Understanding Divide-Conquer Scanning Worms

Yubin Li
Florida International University

Zesheng Chen
Florida International University

Chao Chen
Indiana University – Purdue University Fort Wayne

Outline

- Background and motivations
- Mathematical model
- Simulation study
- Countermeasures
- Conclusions & future work

Background

- **Internet worms: one of the top 4 security problems**
  - **Witty worm infected 12,000 hosts in 45 minutes in 2004**
  - **Storm worm affected tens of million of hosts in 2007**

- **Scanning method**
  - **A key factor for an efficient worm attack.**

Scanning Methods \ Worms	Slammer	CodeRed V2	CodeRed II	Witty	Warhol	Flash
Random scanning	X	X		X		
Localized scanning			X			
Hitlist scanning				X	X	X
Permutation scanning					X	

Table 1. Worm Scanning Methods

Background

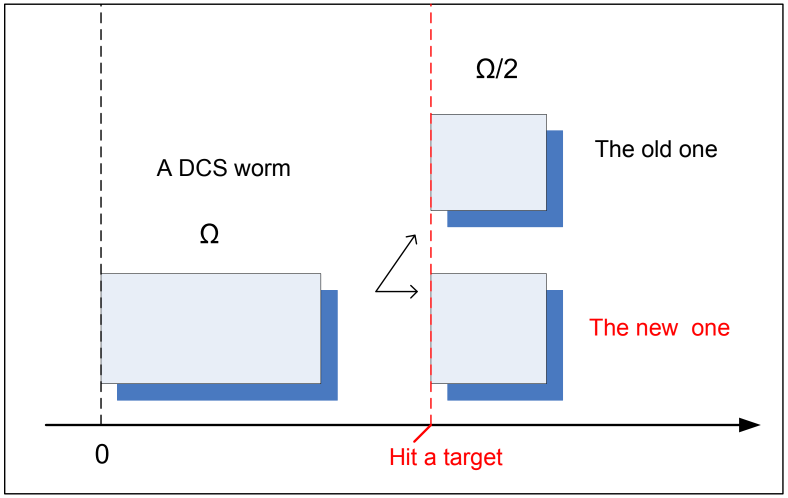
Scanning Methods / Parameters	Random/ Camouflaging	Localized/ Importance	Routable/ Divide-conquer
Scanning rate	X		
Scanning probability		X	
Scanning space			X

Table 2. Three Parameters of Scanning Methods

Background

❖ Divide-conquer scanning

- Named after divide-and-conquer algorithm
- Divides scanning space into half after infecting a target



a. ***Efficient***: avoids that different infected hosts attack the same target

b. ***Fast***: spreads much faster than random scanning

c. ***Stealthy***: weakens the detection of some defense systems

Background

- **Related work** [1][2]
  - Both works assume that vulnerable hosts are uniformly distributed
  - **Conclusion:** Divide-conquer scanning worms have a similar propagation speed as random-scanning worms

[1] J. Xia, S. Vangala, J. Wu, L. Gao, and K. Kwiat, "Effective worm detection for various scan techniques," ***Journal of Computer Security*****, vol. 14, no. 4, 2006, pp. 359-387.**

[2] C. C. Zou, D. Towsley, and W. Gong, "On the performance of Internet worm scanning strategies," ***Elsevier Journal of Performance Evaluation*****, vol. 63. no. 7, July 2006, pp. 700-723.**

Motivations

- **Bridge random scanning and divide-conquer scanning**

***/l* divide-conquer scanning (*/l* DCS)**

a. If an infected host is scanning subnet *a.b.c.d/k* and *k*<*l*, this hosts would divide its scanning space into halves after it compromises a target

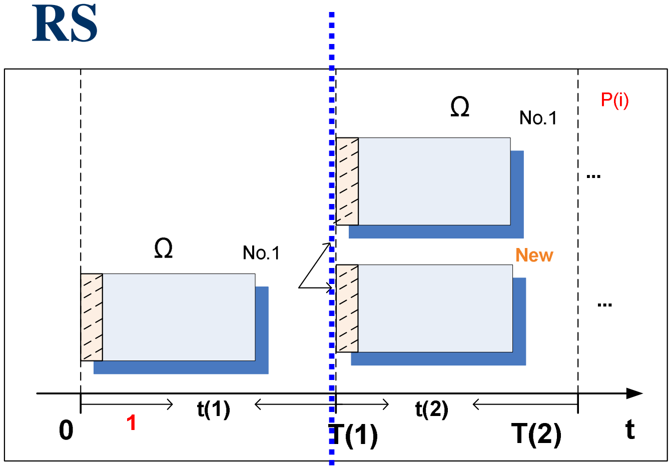
b. Otherwise, *k*=*l*, and the host will **not** divide its scanning space and will still scan *a.b.c.d/l* even after infecting other hosts. The new victims by this host will also scan subnet *a.b.c.d/l*

