The Global Cyber-Vulnerability Report

V.S. Subrahmanian

University Of Maryland

vs@cs.umd.edu @vssubrah

Parts of this talk are joint work with

Michael Ovelgonne, Tudor Dumitras, Aditya Prakash, Chanhyun Kang, Noseong Park, Edoardo Serra

UMIACS
University of Maryland Institute for Advanced Computer Studies

CDIG CENTER for DIGITAL INTERNATIONAL GOVERNMENT

Key Question

Can we use real-world data in order to:

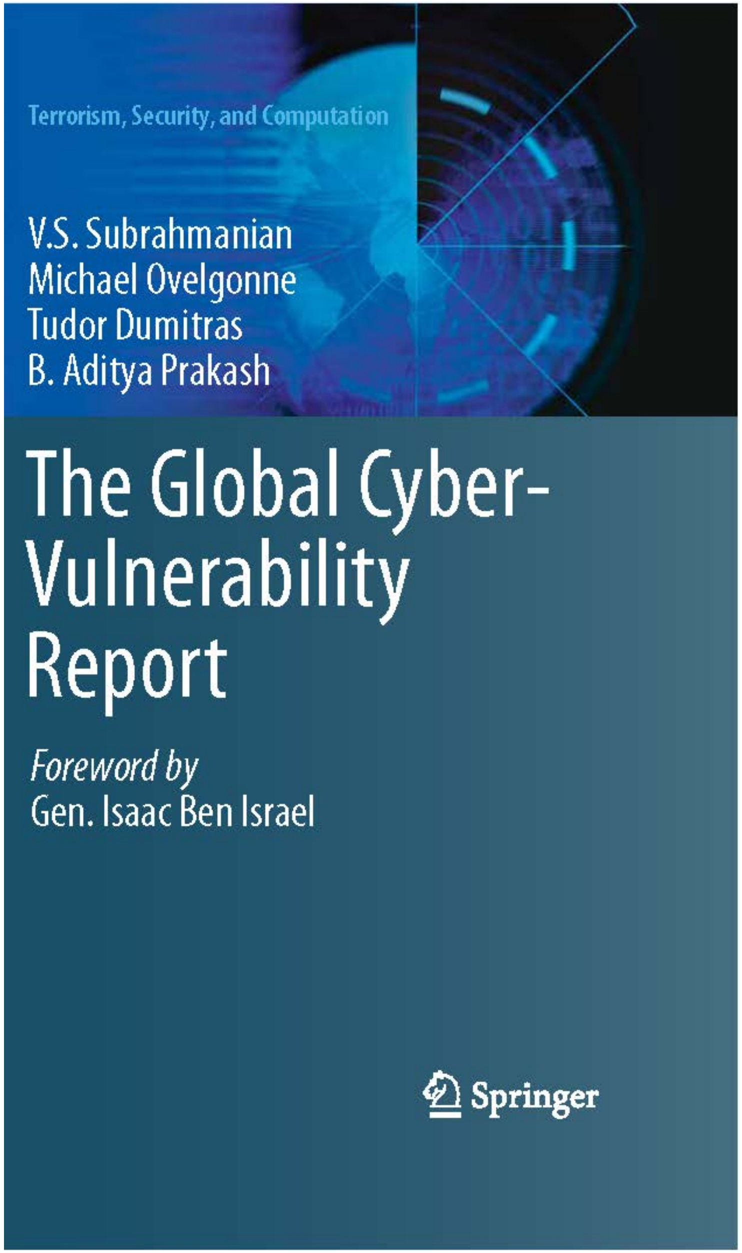
- **quantify the cyber-vulnerability of consumers in different countries?**
- **help governments/companies better ensure safer user behavior**
- **Forecast how different countries will react to highly infectious malware?**

YES.

- Our study:
  - Two years of data, Sep 2009-Aug 2011.
  - Data provided by Symantec's WINE system. **Thanks Symantec !**
  - All data collected on an "opt-in" basis
  - All data was anonymized

Reference

http://www.springer.com/us/book/9783319257587



Terrorism, Security, and Computation

V.S. Subrahmanian
Michael Ovelgonne
Tudor Dumitras
B. Aditya Prakash

The Global Cyber-Vulnerability Report

Foreword by
Gen. Isaac Ben Israel

Springer

Study Methodology

- **Human Vulnerability Study (HVS)**
  - Identify behaviors of users that are correlated with the number of attacks on those users.
- Country Vulnerability Study (CVS)
  - Uses the results of the previous sub-study in order to identify the vulnerability of 44 countries to cyber-attack
- Country Forecast Study (CFS)
  - Uses the raw Symantec data to identify how to predict extent of malware spread in 40 countries.
- Recommendations

Data-set for Human Vulnerability Study

- Detailed study of 8 months of data from Symantec's WINE data set
- Study involved over 3.7M machines with no sampling. However, we excluded machines for which there was less than 200 days of data.
- The resulting data set included
  - 1.6M machines
  - Over 13.7B malware/telemetry reports.

M. Ovelgonne, T. Dumitras, B.A. Prakash, V.S. Subrahmanian, B. Wang. Understanding the Relationship between Human Behavior and Susceptibility to Cyber-Attacks: A Data-Driven Approach, *ACM Transactions on Intelligent Systems & Technology, accepted, to appear.*

WINE Data

- **Binary Reputation Data**
  - Records all binaries (i.e. executable software), malicious and benign, that were present on a host
    - Time stamp of file creation
    - Country in which host is located
    - MD5 or SHA2 hash of the binary
    - URL from where downloaded if applicable
- **Anti-Virus Telemetry**
  - Records malware detected by Norton A/V
    - Detection time
    - Threat Label
    - MD5 or SHA2 hash of the binary
    - Manner of detection (e.g. signature scanning, behavior observation after execution)
  - High confidence in accuracy as Norton A/V shoots for low false positive rates

Cyber-Hygiene: Host Behavior and Vulnerability

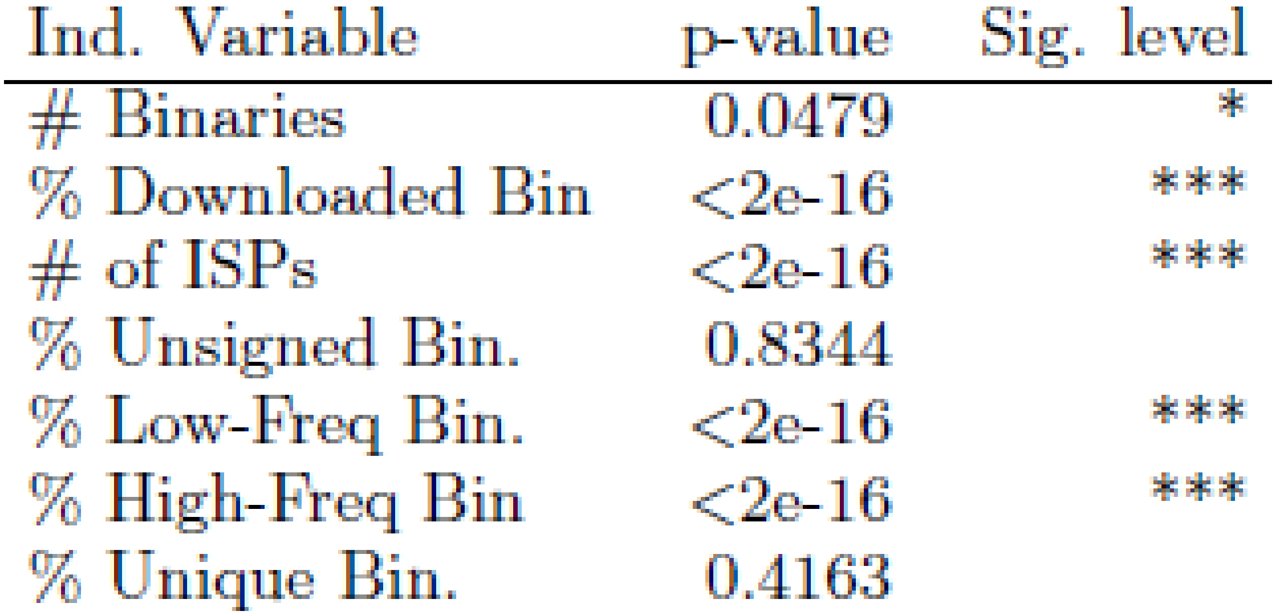
Hosts classified into: gamers, pros, s/w developers, and other.

