Zero Days, Thousands of Nights

The life and times of zero-day vulnerabilities and their exploits

Lillian Ablon

✉ lablon@rand.org

@lilyablon

RAND CORPORATION

black hat USA 2017
JULY 22-27
MANDALAY BAY

- Publicly available research on zero-day vulnerabilities and their exploits is sparse
- Common questions include:
  - **Life Status**: Is a zero-day vulnerability known by others?
  - **Longevity**: How long will a zero-day vulnerability remain undiscovered and undisclosed to the public?
  - **Collision Rate**: What is the percentage of vulnerabilities independently discovered and disclosed in a given time period?
- Answers can help inform decision makers regarding zero-days
- This research provides empirical analysis of zero-day vulnerabilities and their exploits

- Overview of Data
- Research Focus
- Analysis & Findings
- Implications & Recommendations

Overview of our data

207

exploits and their vulnerabilities

14

Year span (2002-2016)

Data consists of information about vulnerability class, source code type, exploit class type, vendor, product, exploit developer, and various dates (vulnerability discovery, exploit developed)

Overview of our data

207

exploits and their vulnerabilities

14

Year span (2002-2016)

BUSBY

Private research group, proxy for nation-state

Data consists of information about vulnerability class, source code type, exploit class type, vendor, product, exploit developer, and various dates (vulnerability discovery, exploit developed)

Data stats: our vulnerabilities are split up into three main types

Memory Corruption

110

- 7 subcategories
- Most common:
  - heap overflow (58)
  - stack overflow (40)

Memory Mismanagement

41

- 13 subcategories
- Most common:
  - null dereference (12)
  - information leak (4)

Logic

67

- 23 subcategories
- Most common:
  - race condition (20)
  - auth bypass (5)
  - privilege errors (4)
  - object injection (4)

Data stats: number of vulnerabilities per source code type

Closed	Open	Mix or N/A
123	74	10

Data stats: number of vulnerabilities found and exploited by vendor

Microsoft	Linux
55	39
Apple	**SUN/Oracle**
14	11

Other

88

- 64 vendors total
- Others include: Mozilla, LinkSys, Google, Adobe, etc.

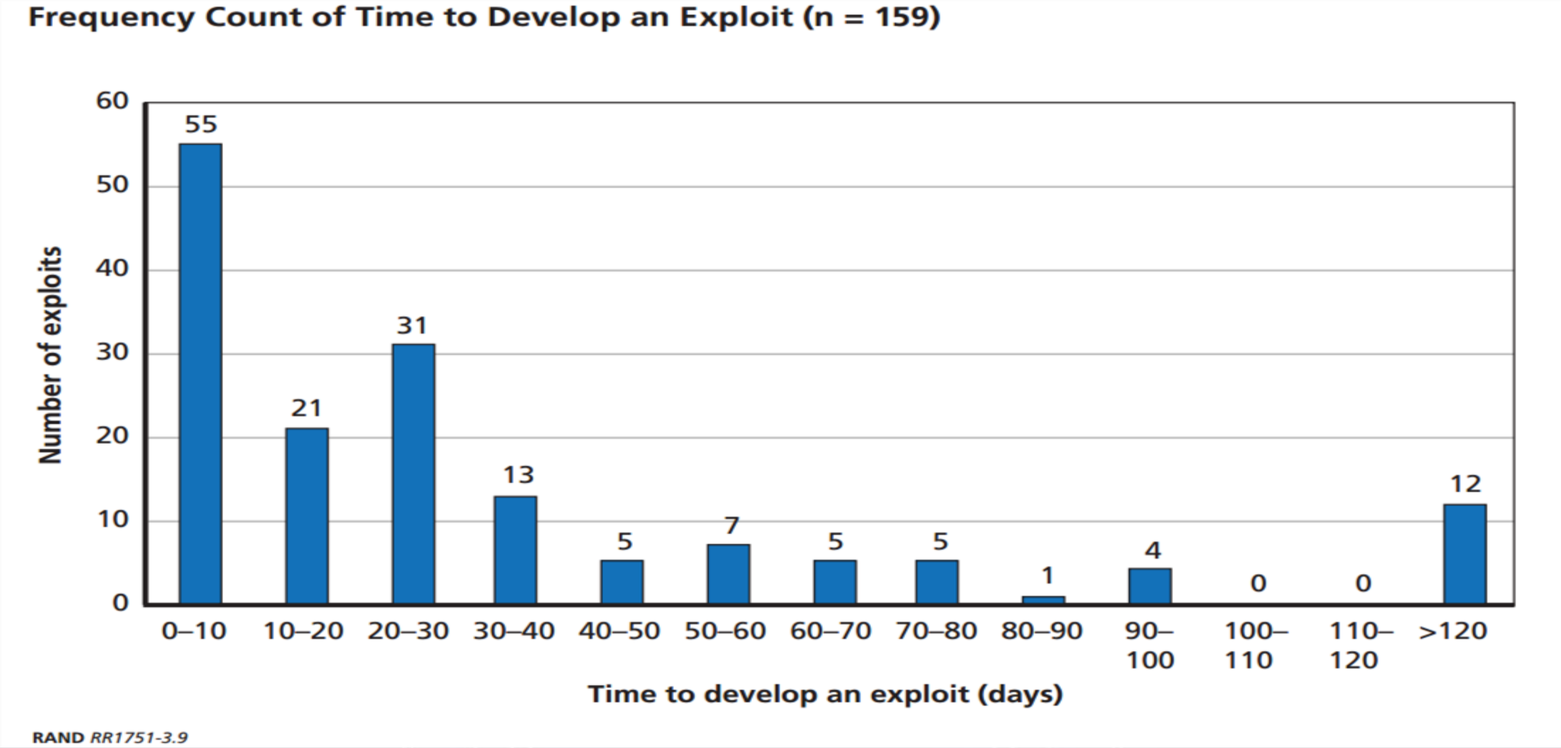
Data stats: number of exploits developed per exploit class type