  - Random scanning: */0* DCS
  - Original divide-conquer scanning: */32* DCS

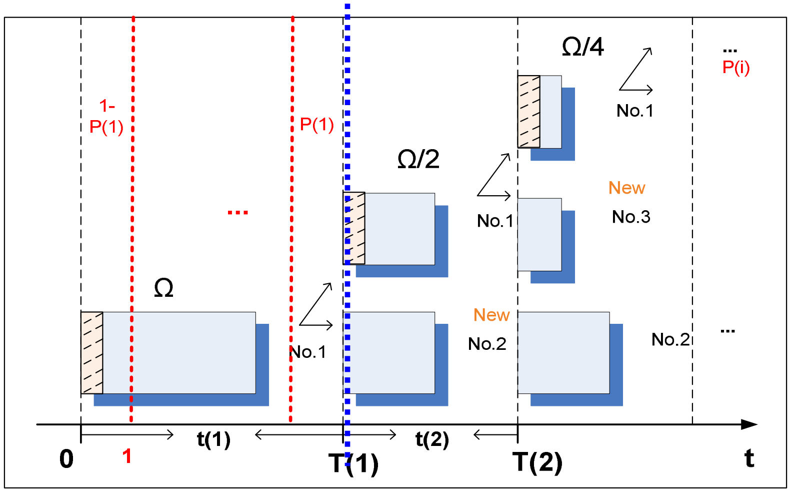
- **Demonstrate a toy example**
  - Compare propagation speeds of /16 DCS and random scanning (RS)
  - Assume vulnerable hosts distributed extremely uneven (65536 vulnerable hosts in a /16 subnet) and hitlist = 1

Motivations

RS



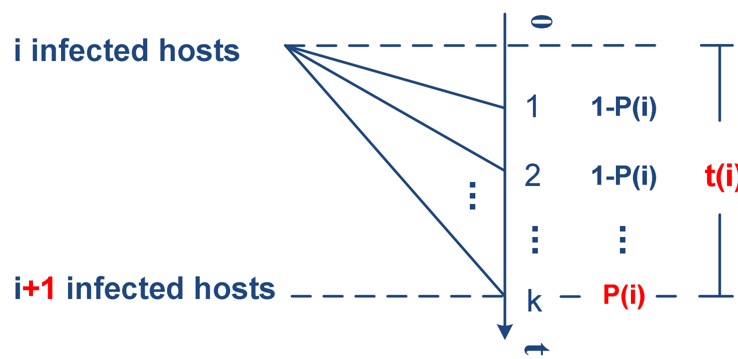
/16 DCS



T(n) : propagation time which is the average time for a scanning method to infect n vulnerable hosts at the early stage.

P(i) : the probability for *i* infected hosts to hit a target in one time ticket

t(i), the average time for *i* infected hosts to find a target, which follows the geometric distribution.



$$P\left[t(i)=k\right]=(1-P(i))^{k-1}P(i)$$

$$E\left[t(i)\right]=\frac{1}{P(i)}$$

Motivations

- **RS propagation time**

the number of scans

prob. for a single scan to hit the target

$$P_{RS}(i) = s \cdot i \cdot \frac{N}{\Omega}$$

$$T_{RS}(n) = \sum_{i=1}^{n} t_{RS}(i) = \sum_{i=1}^{n} \frac{1}{p_{RS}(i)} = \frac{\Omega}{sN} \sum_{i=1}^{n} \frac{1}{i}$$

- **/16 DCS propagation time**
  - If i≤ 16

the number of scans

prob. for a single scan to hit the target

$$P_{DCS}(i) = s \cdot \frac{N}{\Omega / 2^{i-1}}, \quad i \le 16$$

$$T_{DCS}(n) = \sum_{i=1}^{n} \frac{1}{p_{DCS}(i)} = \frac{\Omega}{sN} \sum_{i=1}^{n} \frac{1}{2^{i-1}}, \quad i \le 16$$

