

An Empirical Analysis of Exploitation Attempts based on Vulnerabilities in Open Source Software

new subtitle:

"the boy who [accidentally] kicked the hornet's nest"…

Sam Ransbotham, *Boston College*

BOSTON COLLEGE

Overview

Question: *Does availability of source code affect exploitation attempts?*

Specifically, does open source code affect…

Risk: the likelihood of a vulnerability being exploited?

Diffusion: the diffusion of exploitations based on a vulnerability?

Volume: the volume of exploitations based on the vulnerability?

Methodology

Statistical analysis of intrusion detection system alert and NVD data

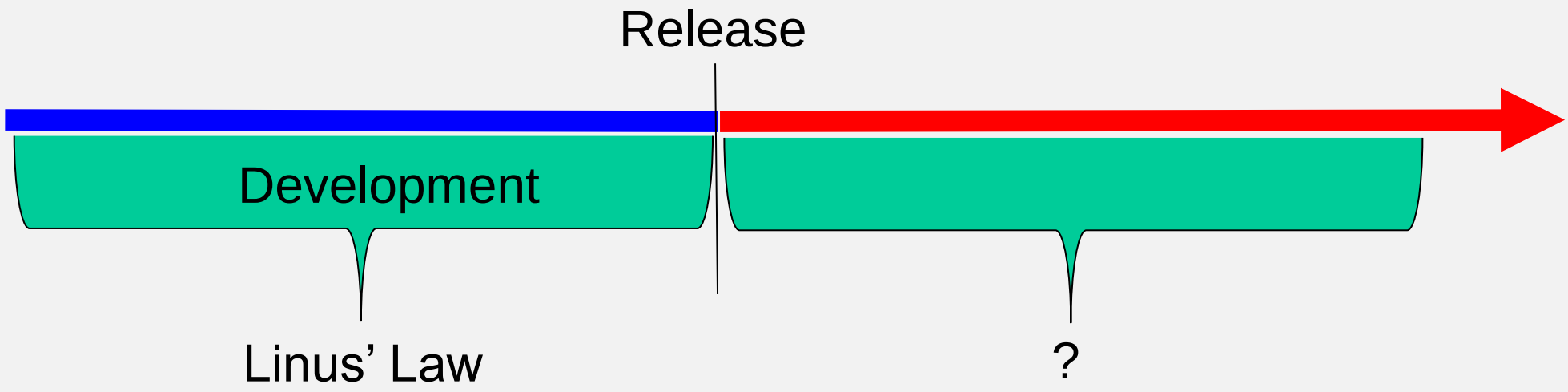
Key Result

Open source can increase the risk, diffusion and volume of exploitation attempts for disclosed vulnerabilities.

NOT the Research Question(s)