Local	Client-side	Remote
76	25	71

Other observations about the data

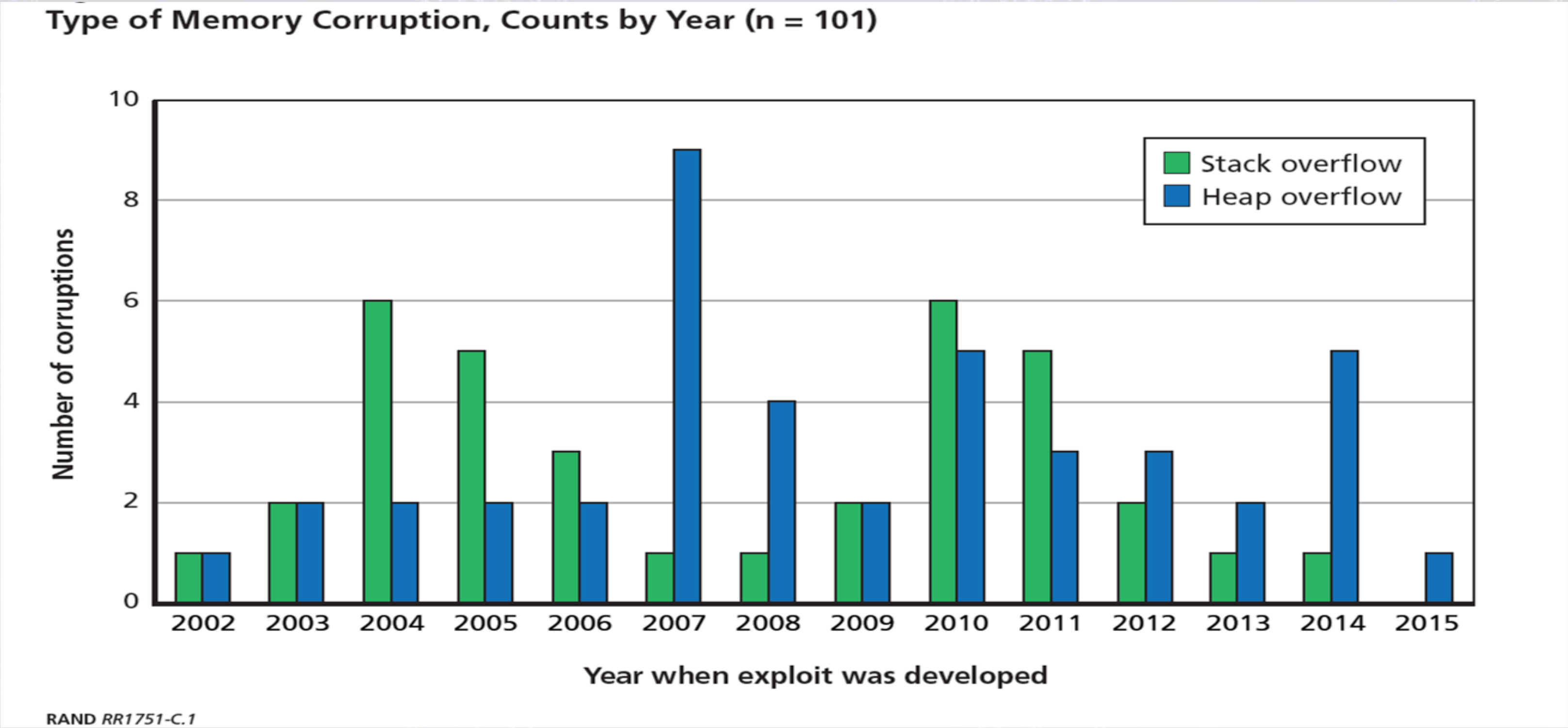
- 4% of the vulnerabilities in the dataset were purchased from an outside 3rd party
- Not all vulnerabilities were exploited
- CVEs do not always provide accurate and complete information about the severity of a vulnerability
- Virtual isolation (hypervisors or VMs) and anti-virus are not necessarily viable mitigations
- Other observations ...

