Blackworm: Analyzing the Spread of a Worm from Poisoned IP Data

Colleen Shannon and David Moore

Cooperative Association for Internet Data Analysis

About Blackworm

- Began to spread January 15, 2006
- 95k Visual Basic executable email attachment run by users
- Also spread to attached network shares
- Malicious: on the 3rd day of every month:
  - searches for files with 12 common file extensions (.doc, .xls, .mdb, .mde, .ppt, .pps, .zip, .rar, .pdf, .psd, and .dmp)
  - replaces those files with the text string "DATA Error [47 0F 94 93 F4 K5]"

So who cares?

- Blackworm is not particularly different from many, many other email viruses, except…
- Every infected computer automatically generates an http request for a web page that displayed a hit count graph (self-documenting code?)
- Logs for the website were available before the first date of payload destruction
- **Some victims could be notified before they lost data**

Log Analysis

- Simple! Just take the logs and look at who connected and you'll have the infected IP addresses!
- Except that the url was publicized…
- Many folks looked at the page to observe the spread of the virus
- Denial-of-service attacks added a large volume of spurious traffic

Log Filtering

- Why not just count IP addresses that were logged once?
- Web traffic aggregators (NAT, proxy servers) obscure victim IP addresses; multiple probes can represent mulitple infections
- DHCP use allows two different computers to have the same IP at the time that they become infected