<

  - If i>16, hitlist scanning, definitely spread faster than RS

Observations

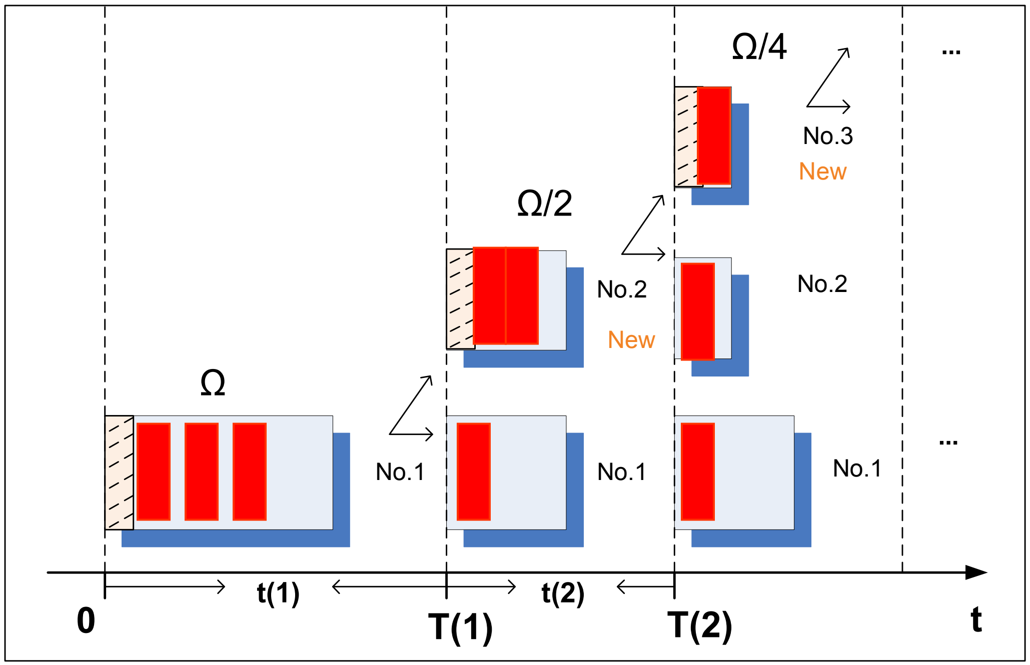
DCS can spread a worm much faster than RS

DCS could lead a worm to spread towards a subnet with many vulnerable hosts

Motivations

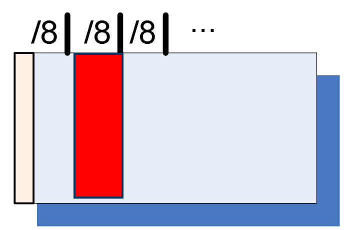
❖ Stealth

Detection system : network telescopes (contain no vulnerable hosts)



It is to better locate the network telescopes most close to the subnet of vulnerable hosts.

CAIDA uses an entire /8 subnet as network telescopes.



a.b.c.d/
2 infected hosts can be detected
1 infected host can be detected
k+1 infected hosts at most

Vulnerable hosts in 1.0.0.0/8

Network telescopes in 2.0.0.0/8

Our Goals

Characterize the relationships between the propagation speeds of DCS worms and the distributions of vulnerable hosts through mathematical analysis and simulations

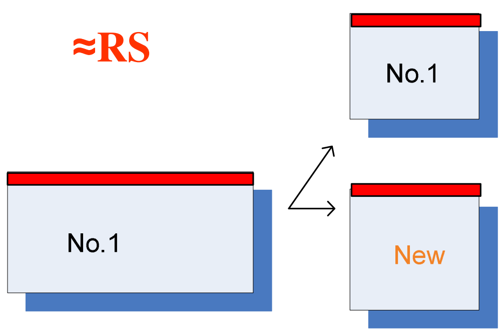
Study the weakness of DCS and defense mechanisms

Outline

- Background and motivations
- Mathematical model
- Simulation study
- Countermeasures
- Conclusions & future work