Some machines fit multiple categories.

Hypotheses studied: Which of these is linked to number of attacks on a machine?

- **number of binaries downloaded**
- **number of unsigned binaries**
- **number of low-frequency (rare) binaries**
- **User travel history** (measured by number of ISPs the host connects from)
- **time of day** when the user logs on (day, evening, night)

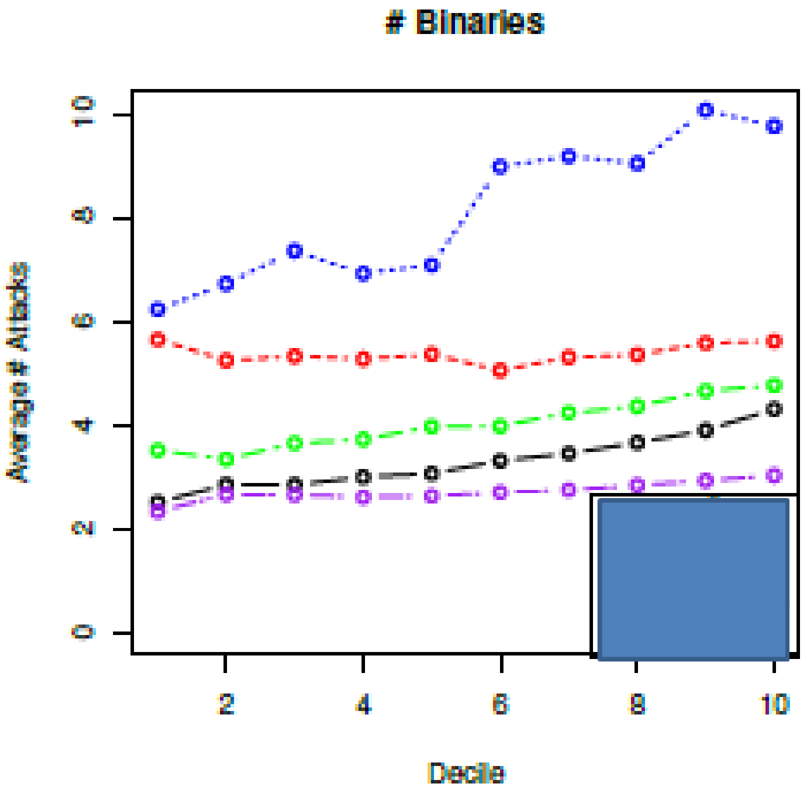
Summary of results



Ind. Variable	p-value	Sig. level
# Binaries	0.0479	*
% Downloaded Bin	<2e-16	***
# of ISPs	<2e-16	***
% Unsigned Bin.	0.8344	
% Low-Freq Bin.	<2e-16	***
% High-Freq Bin	<2e-16	***
% Unique Bin.	0.4163	

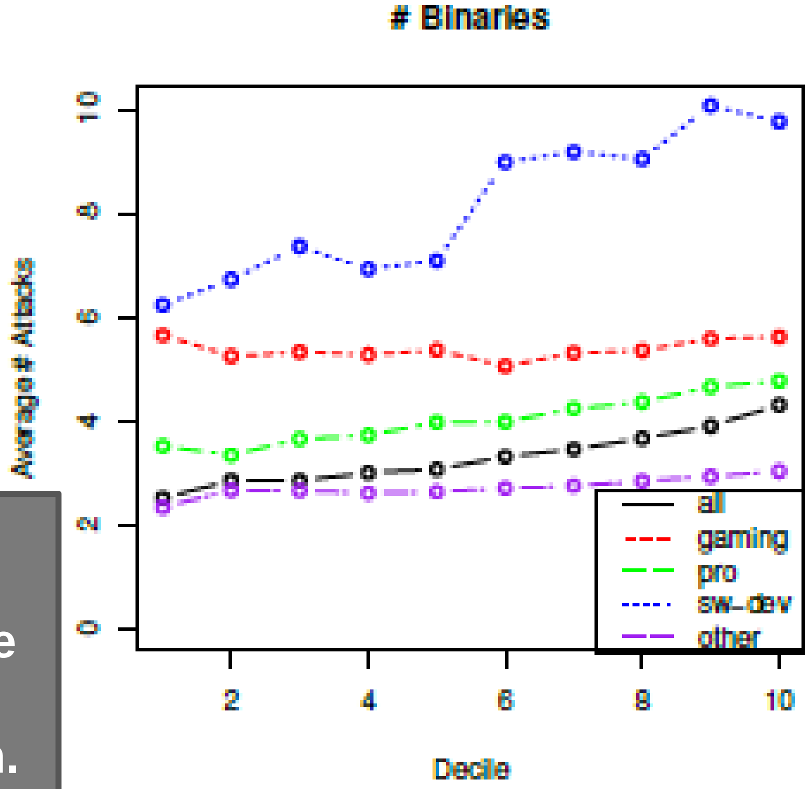
- **Multivariate Poisson regression model shows the important features are:**
  - Percentage of downloaded binaries,
  - Number of ISPs from which user connects,
  - Percentage of low vs. high frequency binaries.

Number of binaries vs. number of infections per host



Curves show all 5 categories: ALL, gamers, pros, s/w developers, others. Can you guess which is which?

Number of binaries vs. number of infections per m/c



Software developers are the most vulnerable (8.1 vs. 3.3), even after discounting for the fact that many binaries may have been produced by them. *All results are statistically significant with p < 0.001* (i.e. with > 99.9% confidence)

Study Methodology

- Two “sub” studies
- Human Vulnerability Study (HVS)
  - Identify behaviors of users that are correlated with the number of attacks on those users.
- **Country Vulnerability Study (CVS)**
  - Uses the results of the previous sub-study in order to identify the vulnerability of countries to cyber-attack
    - Aggregate Statistics
    - Behavior-based Statistics
- Country Forecast Study (CFS)
  - Uses the raw Symantec data to identify how to predict extent of malware spread in 40 countries.
- Recommendations

Data Preparation for CVS

- Number of hosts: 4.23M (2010) and 4.14 (2011)
- Removed all hosts with < 100 days of data.
- Total number of malware/telemetry reports exceeded 20B.
- In-depth studies of the 44 countries for which we had data on at least 500 hosts per year.
- Looked at two dependent variables:
  - % of machines infected in any given country [not discussing this today – similar results to those below]
  - Number of infections per machine in any given country

Question for the audience

- **Which 5 countries have the highest rate of attacks per machine?**

Question for the audience

- **Which 5 countries have the highest rate of attacks per machine?**
  - India
  - S. Korea
  - Saudi Arabia
  - China
  - Malaysia/Russia

More or less tied for most attacks/machine

Risk to India might be greater than for South Korea. Why?

i) S. Korea has an adversary, N. Korea, with significant cyber-capabilities. India's adversary, Pakistan, has not yet reached that point.

ii) N. Korea transferred nuclear weapons to Pakistan. If they transfer cyber-technology, it would pose a major risk, e.g. *DarkSeoul* attacks on banks, TV stations.

iii) India must reach agreements with states interested in containing cyber weapons.