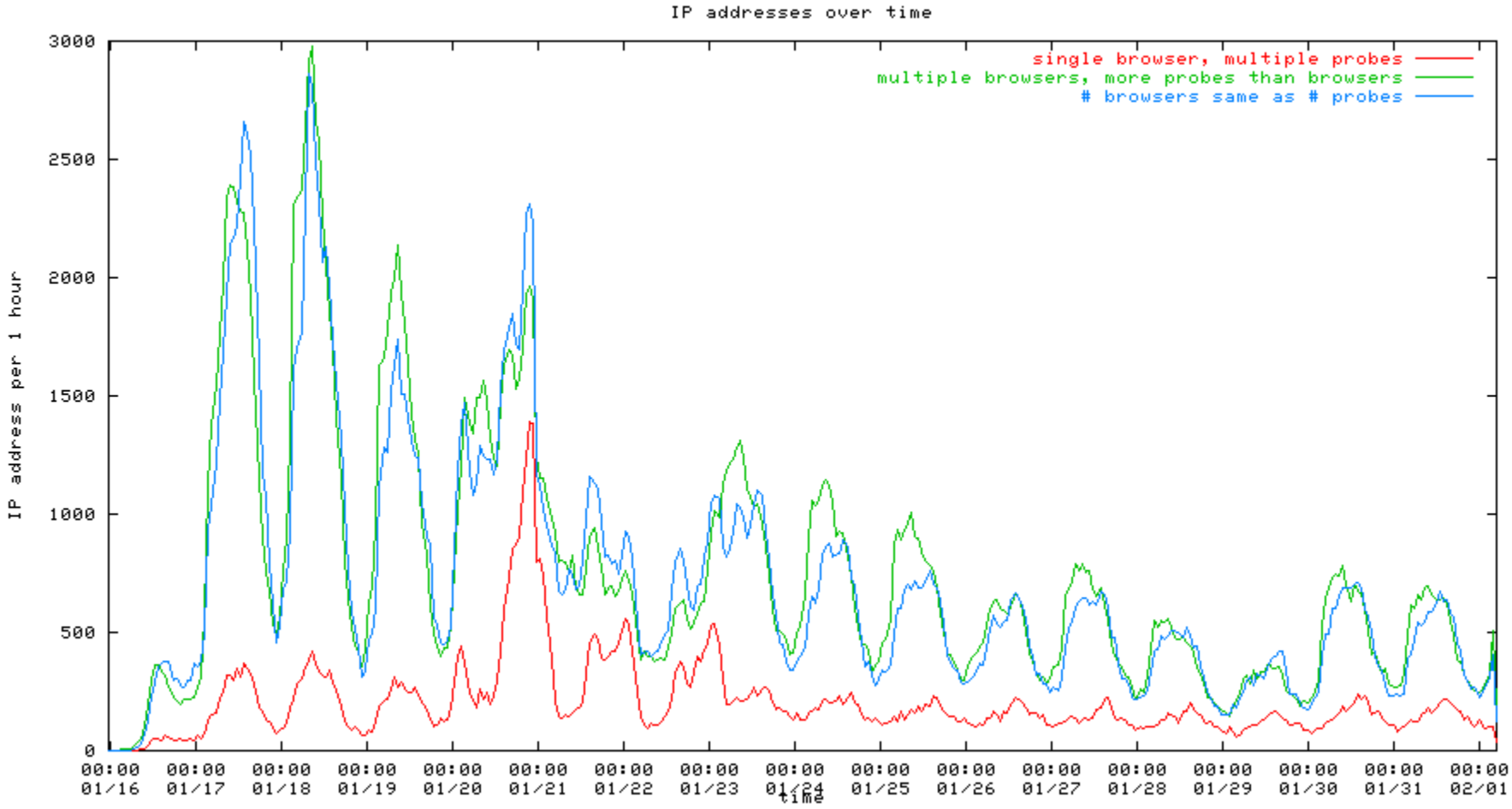
Log Filtering: DoS Attacks

- Many denial-of-service attacks use one tool deployed across many compromised computers
- Attack connections share common features: browser type, referer strings
- Those features combined with sharp onset and cessation identify DoS attacks in the log data

Log Filtering Process

- Remove referer/browser strings set by common DDoS tools (91.1% of all hits)
- Remove requests for pages different from the one accessed by the virus (0.2%)
- Remove any request with a referer string (virus did not use one in its probes) (0.8%)
- Remove requests from invulnerable Operating Systems: MacOS, Unix, cell phone, and PDA devices (0.03%)