Mathematical Analysis

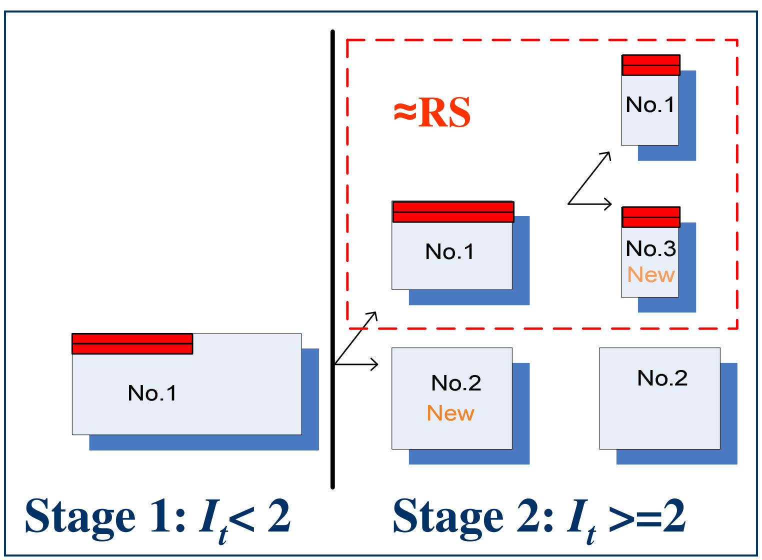
❖ How can we model different distributions of vulnerable hosts?



a. Uniform in the IPv4 (/*0* subnet)

$$I_{t+1} = I_t + (N - I_t)[1 - (1 - \frac{I_t}{\Omega})^s]$$

$$\approx I_t + \frac{s}{\Omega} I_t (N - I_t) \qquad (1)$$



b. Uniform in half of the IPv4 (/*1* subnet)

Equation (1)

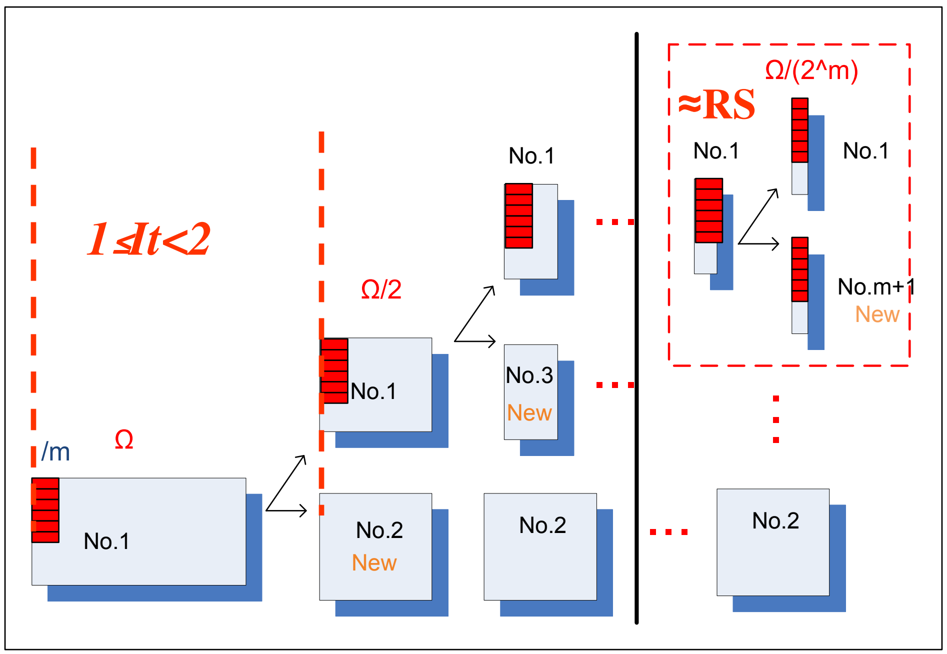
$$I_{t+1} = I_t + (N - I_t)\left\{1 - \left[1 - \frac{(I_t - 1)}{\Omega / 2}\right]^s\right\}$$

$$\approx I_t + \frac{2s}{\Omega}(I_t - 1)(N - I_t)$$

Mathematical Analysis

"locally evenly distributed, and global unevenly" -- a general case

c. Uniform in a /*m* network



Stage 1: $I_t < m+1$,

Specifically $i \le I_t < i+1$

$$I_{t+1} = I_t + (N - I_t)\left\{1 - \left[1 - \frac{(I_t - i + 1)}{\Omega / 2^{i-1}}\right]^s\right\}$$