Question for the audience

- **Which 5 countries have the lowest rate of attacks per machine?**

Question for the audience

- **Which 5 countries have the lowest rate of attacks per machine?**
  - Norway
  - Denmark
  - Finland
  - Sweden
  - Switzerland/Japan

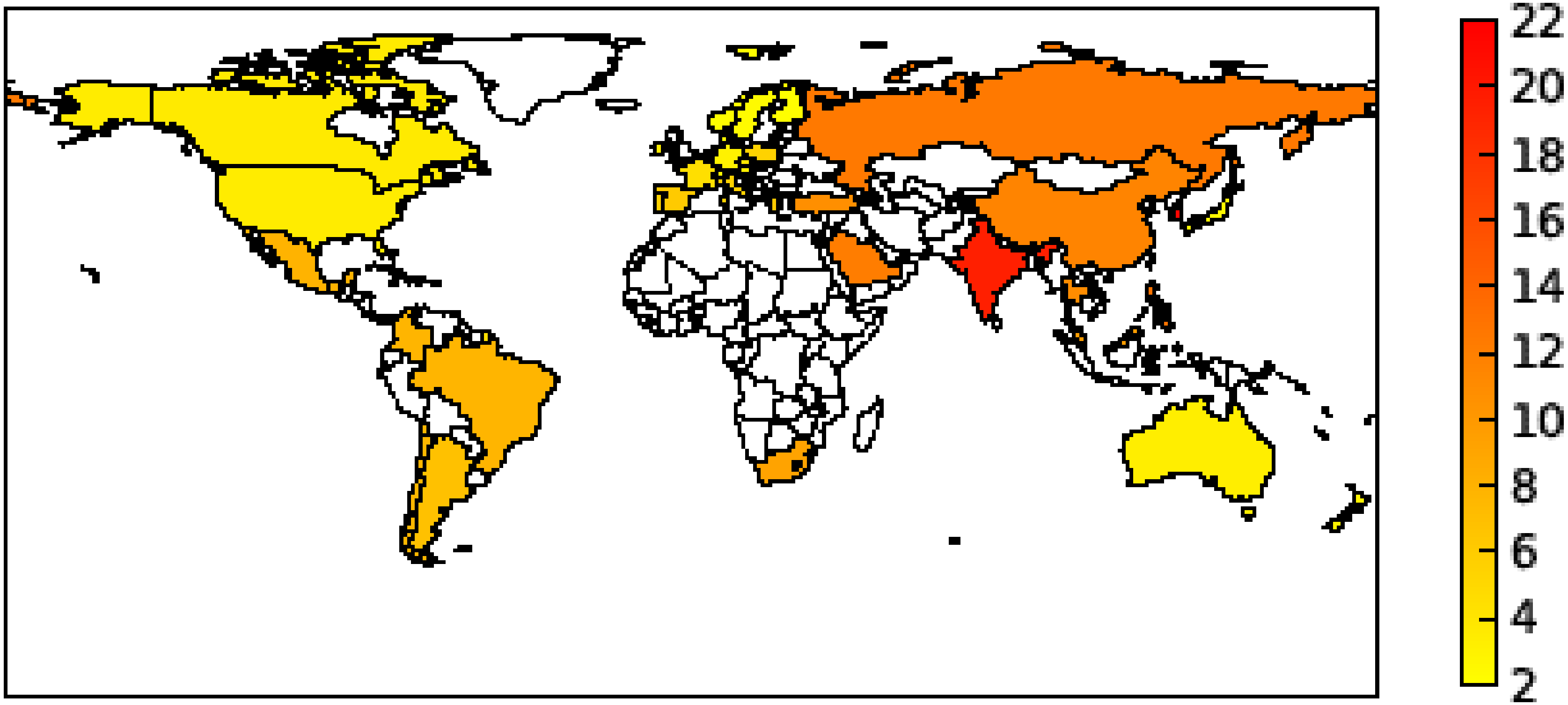
More or less tied for fewest attacks/machine

Cyber-Security Collaboration Opportunities

i) Cyber-defense: Japan, a strong ally of India, has considerable cyber-defense capability, and strong shared security concerns.

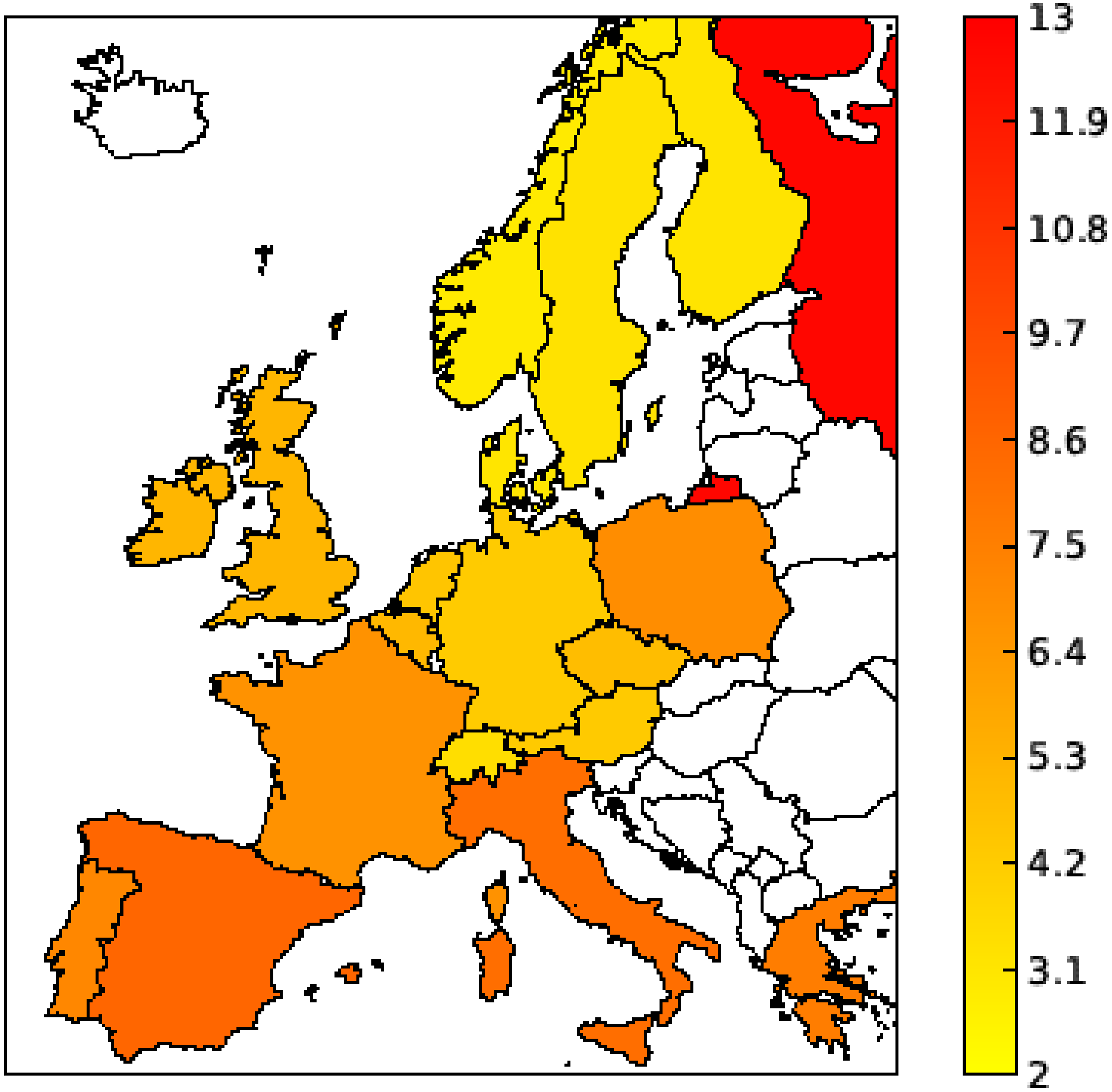
ii) Cyber-offense: *Hard to identify countries with the best cyber-offense capabilities*. Examples include Israel, UK, and USA.