Exploit Development time is relatively short



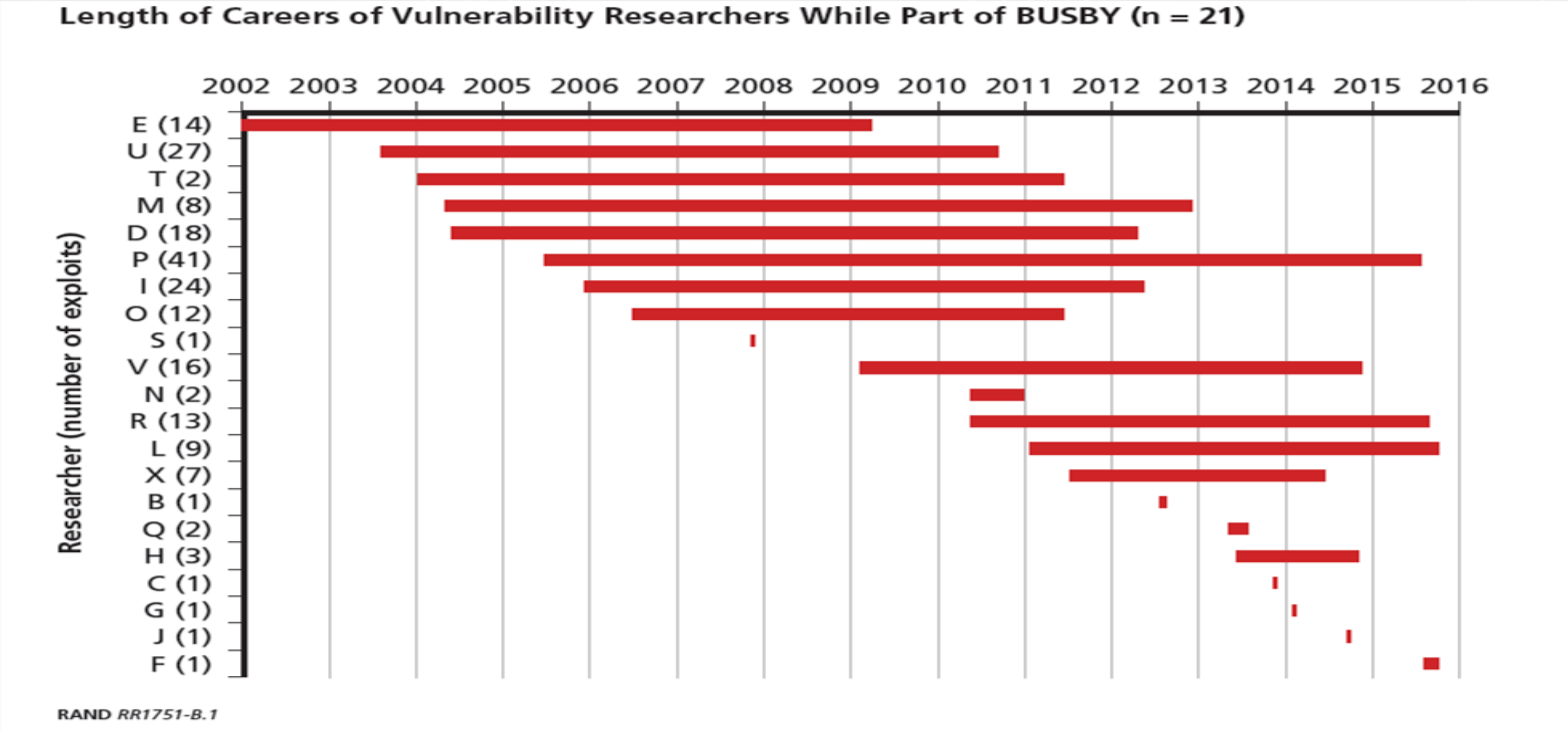
Over 70% of exploits are developed in a month (31 days) or less