$$\approx I_t + \frac{2^{i+1} \cdot s}{\Omega}(I_t - i + 1)(N - I_t)$$

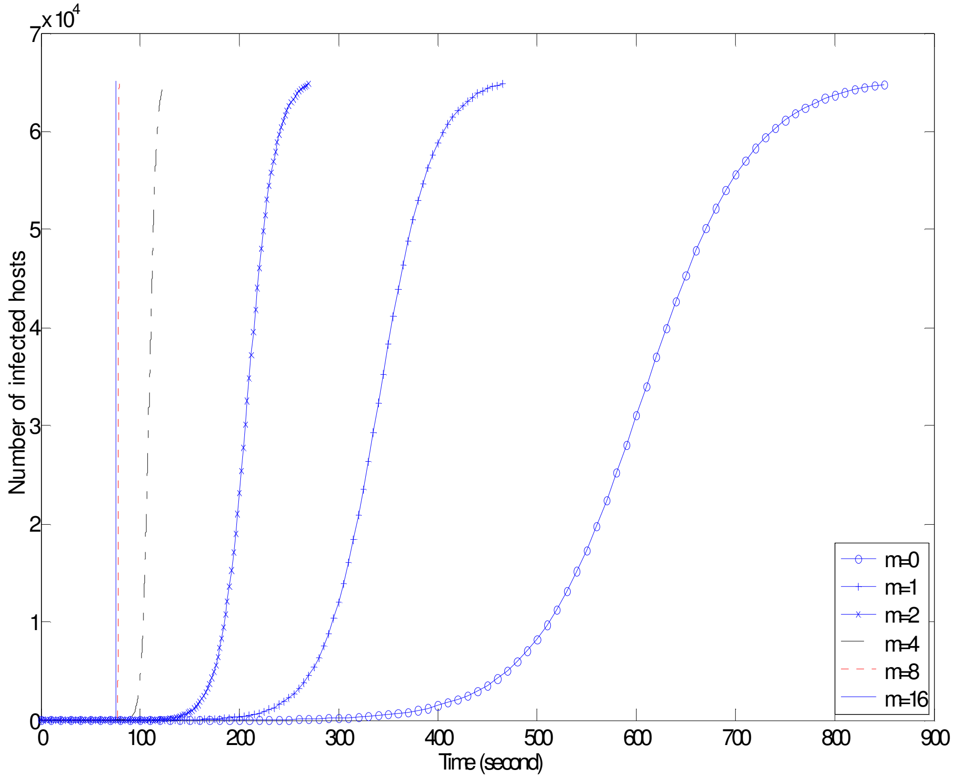
Stage2: $I_t \ge m+1$

$$I_{t+1} = I_t + (N - I_t)\left\{1 - \left[1 - \frac{(I_t - m)}{\Omega / 2^m}\right]^s\right\}$$

$$\approx I_t + \frac{2^m \cdot s}{\Omega}(I_t - m)(N - I_t)$$

Mathematical Analysis

- **Results by applying above equations and varying *m*.**



Hitlist: 1

Scanning rate: 1200/s

Vulnerable population: 2^{16}

Curves from right to left:

***m* = 0, 1, 2, 4, 8, 16**

Conclusion : When *m* is larger, the distribution of vulnerable hosts is more uneven, and DCS spreads a worm faster.

Outline

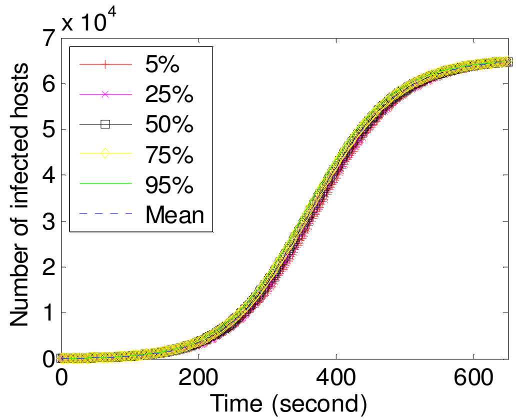
- Background and motivations
- Mathematical model
- Simulation study
- Countermeasures
- Conclusions & future work

