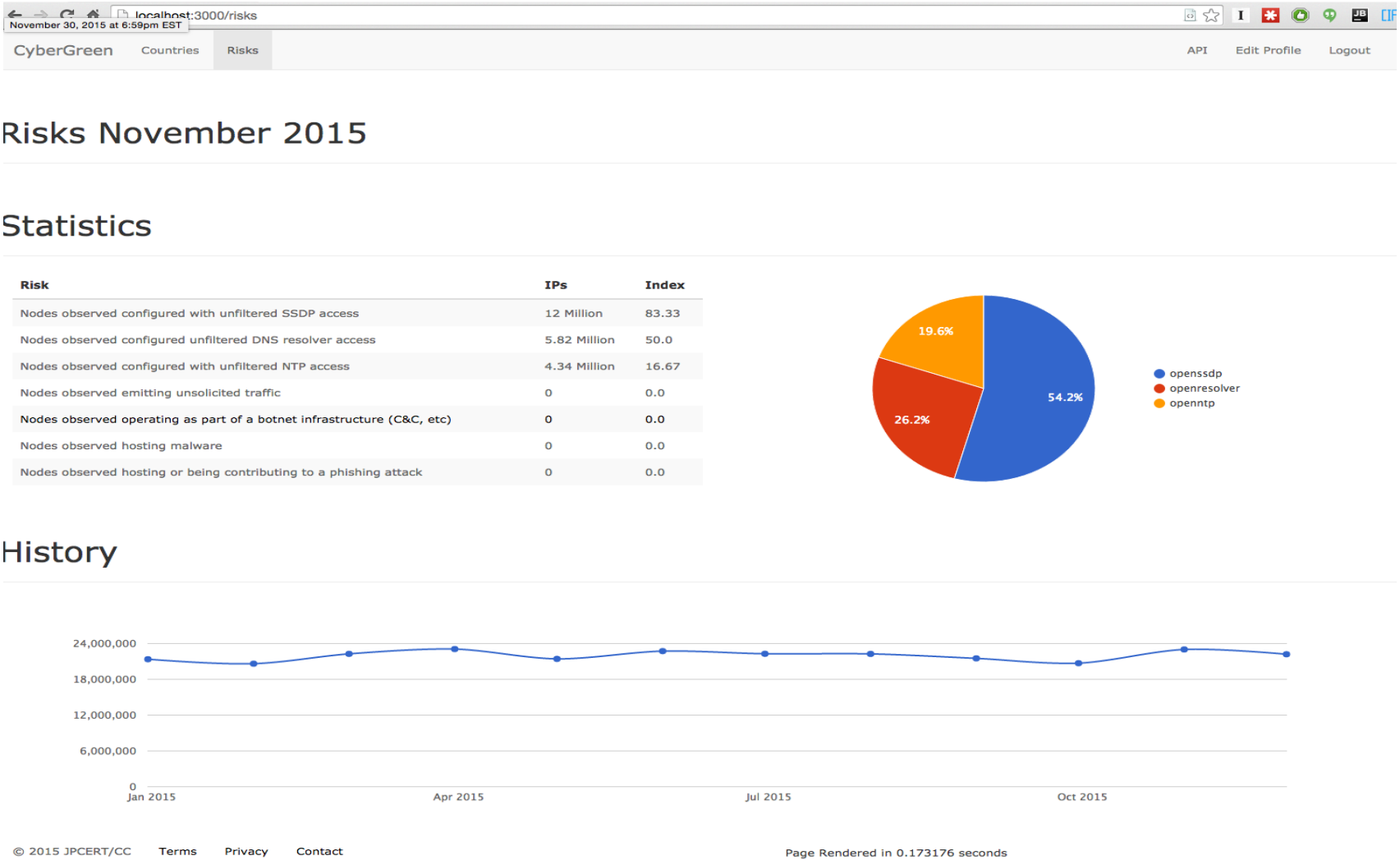
- Start simple, provide ability to add more as users need
  - Display less information to start by default
  - For example, start with total count of “infected” machines in a region or ASN
  - Total count is an easy to comprehend metric
  - Provide users ability to display additional statistics as they see fit

- Make it as easy as possible to get to “what needs to be remediated on my network” (as few clicks as possible)