Mitigations have affected exploitability (ex: heap vs stack overflow)



Mitigations introduced c. 2007 caused a shift in type of buffer overflow exploited

Exploit development career lengths vary



Low hanging fruit may account for a higher number of exploits developed early on

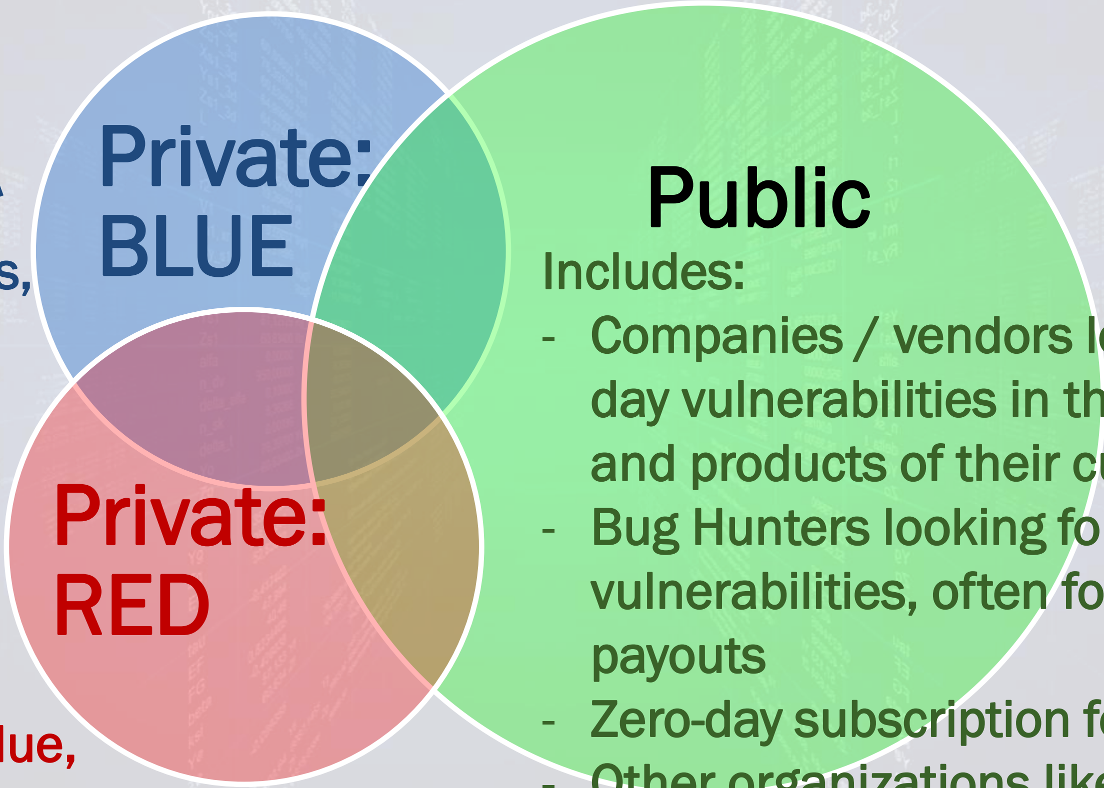
Caveats on the data

- Results from our data can be generalized only to similar datasets

- We are comparing private data to public data (ideal would be comparing multiple private datasets)