Simulation Study

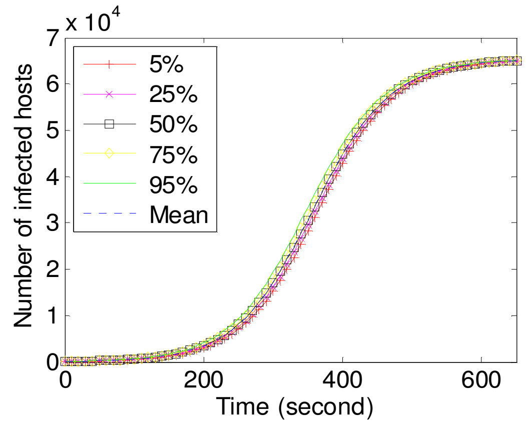
- **Why perform simulation study?**
  - Provide more realistic scenarios
  - Consider an arbitrary distribution of vulnerable hosts
  - Give the variation of the number of infected hosts
- **Implement a simulator**
  - Run 100 times with different seeds
  - Start from 100 initially infected hosts
  - Set scanning rate = 1200 scans/second and vulnerable population = 2^{16}
  - Follow /16 DCS
- **Design vulnerable-host distributions**
  - "nonuniform-*u*" distribution: A higher value of *u* gives a more uneven distribution of vulnerable hosts

    (u=0: uniform distributed ; u=16:extremely unevenly distributed)
  - Compare the simulation results with *u=0, 4, 8, 12,16* and witty-worm victim distribution