World Cyber-Vulnerability Map



Feb 2016, @vssubrah

Europe Cyber-Vulnerability Map

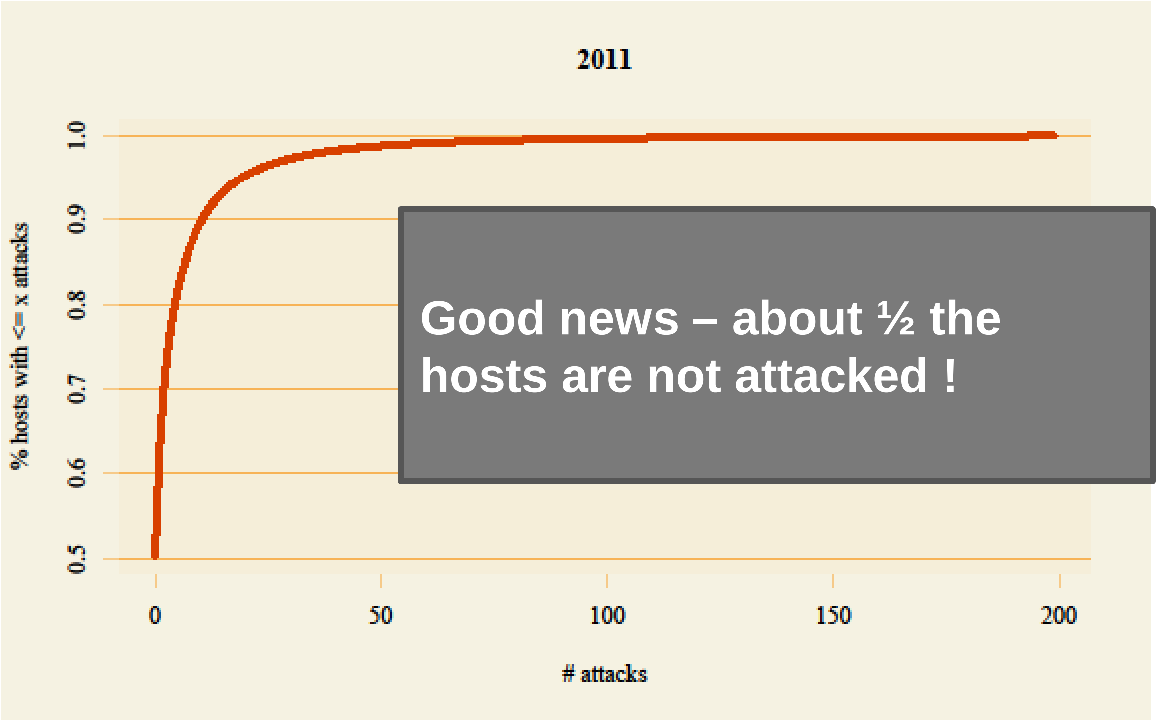


Average Number of Attacks per host

- **What is the average number of attacks per host?**

Average Number of Attacks per host

- **What is the average number of attacks per host?**



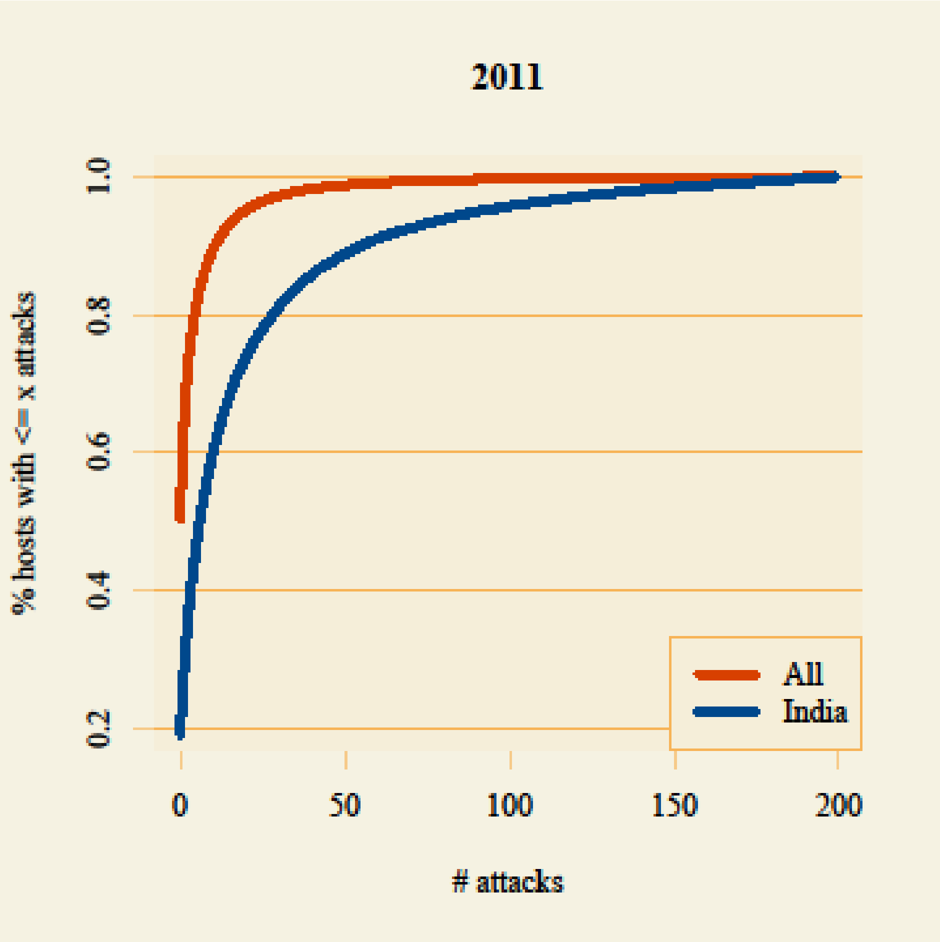
Good news – about ½ the hosts are not attacked !