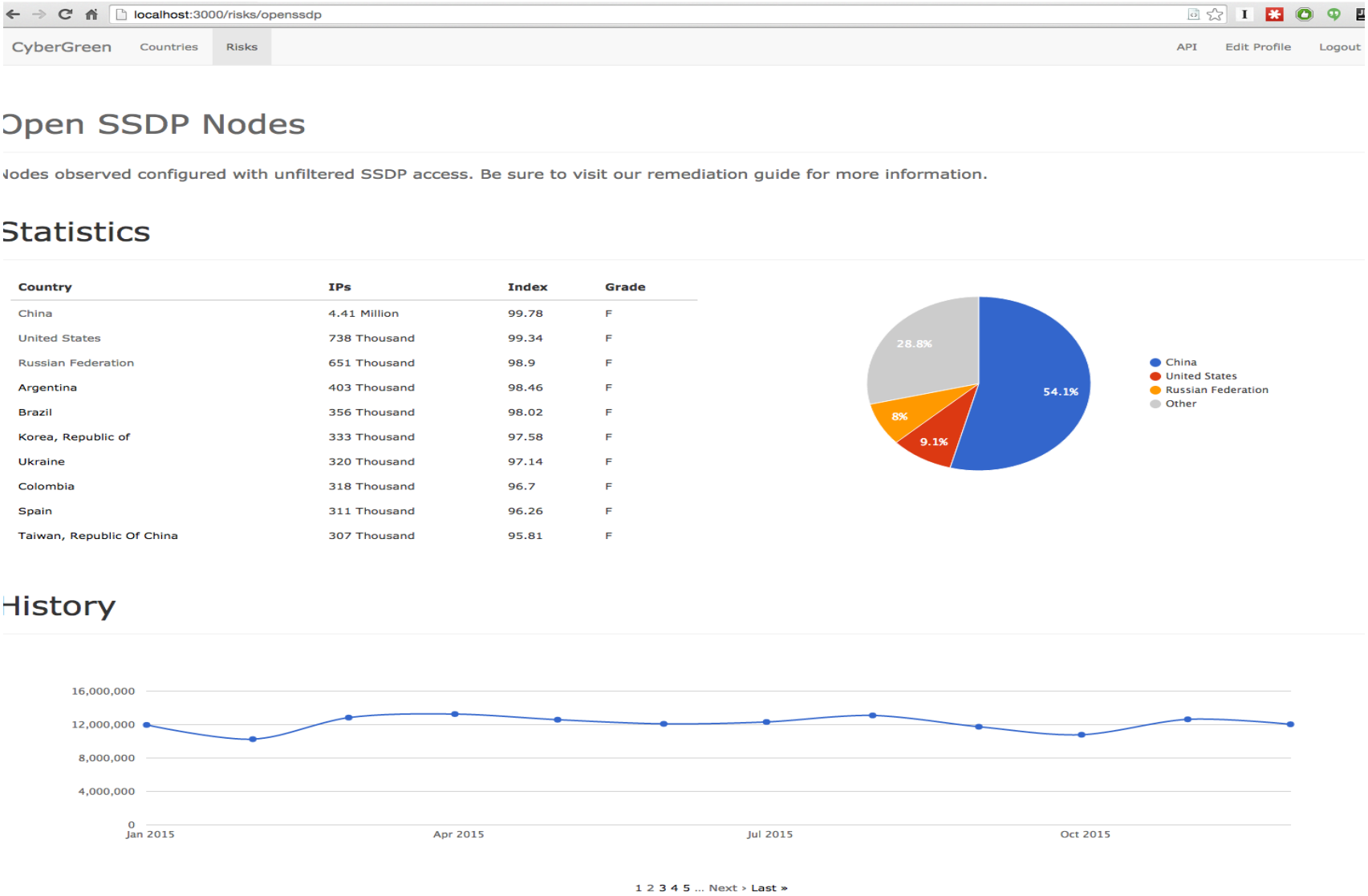
Deeper dive into the Index (currently in the works)