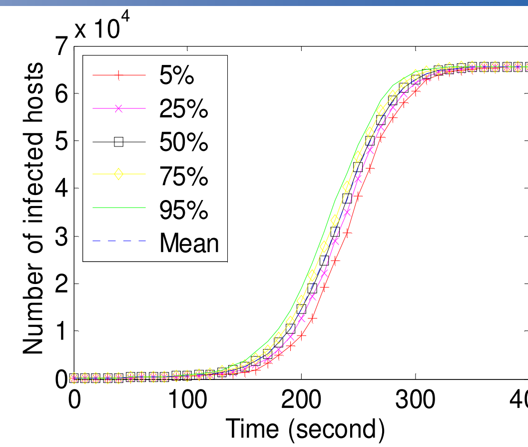
Simulation Study



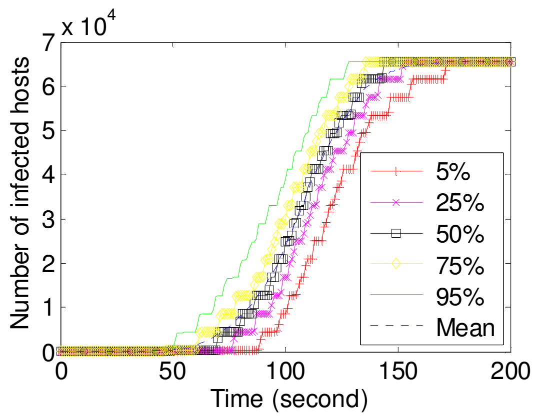
Nonuniform-0



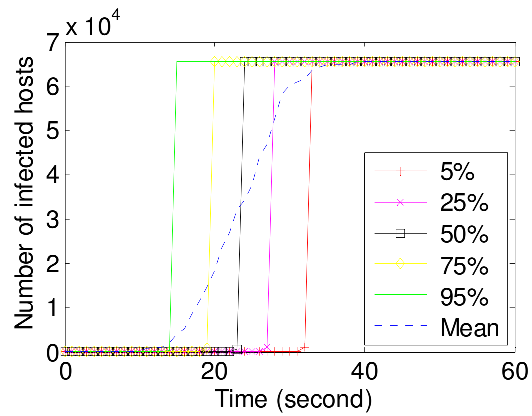
Nonuniform-4



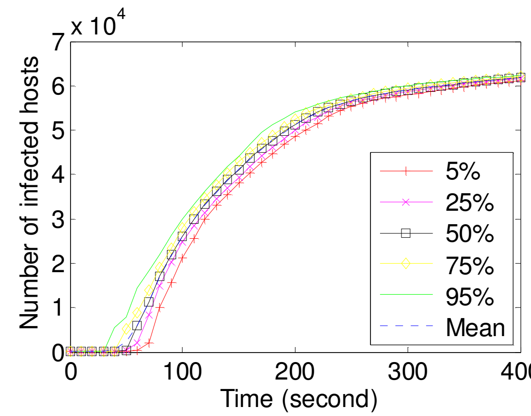
Nonuniform-8



Nonuniform-12



Nonuniform-16

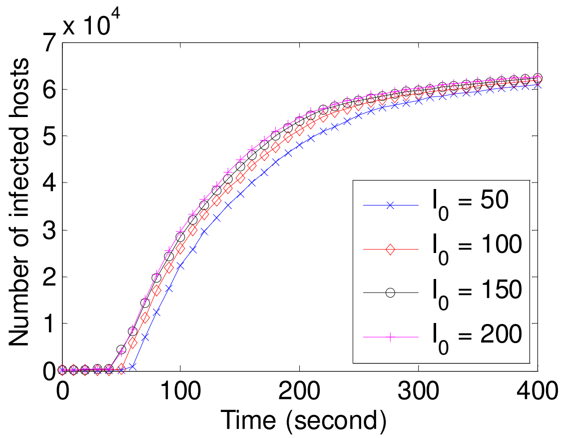


Witty

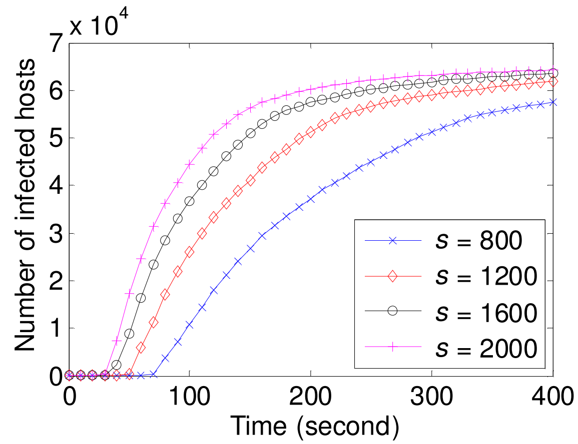
Conclusion: When *u* increases, DSC worms spread faster.

Simulation Study

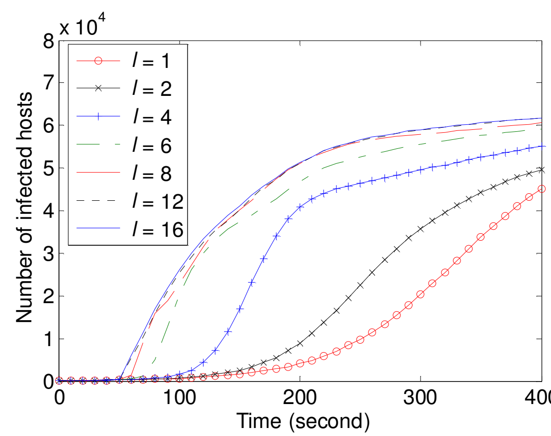
- **Propagation speed & important parameters**
  - Vulnerable hosts distribution
  - Hitlist
  - Scanning rate
  - Degree of divide and conquer