Sanity Check



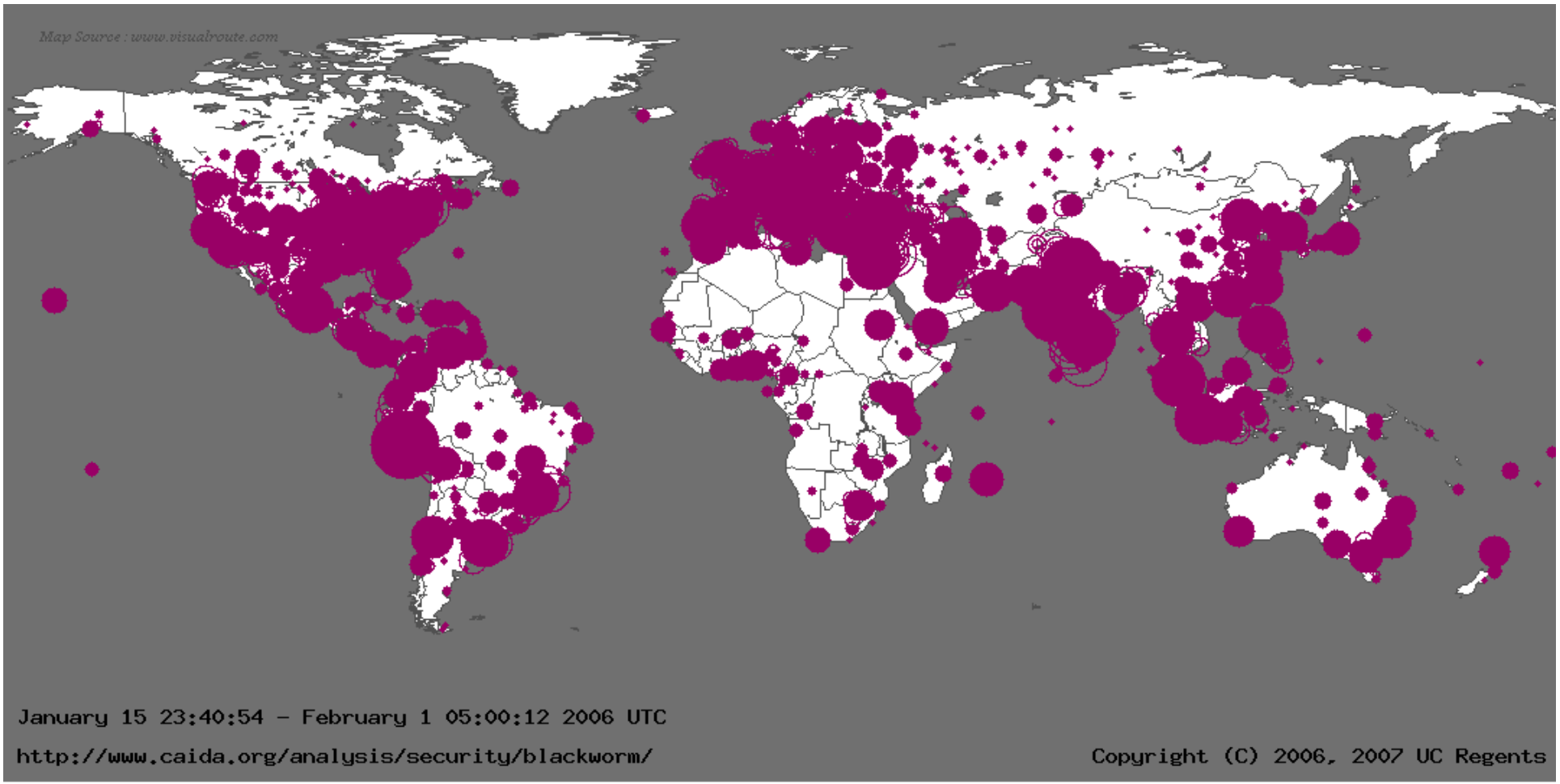
Sources of Error and Uncertainty

- Infected computers that failed to send the probe
- Network firewalls or outages that prevented victims from reaching the web page
- Denial-of-Service attacks preventing infected computers from reaching the web page