Average Number of Attacks per host

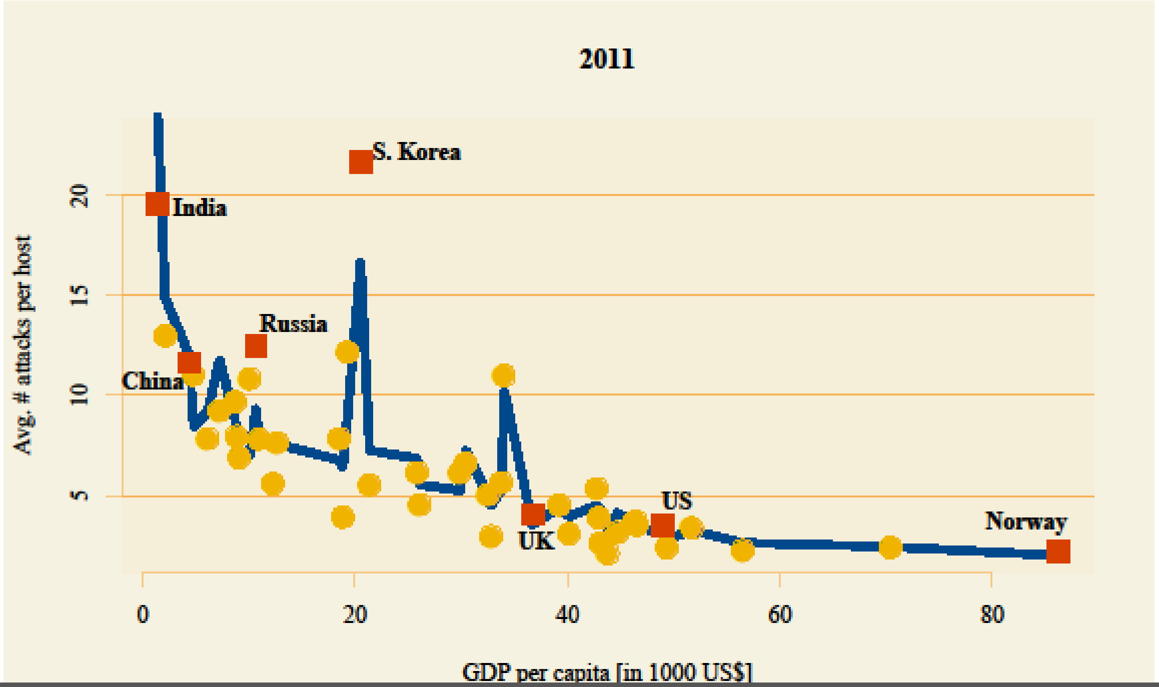
- **But in India…..**

Number of Indian hosts with no attacks is about 20%.

So about 80% of Indian hosts are attacked.

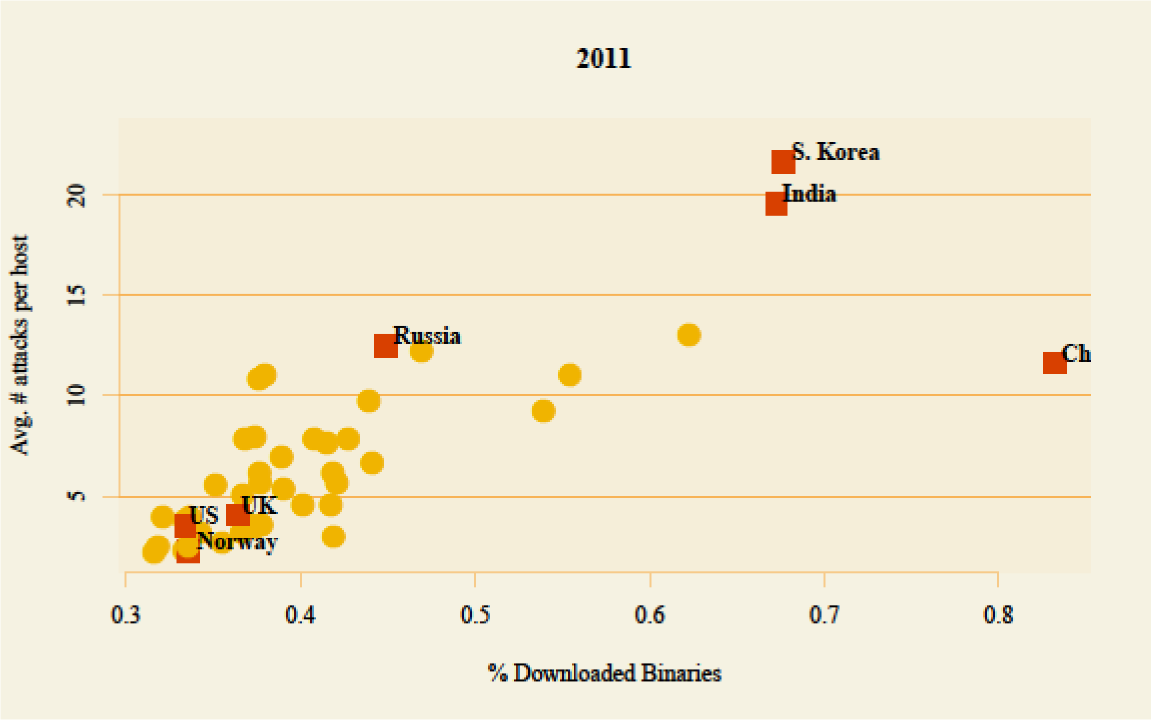


Per Capita GDP and attacks



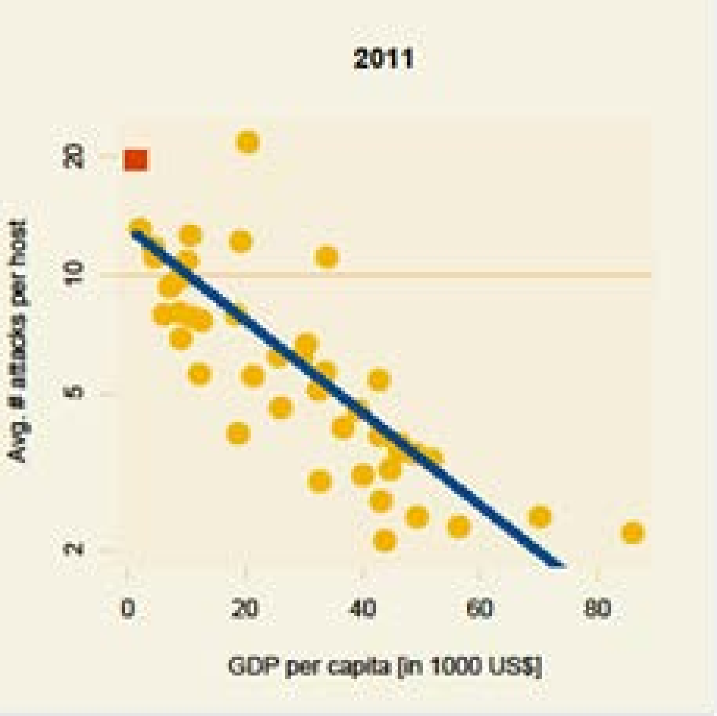
In both 2010 and 2011, we saw that increased per capita GDP was generally associated with a decreased number of attacks per host. In 2011, there is a Pearson correlation coefficient of -0.83 for this !