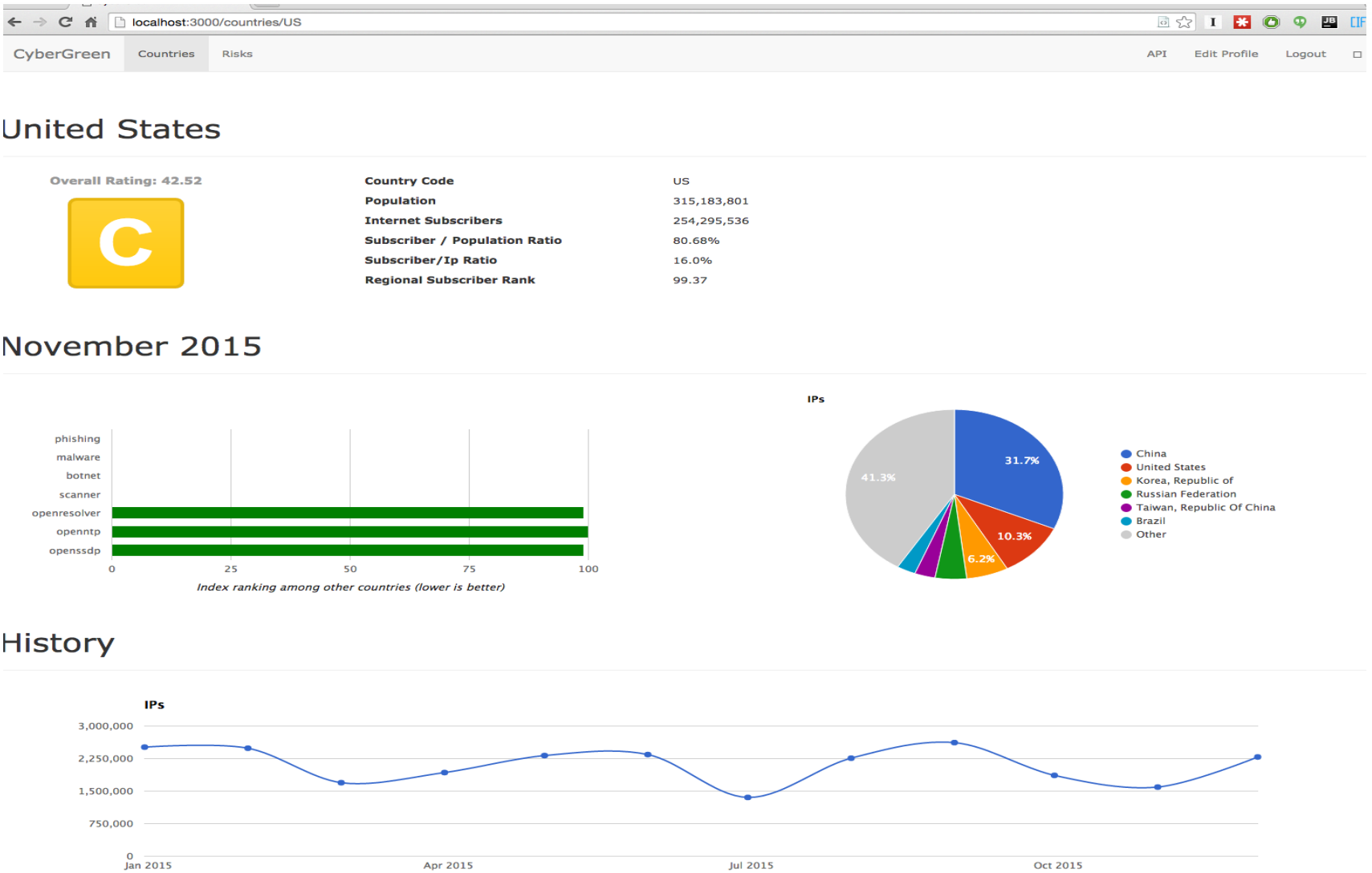
Changes to the Portal (currently in the works)



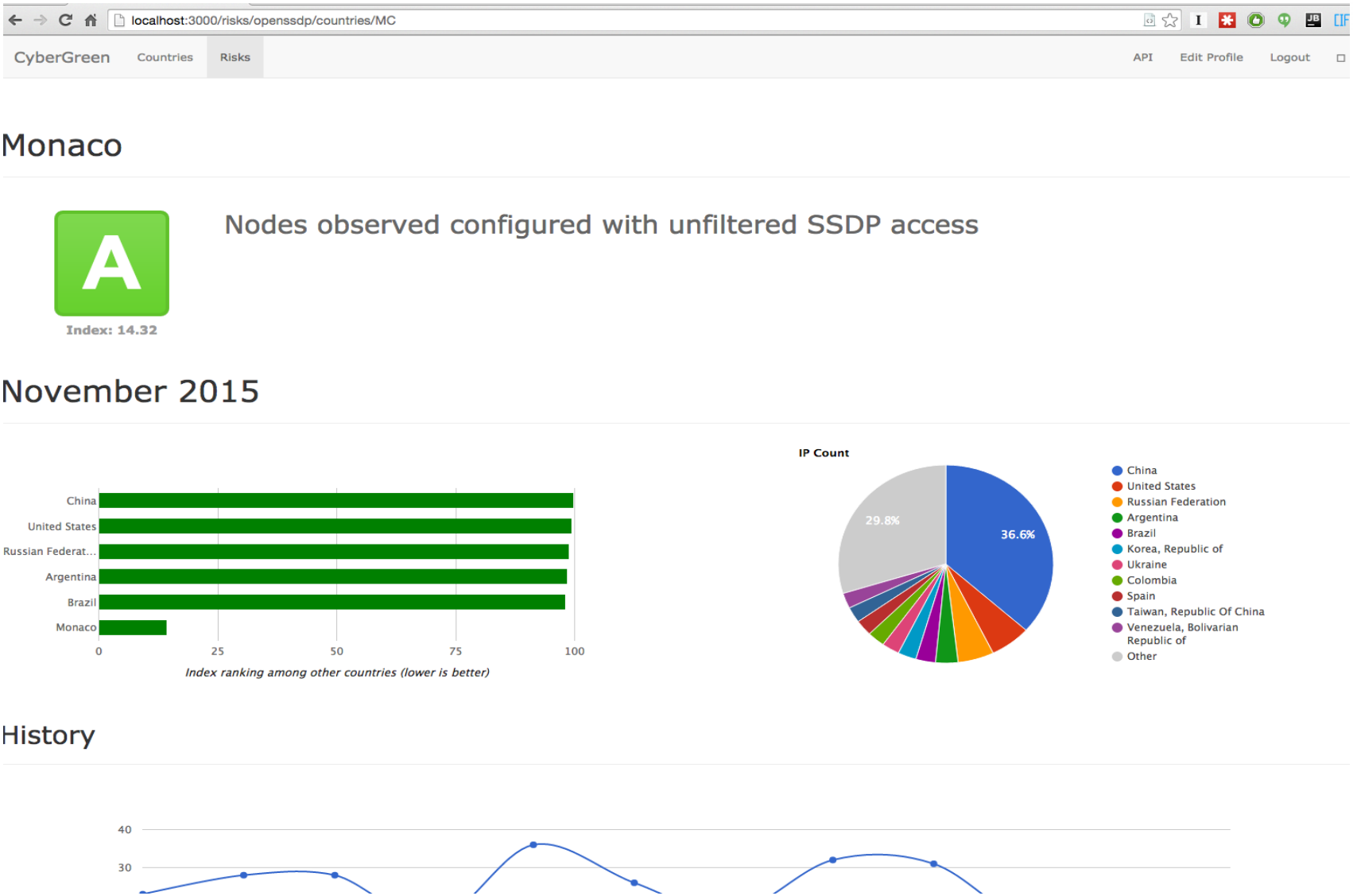
Changes to the Portal (currently in the works)