- People who viewed the counter only once using a vulnerable browser, but were not infected

Estimating a Victim Count

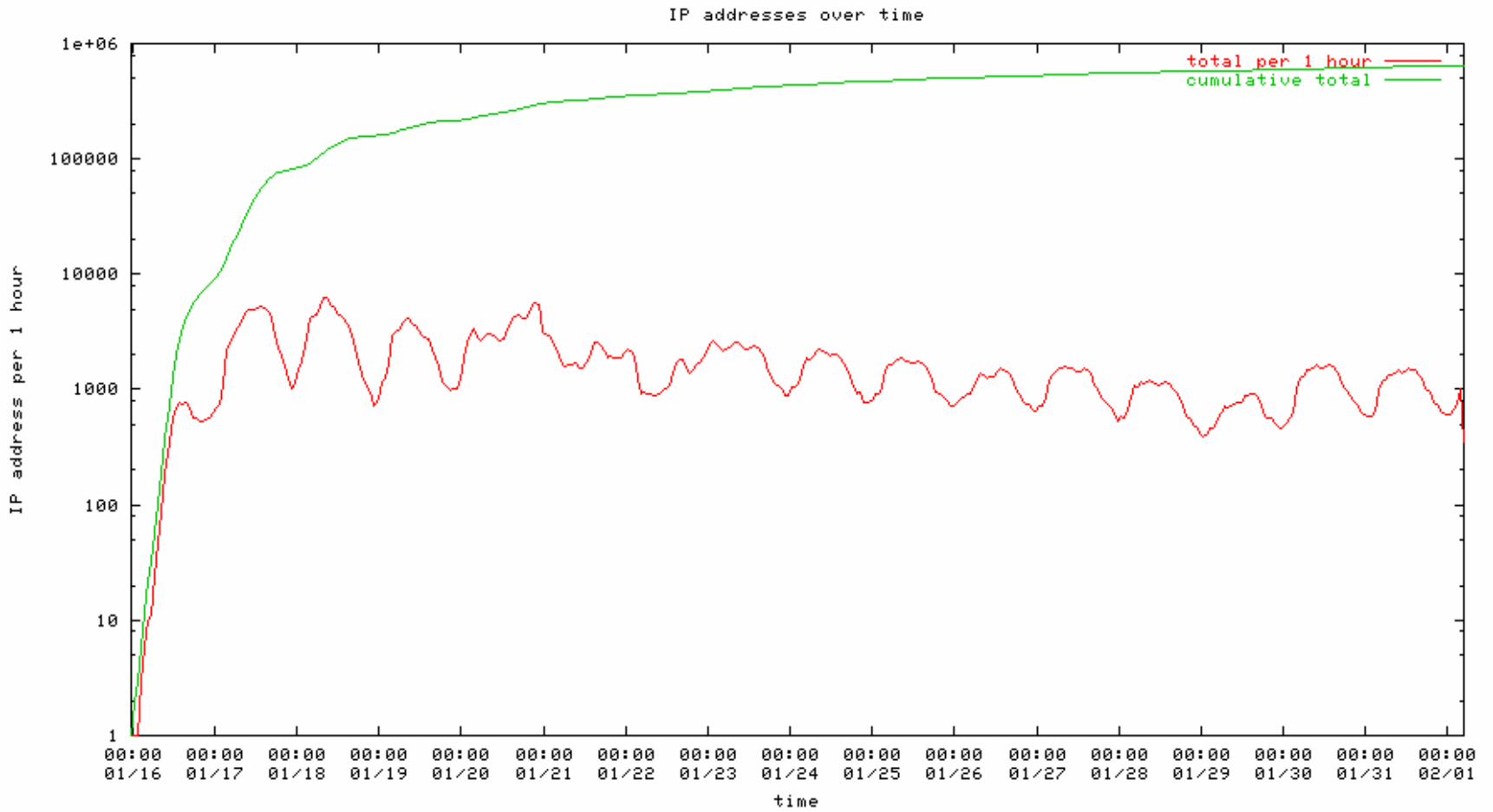
- Lower bound: for each IP address, the number of unique, vulnerable browser types received from that IP address