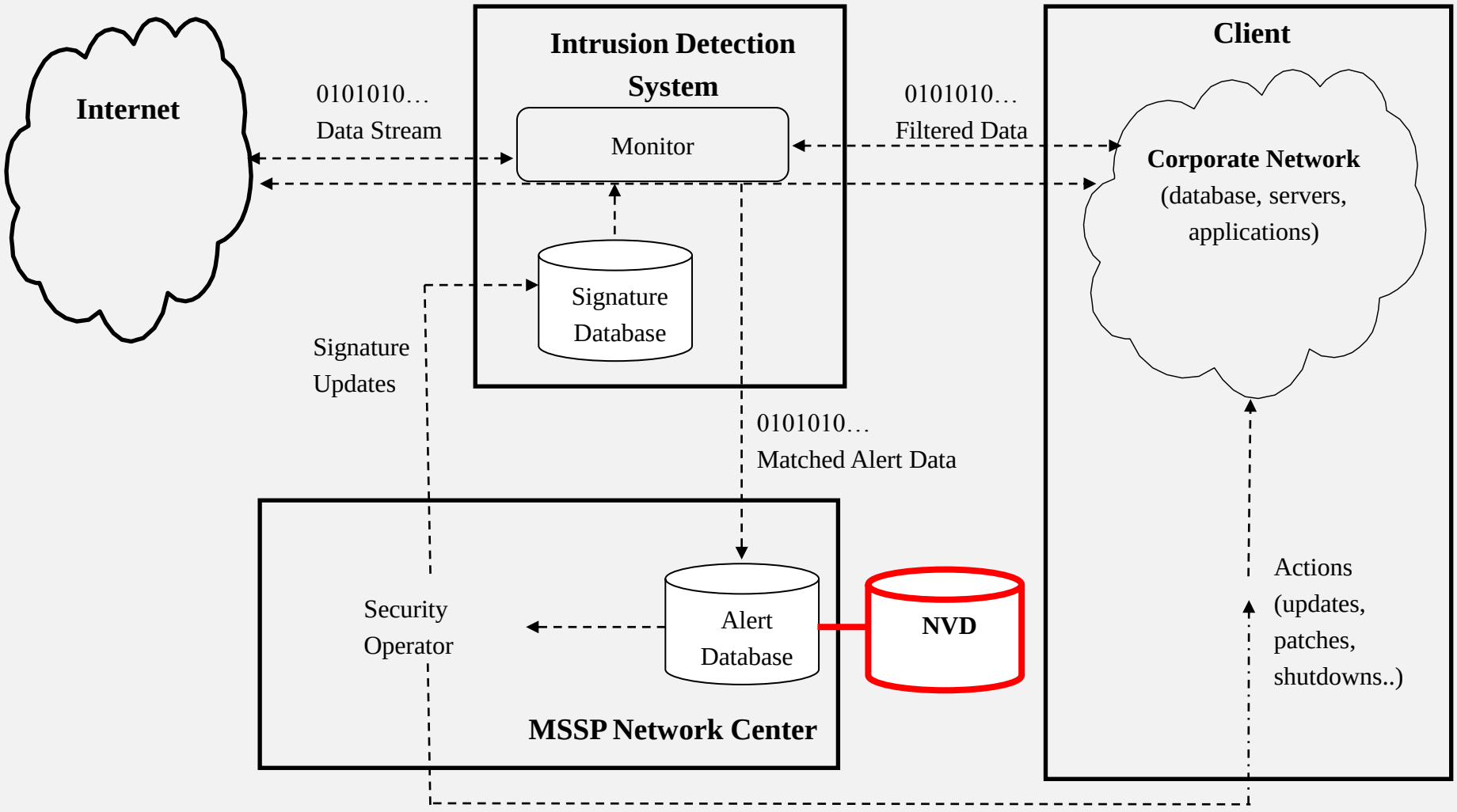
- Are more bugs in open source caught before release?
- Is open source more secure?
- Is open source better than closed source?

Simplified Software Development Process



Core argument: Information

Research Environment



Summary of Data

1. Alert data from MSSP
2. CERT/NVD vulnerability information
3. Manual coding of open/closed source
    - 13,101 distinct software products
        - 3,369 (26%) open source
        - 3,121 (24%) closed source
        - 6,611 (50%) unknown

Key unique features

- Not single client; multiple client
- Extended time period (two years)
- Real, not honeypot

Key Control Variables

1. Common Vulnerability Scoring System (CVSS) Assessment
   A. Access required: (local, adjacent, remote)
   B. Complexity: (low, medium, high)
   C. Authentication: (required or not)
   D. Impacts: (confidentiality, data integrity, availability of system resources)
   E. Type
      1. Access Validation: incorrect allowance of privileges
      2. Input Validation: failure to handle incorrect input
      3. Design Error: shortcomings in design of software
      4. Exception Error: Insufficient response to unexpected conditions
      5. Configuration Error: weak configuration of settings
      6. Race Condition: errors due to sequencing of events
2. Patch available
3. Signature available
4. Age of vulnerability (days since publication)

Data details

Measure		Open Source		Closed Source	
Exploited	No	329	92%	457	87%
	Yes	30	8%	67	13%
Access Required	Requires Local	50	14%	63	12%
	Requires Adjacent	3	1%	8	2%
	Network	306	85%	453	86%
Complexity	Low	187	52%	245	47%
	Medium	131	36%	225	43%
	High	41	11%	54	10%
Authentication	Not required	337	94%	508	97%
	Required	22	6%	16	3%
Confidentiality Impact	No	104	29%	105	20%
	Yes	255	71%	419	80%
Integrity Impact	No	103	29%	94	18%
	Yes	256	71%	430	82%
Availability Impact	No	69	19%	73	14%
	Yes	290	81%	451	86%
Signature Available?	No	348	97%	380	73%
	Yes	11	3%	144	27%

Does open source code affect the risk of exploitation?