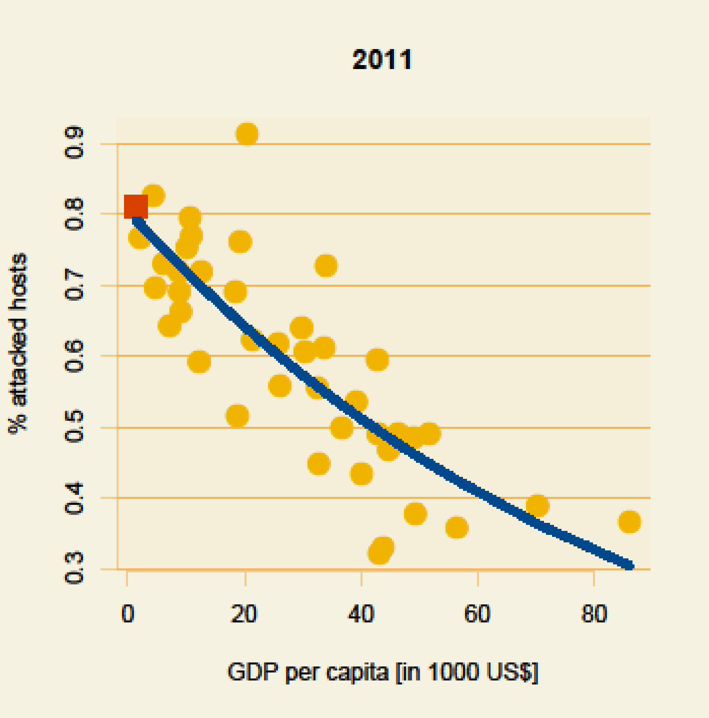
Downloaded binaries and risk



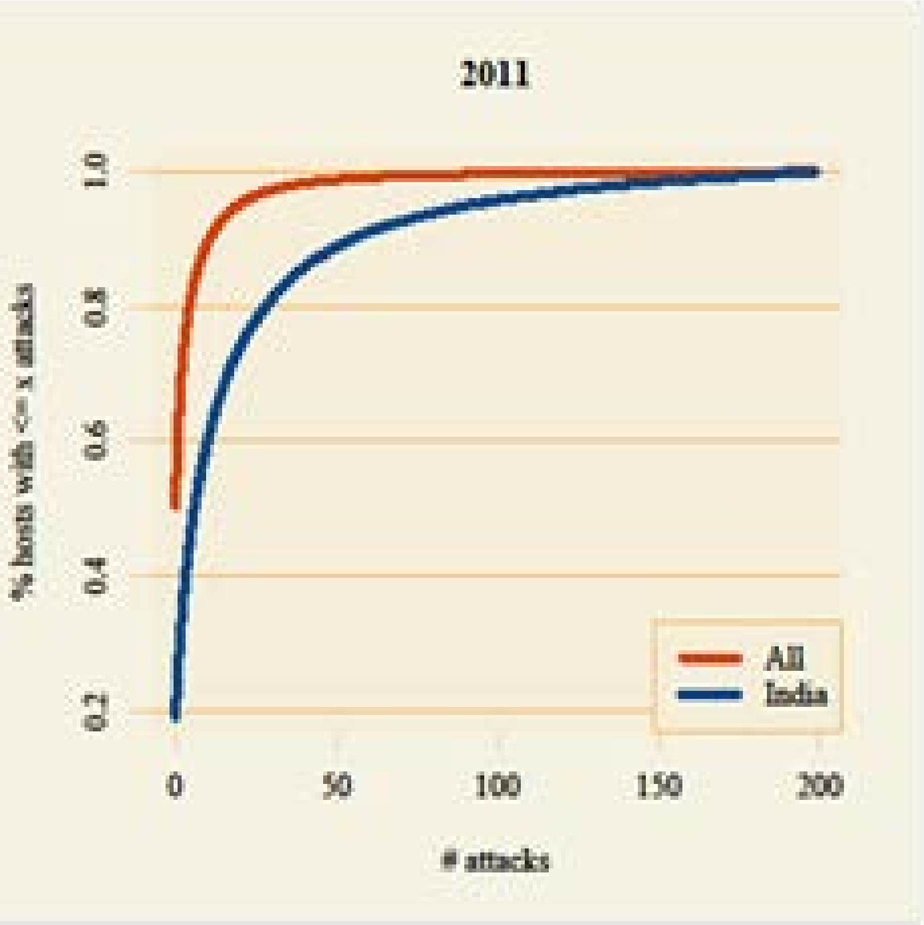
As the number of downloaded binaries increases on a host, the number of attacks on that host also increases. But there are outliers, e.g. China.