Changes to the Portal (currently in the works)



Changes to the Portal (currently in the works)



Changes to the Portal (currently in the works)

- Changes aimed towards more transparency
  - —Show whatever we can, without overwhelming users with "too much, too soon"
- Improvement on visualization of monthly trends
  - —Overall
  - —Per region
  - —By threat type
  - —Cross-compare with other regions
- Look into a calculation that "weighs" the index to better reflect the ranking
  - —enhancement or addition to the percentile rank calculation

Changes to the Portal (currently in the works)

- Working on a way to show “not all threats are equal”
  - —Weighing threats (e.g. Open Resolver >> Conficker)
  - —Want to prioritize in addressing threats that can cause significant damage to the internet

- Vetting function, that allows certain parties (ASN owners, national CSIRTs) to access more detailed statistics and data

- Considering “weighing data source” option
  - —Allows users to trust certain data sources over others
  - —CyberGreen will provide defaults in this case

Analyzing the Data (Future Works)

- Some questions that we would like to answer…
  - —Is there a correlation between number of incidents in a region and how well / not well an economy is doing?
    - How does addressing an incident affect the economy?
  - —How much does an incident ‘cost’?
    - In terms of economy
    - In terms of man/hours
  - —Is there a correlation between growth in internet subscriptions and number of incidents?
    - Compare with population / IP address counts also?
  - —How many people (in a region) are actually affected when something goes wrong?
  - —What is the average lifespan of a compromise?

Economic Factors

- Use GDP?
  - —OECD?
  - —World Bank?
  - —Combination?
  - —Other?
- Use Stock Market changes?
  - —Reuters Global Market Data?
  - —What to do with regions with small / no markets?
- Any other economic factors that can be used?
  - —Only requirement is that it can be applied to all regions

Population Factors

- Internet Users data
  - —Number of internet subscriptions
  - —Ratio of internet subscriptions to population
  - —Ratio of internet subscriptions to number of IP addresses
  - —Rates of change of the above

Anything else?

- Any hypotheses that you would like us to test out?
  - —Can be anything
  - —The hypothesis does not have to be correct
  - —Just trying to allow the data “to do the talking”

Moving Forward

- Not a ‘static’ project
  - — Any comments / feedback can be pushed into an update fairly quickly
- Remediation routing features in the portal
  - — Upcoming addition within the next few weeks
  - — A way for CSIRTs to be ‘routed’ to the data source
- Best practices catalog
  - — For CSIRTs looking for a starting point / a little help on something that they are stuck on, etc.
- Statistical expert group to discuss metrics (started from July 2015)
  - — Updates / information will be blogged

- The CyberGreen Project appreciates your participation.
  - mailto: contact@cybergreen.net
  - https://stats.cybergreen.net/
- Thank you!!