Variable	Control Model		Test Model	
Complexity: Medium	-0.138***	(0.023)	-0.122***	(0.023)
Complexity: High	0.187***	(0.025)	0.169***	(0.025)
Confidence Impact	-0.133***	(0.027)	-0.143***	(0.027)
Integrity Impact	0.074*	(0.031)	0.107***	(0.030)
Availability Impact	0.361***	(0.034)	0.356***	(0.034)
Patch Available	-0.022	(0.018)	-0.031	(0.018)
Signature Available	0.979***	(0.020)	1.095***	(0.020)
Vulnerability Types	indicators		indicators	
Open Source			0.259***	(0.019)

Increased risk of exploitation attempt

Cox proportional hazard model of 12,661 exploitation attempts across 847,126 observations of 883 vulnerabilities in 960 firms; robust standard errors in parentheses; analysis stratified across 960 firms; significance levels: * p<0.05; ** p<0.01; *** p<0.001

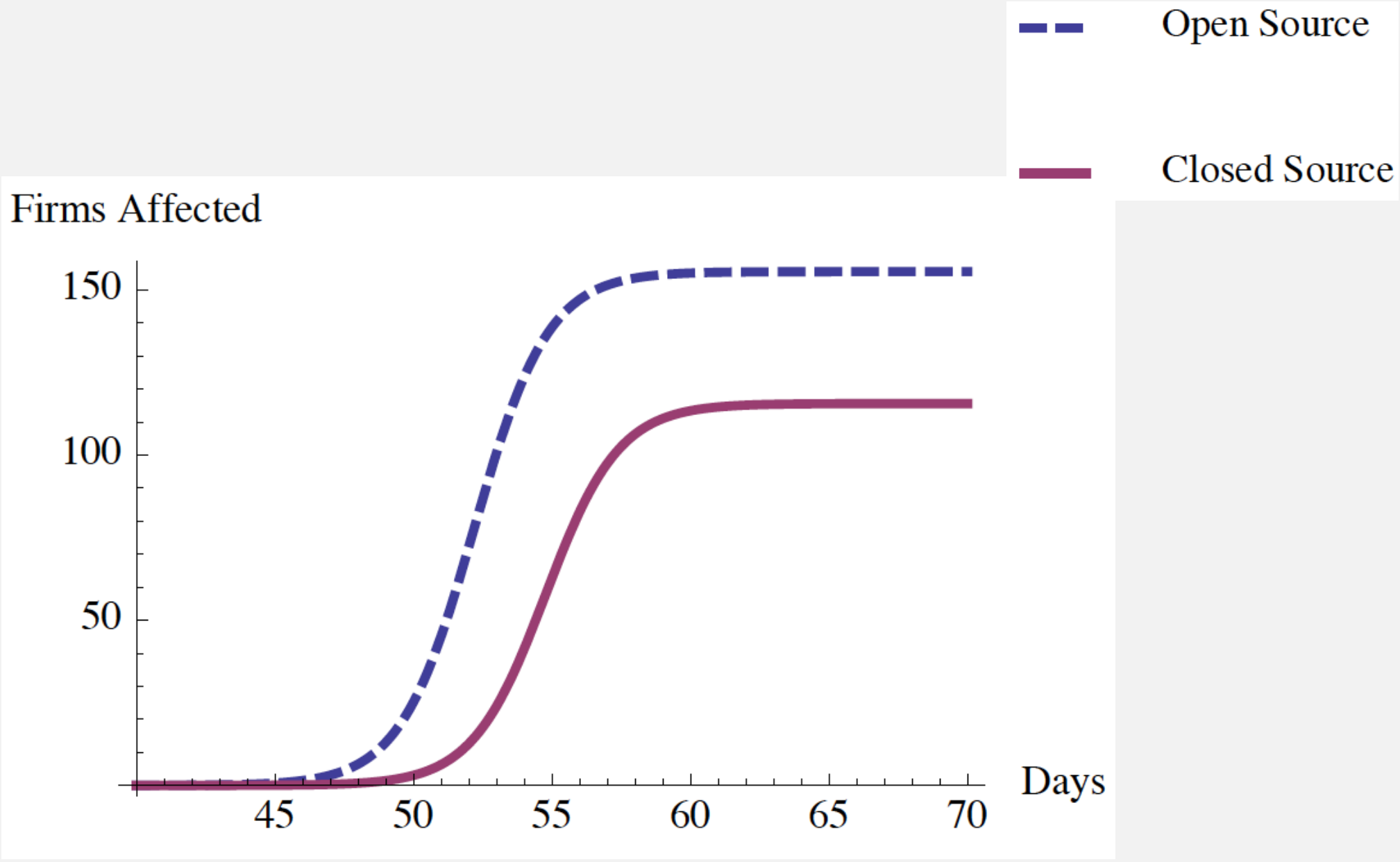
Does open source affect the diffusion of exploitation?