The Case of India: Attacks per Host





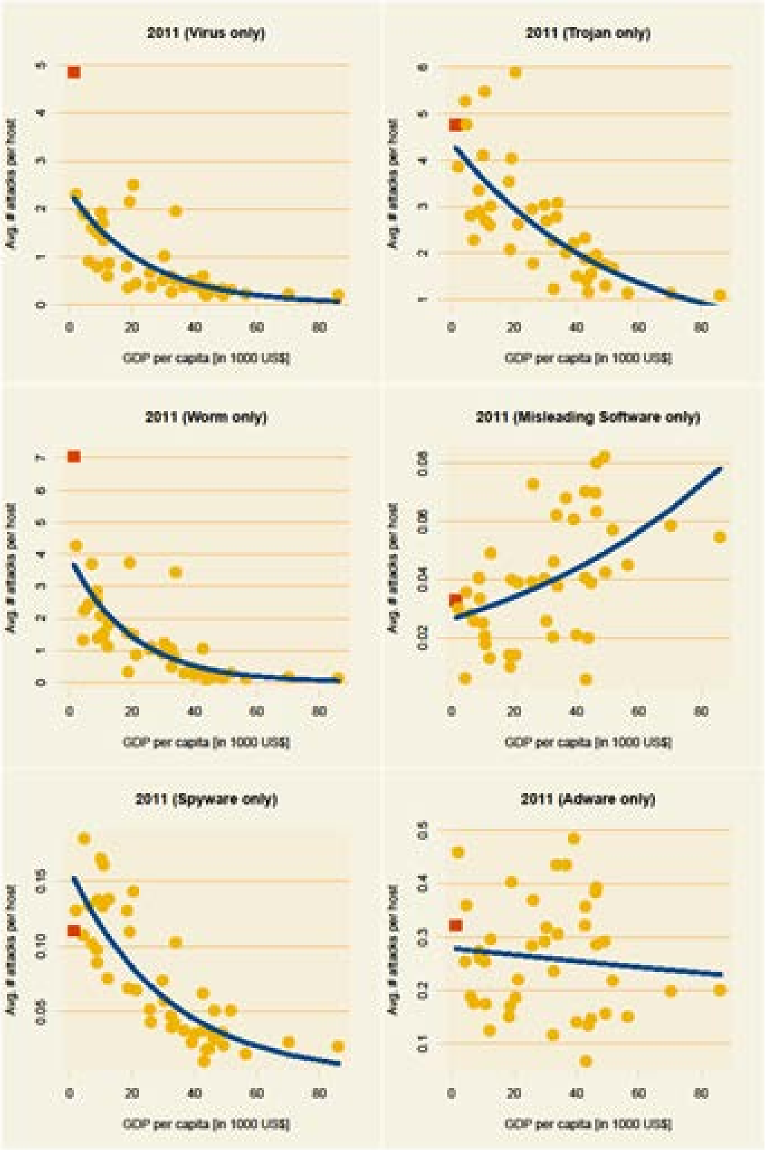
The Case of India: Attacks per Host



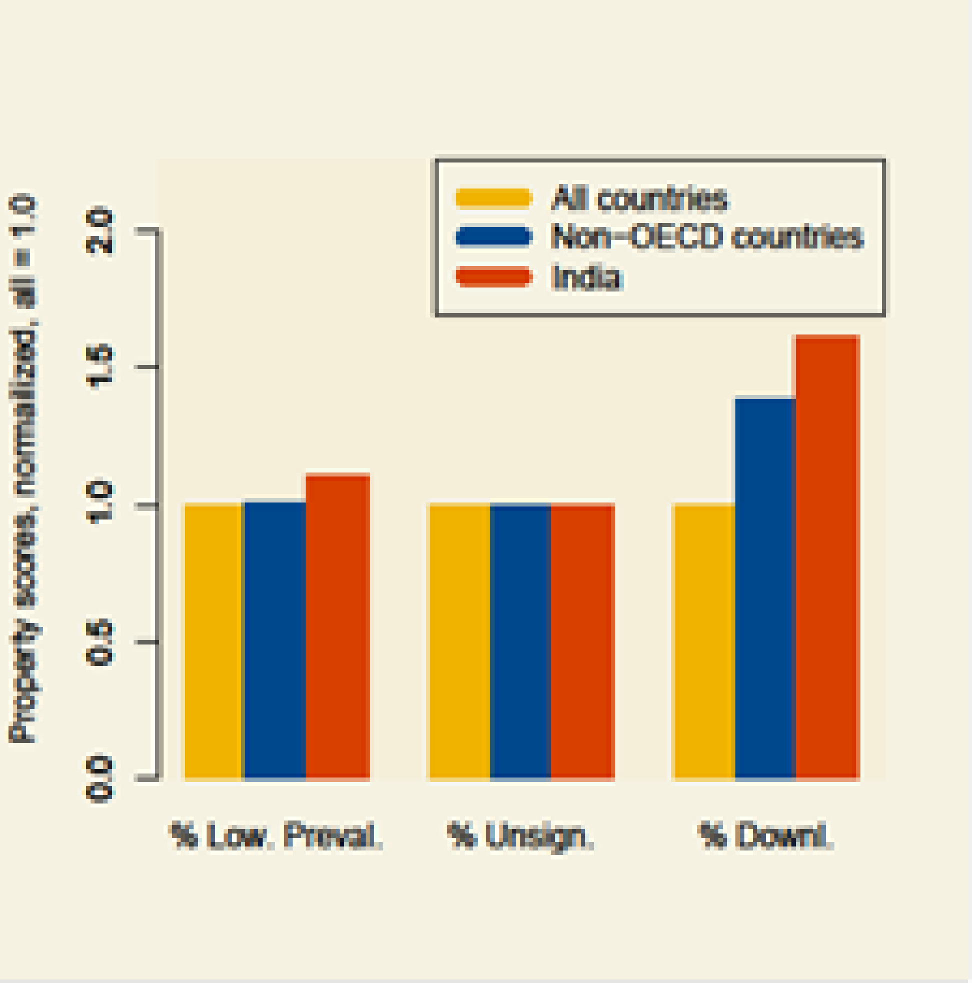
In general, the percentage of hosts in India with less than or equal to *x* attacks is smaller than for the entire world.

Not a good sign.

The Case of India: Types of Malware



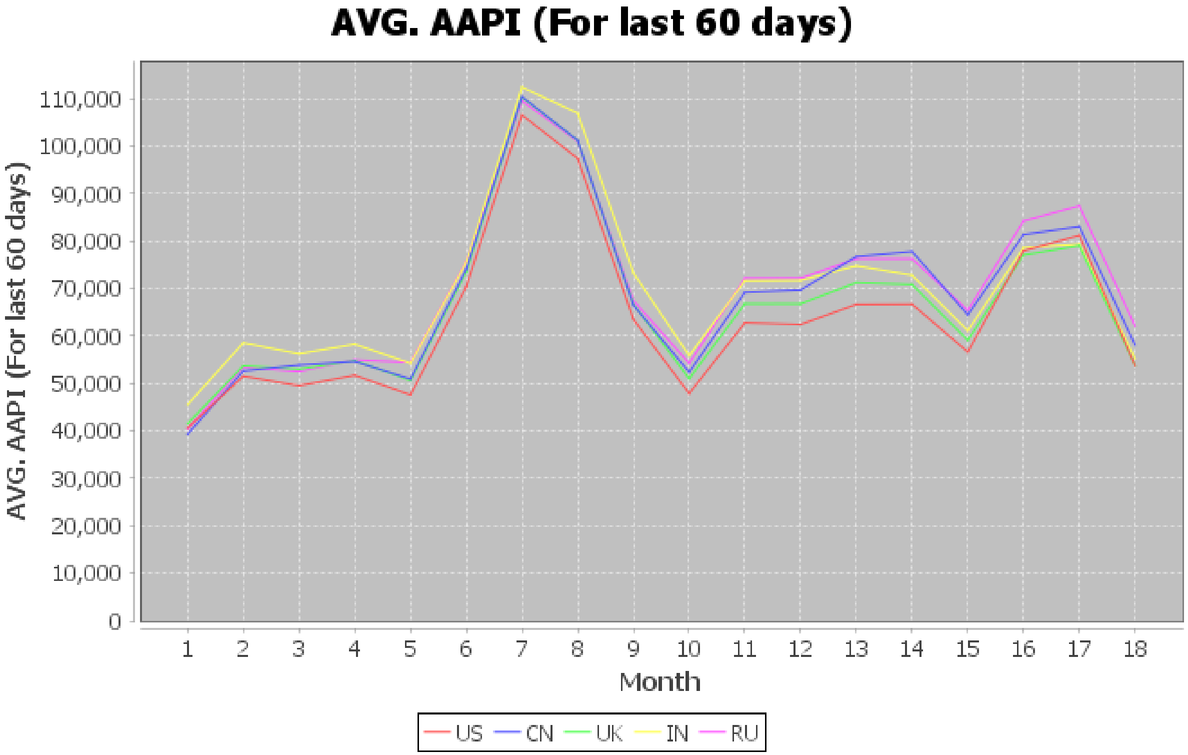
The Case of India: Cyber-Hygiene



Cyber-hygiene education in India must teach the risks of

- Installing low prevalence binaries
- Risks of drive-by attacks from visiting questionable sites and/or downloading binaries.

Case of India: Host Behavior



Average absolute patching incompetence (AAPI). Trends over 18 months similar for all 5 countries shown – but India is the worst of these 5 countries.

Study Methodology

- Two “sub” studies
- Human Vulnerability Study (HVS)
  - Identify behaviors of users that are correlated with the number of attacks on those users.
- Country Vulnerability Study (CVS)
  - Uses the results of the previous sub-study in order to identify the vulnerability of countries to cyber-attack
    - Aggregate Statistics
    - Behavior-based Statistics
- **Country Forecast Study (CFS)**
  - Uses the raw Symantec data to identify how to predict extent of malware spread in 40 countries.

Study Data

- Looked at the entire 2 years of data.
- Considered all machines from 40 countries. Reporting on results today for
  - China
  - India
  - S. Korea
  - UK
  - US
- Considered the 50 most commonly occurring malware.
- Tried an approximately 80-20 forecast. Checked to see how well we can forecast the number of attacked machines on the last 20% of days (by one of these 50 malwares) from the first 80% - both on a per-malware basis and in aggregate form.