Various groups search for vulnerabilities

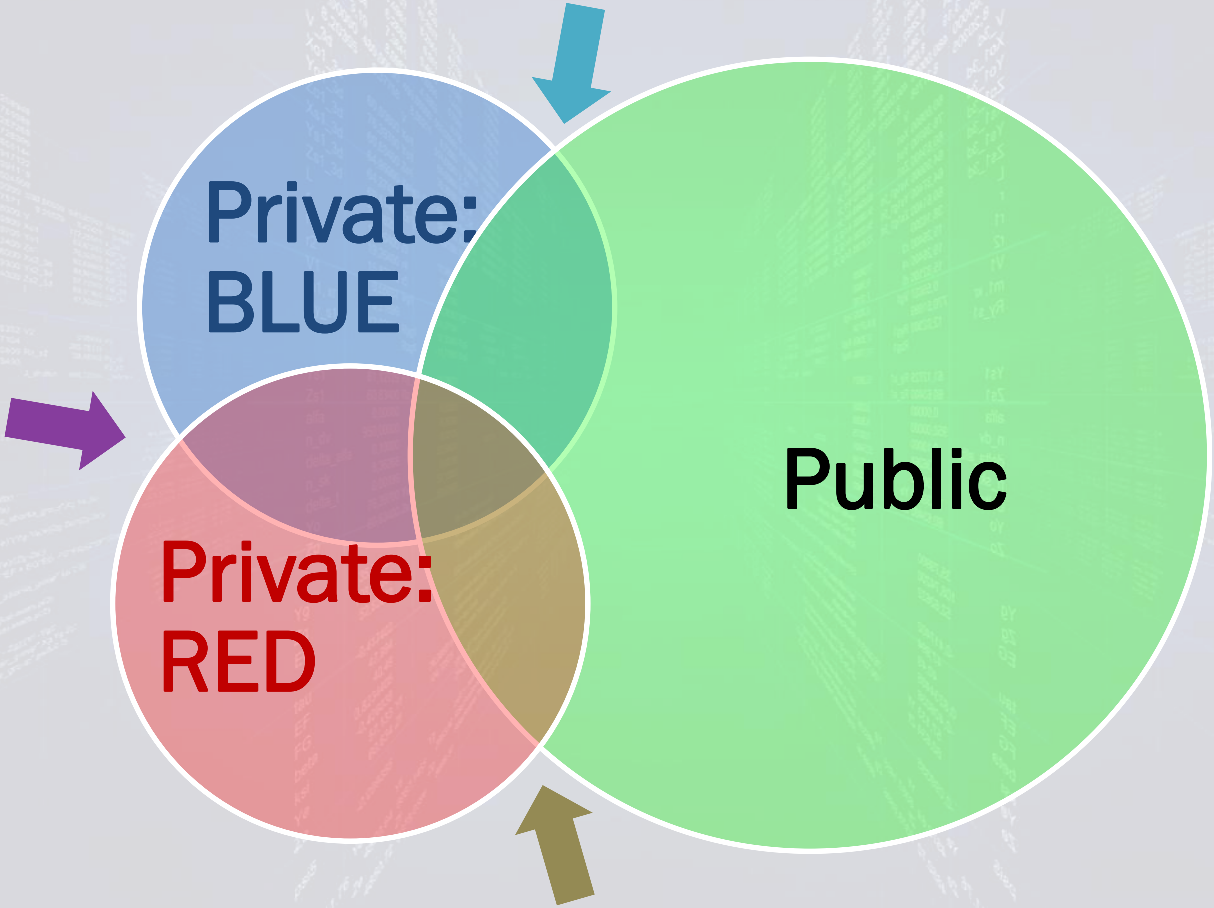
Governments, defense contractors, exploit developers, vulnerability researchers



Adversaries of Blue, Malicious Actors

Includes:
- Companies / vendors looking for zero-day vulnerabilities in their own products and products of their customers
- Bug Hunters looking for zero-day vulnerabilities, often for bug bounty payouts
- Zero-day subscription feed businesses
- Other organizations like Project Zero