- Upper bound: for each IP address, the total number of probes received from that IP address

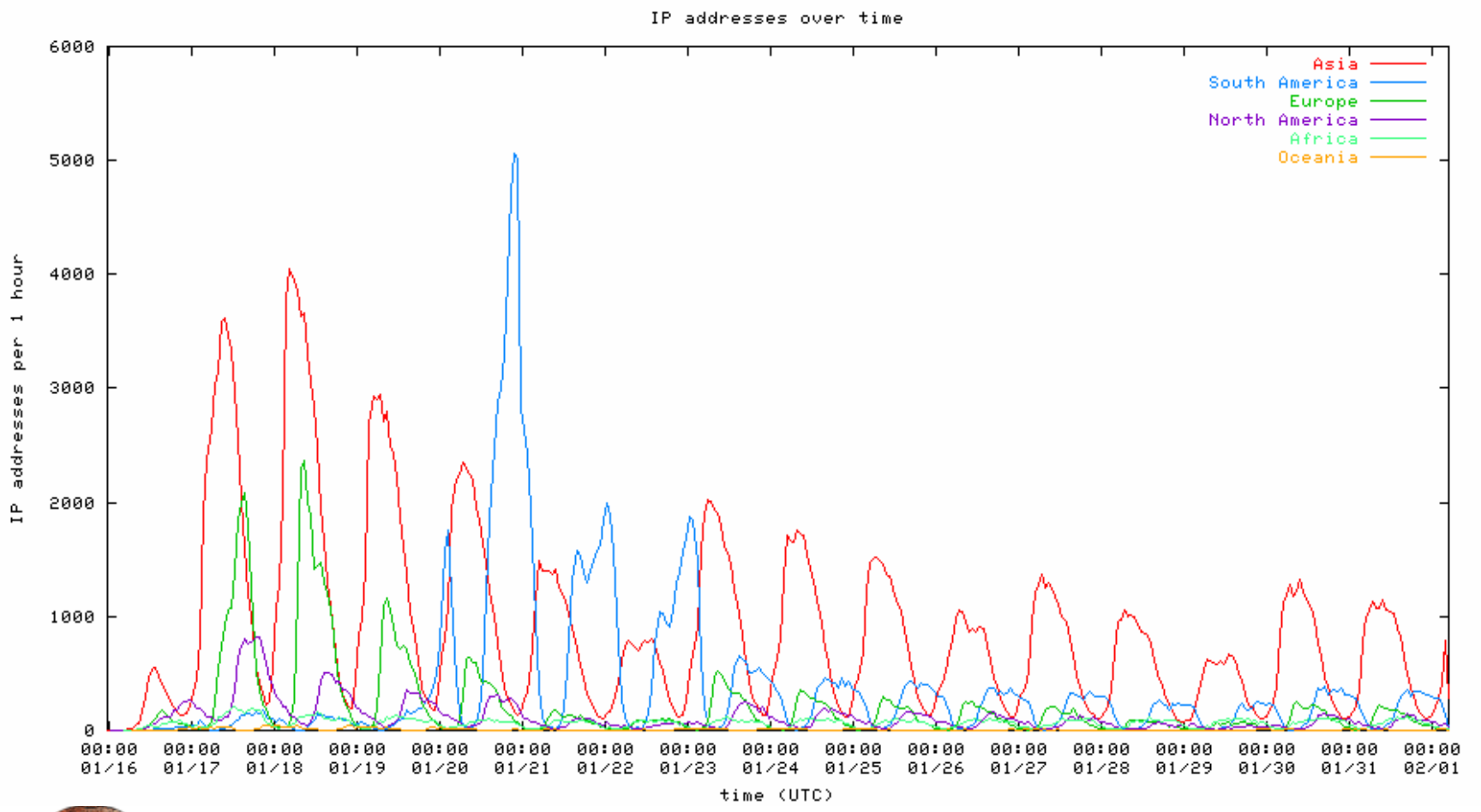


- Blackworm victim estimate: between 469,507 and 946,835 (3.2%-6.4% of original log entries)

Blackworm Overall



Blackworm by Continent



Blackworm by Country (>2%)

Country	Min. Count	Min %	Max Count	Max %
India	151341	32	273013	29
Peru	87599	19	150785	16
Italy	38216	8	58002	6
Turkey	28264	6	43437	5
USA	26315	6	58791	6
Egypt	12201	3	25104	3
Malaysia	11160	2	19942	2

Blackworm by TLD (>1%)

TLD	Min. Count	Min %	Max Count	Max %
Unknown	173510	37	367750	39
net	77706	17	141308	15
pe	71881	15	123960	13
it	31367	7	45923	5
in	25127	5	52818	6
com	18516	4	39283	4
tr	16162	3	24204	3

Concurrent Infections

- 45,401 Blackworm victims (10%) had concurrent spyware and/or botnet infections advertised in their browser string
  - `Mozilla/4.0 (compatible; MSIE 5.5; Windows 98; Sgrunt|V109|29|S493689067|dial; FunWebProducts; XBE|29|S0406967952114 3#398|isdn; snprtz|S04138822910124)`

Cuttlefish Animation…

- See http://www.caida.org/analysis/security/blackworm/#Animations

Conclusions

- Log analysis allows insight into email virus spread given sufficient data mining
- Email viruses spread in a slower and steadier pattern than Internet worms, which infect the vast majority of their victims in the first day
- Diurnal patterns are strongly apparent in spread data (people read their email when they are awake)

Conclusions (2)

- Country distribution of victims does not correlate with web infrastructure development
- Spread strongly influenced by geographic location (based on social and linguistic similarity)
- TLD distribution reflects geographic distribution rather than # of vulnerable hosts/TLD
- 10% of victims had concurrent botnet or spyware infection

Acknowledgements

- Thanks to our sponsors:

- Thanks also to: Gadi Evron, Paul Vixie, Joe Stewart, Mikko Hypponen, Swa Frantzen, Randy Vaughn, Chris Jackman, Jason Nealis, Rob Thomas, and Lorna Hutcheson for providing us with data and insight into the spread of the virus.

Internet Measurement Data Catalog

http://imdc.datcat.org