Kang, C., Park, N., Prakash, B.A., Serra, E. and Subrahmanian, V.S., Ensemble Models for Data-driven Prediction of Malware Infections. *Proc. 9th ACM Conference on Web Search & Data Mining*, San Francisco, Feb 2016.

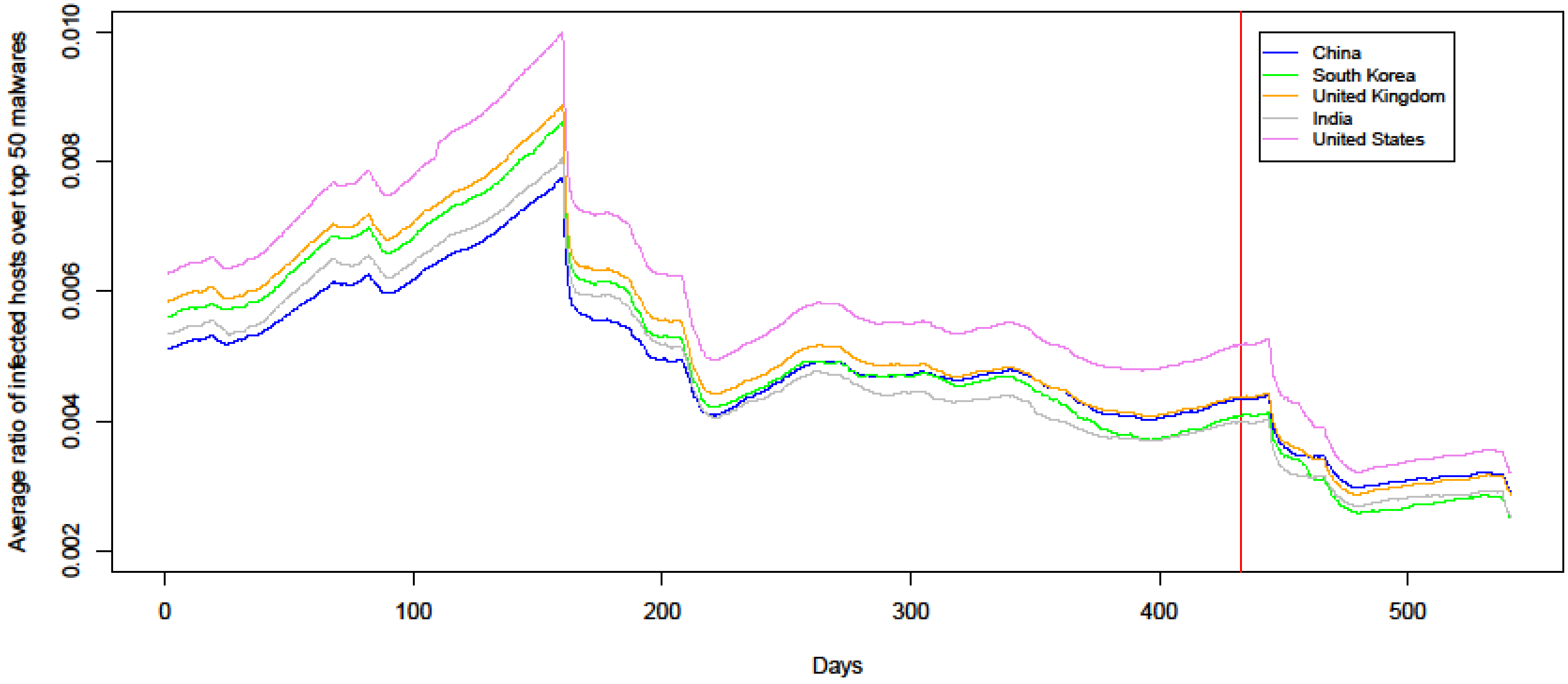
Question for the Audience

- Of the 5 countries, which ones would be most heavily affected by the 50 most common malware?
- Here are the 5 countries listed in alphabetical order
    - China
    - India
    - S. Korea
    - UK
    - US

Question for the Audience

- Of the 5 countries, which ones would be most heavily affected by the 50 most common malware?
- Order is more or less reversed for the 50 *most common malware.* Here are the 5 countries listed in descending order of targeting by the 50 most common malware we examined.
  - US
  - UK
  - S. Korea
  - India
  - China
- **Why?**

Percentage of hosts infected by top 50 malware: by country



Possible Answer

- S. Korea, India, and China may be doing a better job *patching* against the most common 50 types of malware. Or maybe in the US, people distribute patching effort across lots of malware components. *Our model took patching behavior into account.*
- But overall, US and UK did better patching across the board, which is why the overall percentage of infected hosts in the US and UK is much smaller than in S. Korea, India, and China.

Study Methodology

- Two “sub” studies
- Human Vulnerability Study (HVS)
  - Identify behaviors of users that are correlated with the number of attacks on those users.
- Country Vulnerability Study (CVS)
  - Uses the results of the previous sub-study in order to identify the vulnerability of countries to cyber-attack
    - Aggregate Statistics
    - Behavior-based Statistics
- Country Forecast Study (CFS)
  - Uses the raw Symantec data to identify how to predict extent of malware spread in 40 countries.
- **Recommendations**

Recommendation to Indian Security Managers

- **Everything must be patched.**
  - Attackers will use multiple attack vectors, not just ones with the highest impact, or the most common malware to compromise your system.
  - Use of lesser known malware which enterprises didn’t protect against allows attackers easy access to enterprise networks.
- **Must closely monitor user activity.** Users are a huge weak link in an enterprise. Must monitor:
  - Number of downloaded binaries.
  - Number of low-frequency binaries.
  - Other activity (e.g. exfiltration of sensitive customer data).

Recommendation to Indian CEOs

- **Prepare Response**
    - Run cyber-security red team exercises regularly
    - Have cyber incident response teams ready at all times
        - Technical – shut down attack
        - Legal – understand liability
        - PR – manage response to press and public
        - Liaise – work with police/government security staff
- **Share Security Attack Information**
    - You need to know what is going on around you.
    - Share security attack information with trusted partners.
    - If you are attacked, your partners could be next (and vice versa)

Recommendation to State/Central Government

- **Improve Education on Cyber-Hygiene**
  - Basic do's and don't's of cyber-security should be taught at the school level.
  - Need excellent cyber-security programs at universities.
  - Need training programs to significantly increase the number of incident response teams.
- **Cyber-Situation Awareness**
  - India cannot protect itself if it doesn't know who has been attacked (successfully or unsuccessfully). Should require reporting of certain types of cyber-incidents.
  - But government policy must incentivize and encourage sharing of attack data by victims, not inhibit or unnecessarily penalize/publicize it.
  - Set up framework to predict impact of new cyber-attacks on other parts of the Indian Internet infrastructure and take immediate corrective action.
- **Strategic Cyber-Security Partnerships**
  - Strength in cyber-offense does not necessarily imply strength in cyber-defense and vice-versa. Choose partners carefully.
  - Cyber-defense partners could include Japan.
  - Cyber-offense partners could include Israel, UK, and USA.
- **Strong Cyber-Deterrence**
  - Respondents must know that India has strong cyber-deterrence capabilities.
  - Build up strong cyber-attack capability.

Contact Information

V.S. Subrahmanian

Center for Digital International Government

Lab for Computational Cultural Dynamics

University of Maryland, College Park, MD.

vs@cs.umd.edu

@vssubrah

http://www.cs.umd.edu/~vs/

http://www.umiacs.umd.edu/research/CDIG/

http://www.umiacs.umd.edu/research/LCCD/