Variable	Penetration (P)		Rate (R)		Delay (D)	
Complexity: Medium	98.721***	(1.488)	0.220***	(0.026)	-0.466	(0.410)
Complexity: High	-1.211	(1.349)	-0.462***	(0.052)	1.709*	(0.593)
Confidence Impact	86.082***	(3.419)	-0.363***	(0.040)	26.190***	(3.014)
Integrity Impact	-193.342***	(3.792)	-0.620***	(0.067)	-9.945***	(1.449)
Availability Impact	-112.635***	(3.292)	-0.406***	(0.047)	-99.260***	(11.023)
Patch Available	32.821***	(1.117)	-0.150***	(0.018)	-2.746***	(0.510)
Signature Available	133.025***	(1.437)	1.370***	(0.149)	-31.127***	(3.554)
Vulnerability Types	indicators		indicators		indicators	
Open Source	46.995***	(1.261)	0.251***	(0.029)	-13.992***	(1.708)

$$N(t) = \frac{P}{1+e^{(-Rt-D)}}$$

Nonlinear regression on the cumulative number of affected firms; 83,806 daily observations of vulnerabilities exploited in at least one of 960 firms. Robust standard errors in parentheses; significance levels: *p<0.05; **p<0.01; ***p<0.001

Does open source affect the diffusion of exploitation?



Does open source code affect the volume of alerts?

Variable	Stage 1		Stage 2	
Complexity: Medium	-0.177***	(0.004)	-0.049***	(0.003)
Complexity: High	0.411***	(0.005)	-0.149***	(0.004)
Confidence Impact	-0.270***	(0.005)	0.122***	(0.004)
Integrity Impact	0.505***	(0.005)	-0.188***	(0.005)
Availability Impact	-0.055***	(0.005)	-0.005	(0.004)
Vulnerability Types	indicators		indicators	
Firm effects	indicators		indicators	
Monthly indicators	Publish month		Alert month	
Age (in days, log)			-0.199***	(0.004)
Patch Available	-0.068***	(0.003)	-0.042***	(0.002)
Signature Available	0.832***	(0.004)	-0.001	(0.004)
Open Source	0.072***	(0.004)	0.148***	(0.003)

increases volume

Heckman two stage regression; n = 896,407; 473,699 uncensored; 883 vulnerabilities; robust standard errors in parentheses; significance levels: * p<0.05; **p<0.01; ***p<0.001

Stage 1: uncensored if exploit attempt for the vulnerability is observed in the sample

Stage 2: natural log of the number of exploitation attempts

Summary

Result

Open source can increase the risk, diffusion and volume of exploitation attempts for disclosed vulnerabilities.

Limitations

- Do not observe vulnerability discovery (pre or post release)
  - Observe exploitation attempts for disclosed vulnerabilities
- Relatively limited set of vulnerabilities
- Open and closed source products may have differences in the types of vulnerabilities found
- High volume of noisy data: IDS and NVD
- Source benefits defenders too
- Patching and deployment are also important.

Thoughts, Extensions, and Creativity Needed

Despite limitations, an effect to be aware of

- What are alternative explanations for results?

Benefits from open source remain at the pre-release stage

- Likely outweigh effects shown in this paper
- Is private disclosure possible in open source?
- Is there a way to combine pre-release benefits while avoiding post-release?

Embedded nature of components interesting