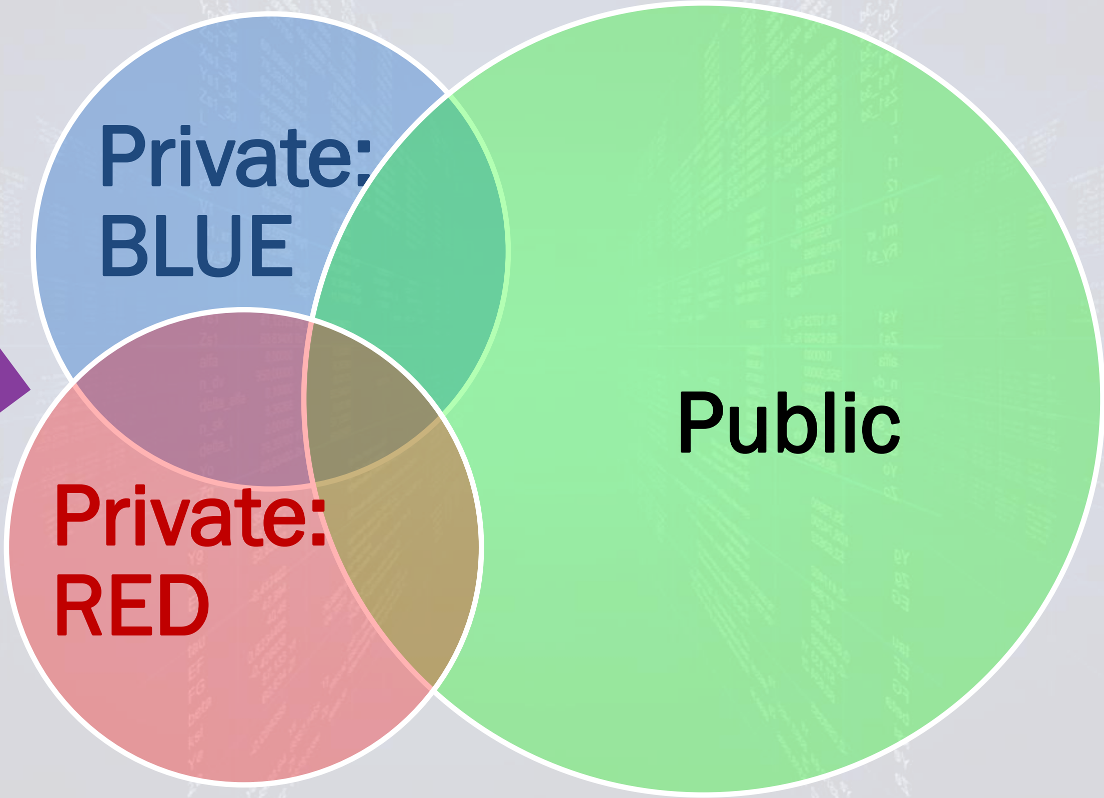
Some vulnerabilities are discovered by more than one group



A big unknown is the overlap between various groups

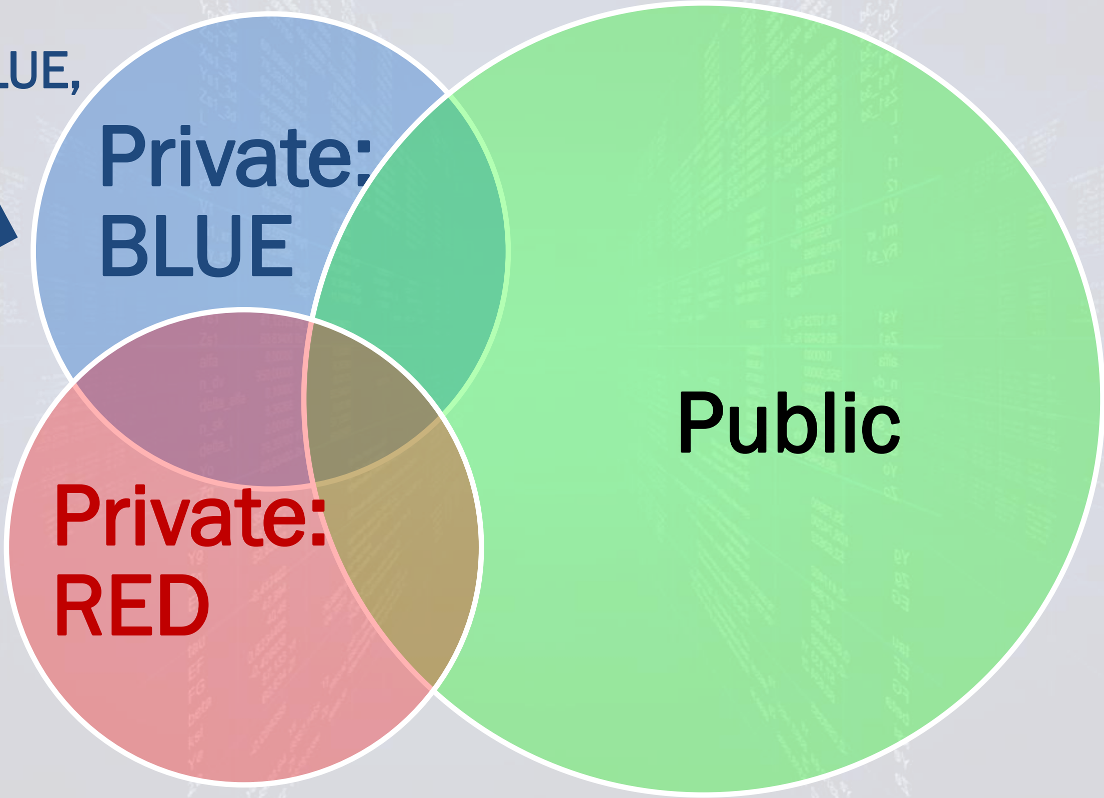
Vulnerabilities known to *both* BLUE and RED