a. Hitlist



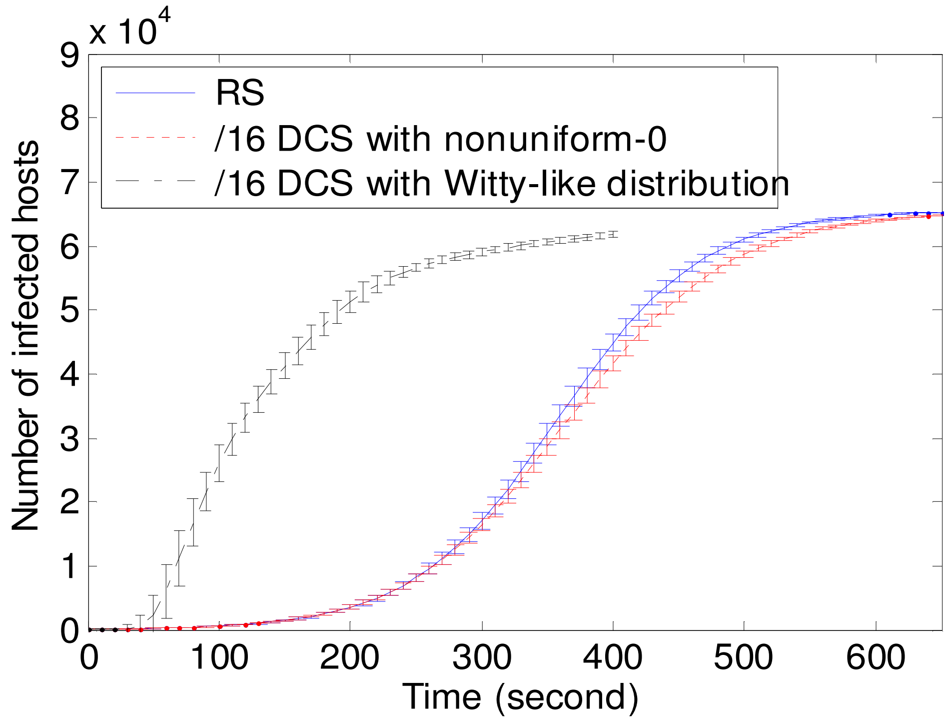
b. Scanning rate



c. Degree of divide and conquer

Simulation Study

❖ **Comparison between RS and /16 DCS**



Conclusion : If the vulnerable hosts distribution is uniform, /16 DCS spreads slightly slower than RS in the late stage. But /16 DCS spreads much faster than RS under witty-like distribution.

Outline

- Background and motivations
- Mathematical model
- Simulation study
- Countermeasures
- Conclusions & future work

Countermeasures

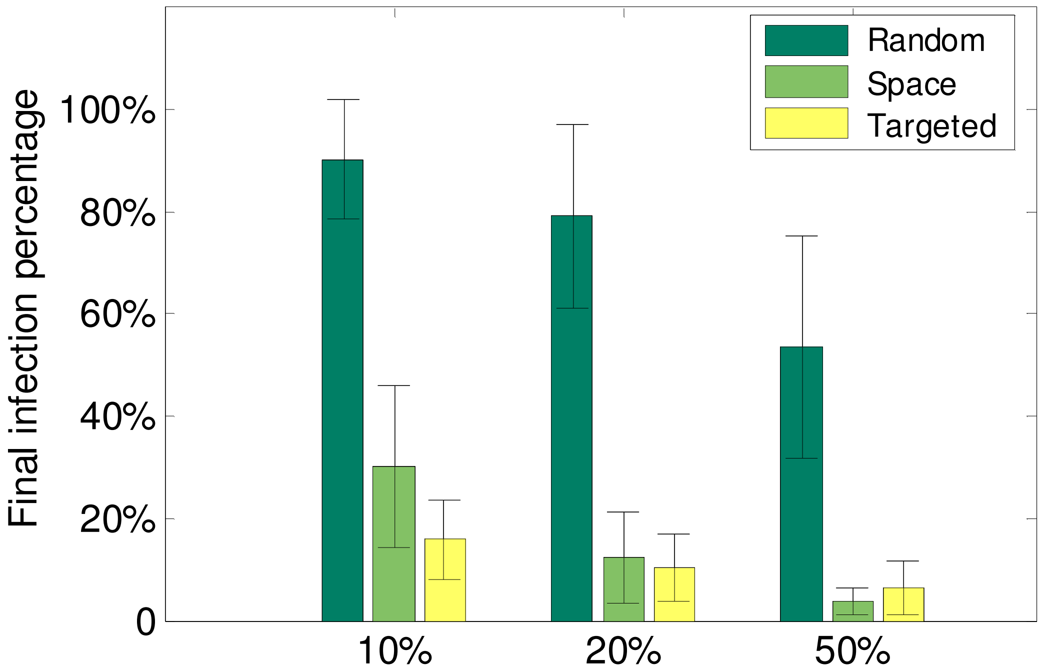
- **How can we defend against DCS worms?**

- **DCS is vulnerable to nodal failures at early stage**

  The hosts in the space which is assigned to the certain removal infected host can be protected.

- **Consider three removing strategies**
  - ***Random*: Remove infected hosts randomly
  - ***Space*: Remove infected hosts that scan the largest address sub-space
  - ***Targeted*: Remove infected hosts that scan address subnets containing the the largest number of vulnerable hosts

Countermeasures



Hitlist = 1

Vulnerable population = 2^{16}

When 100 hosts are infected, remove 10%, 20%, and 50%

Conclusion : “Space” and “targeted” removal can effectively defend against DCS worms, and “targeted” is not always better than “space”

Conclusions & Future Work

- **DCS can be a powerful attacking tool for future Internet epidemics because of its characteristics**
  - *Efficiency*
  - *Faster*
  - *Stealth*

- **Future work**
  - Studying variants of DCS worms
    - Adaptive DCS
  - Developing other effective defense mechanisms
    - Active honeynet

Q & A