disclosure by BLUE may strengthen BLUE's defensive posture



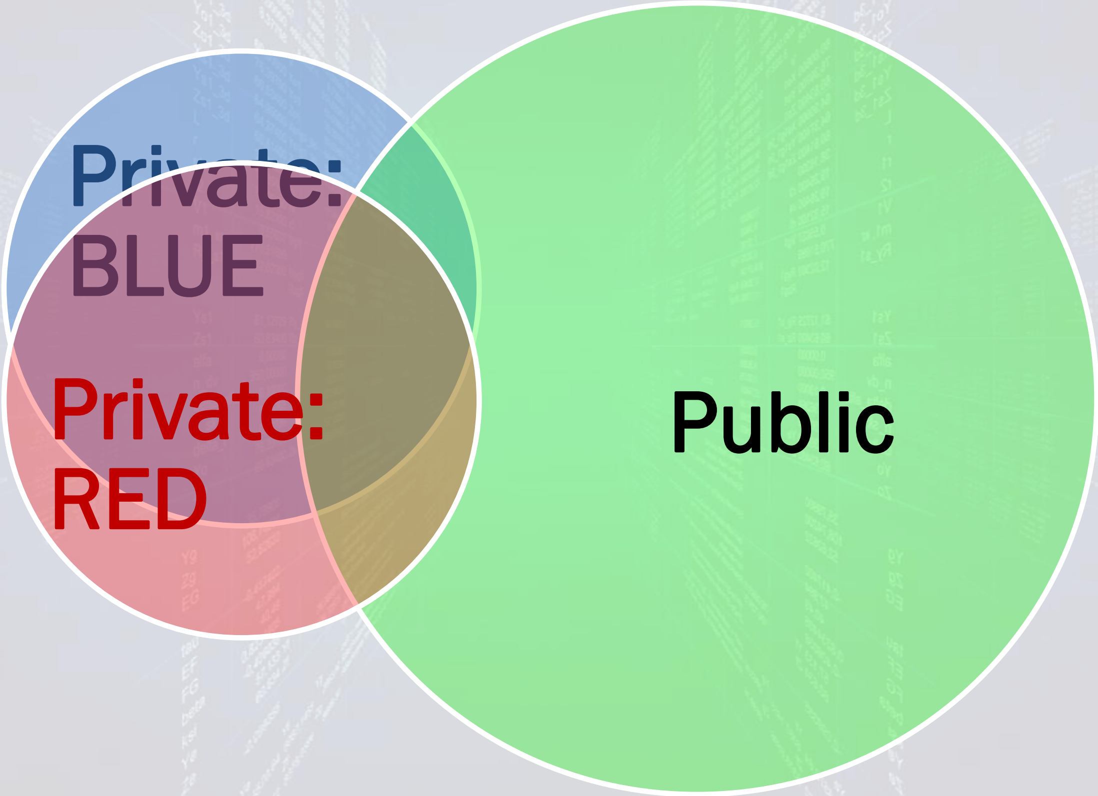
A big unknown is the overlap between various groups

Vulnerabilities known *only* to BLUE, and not to RED:

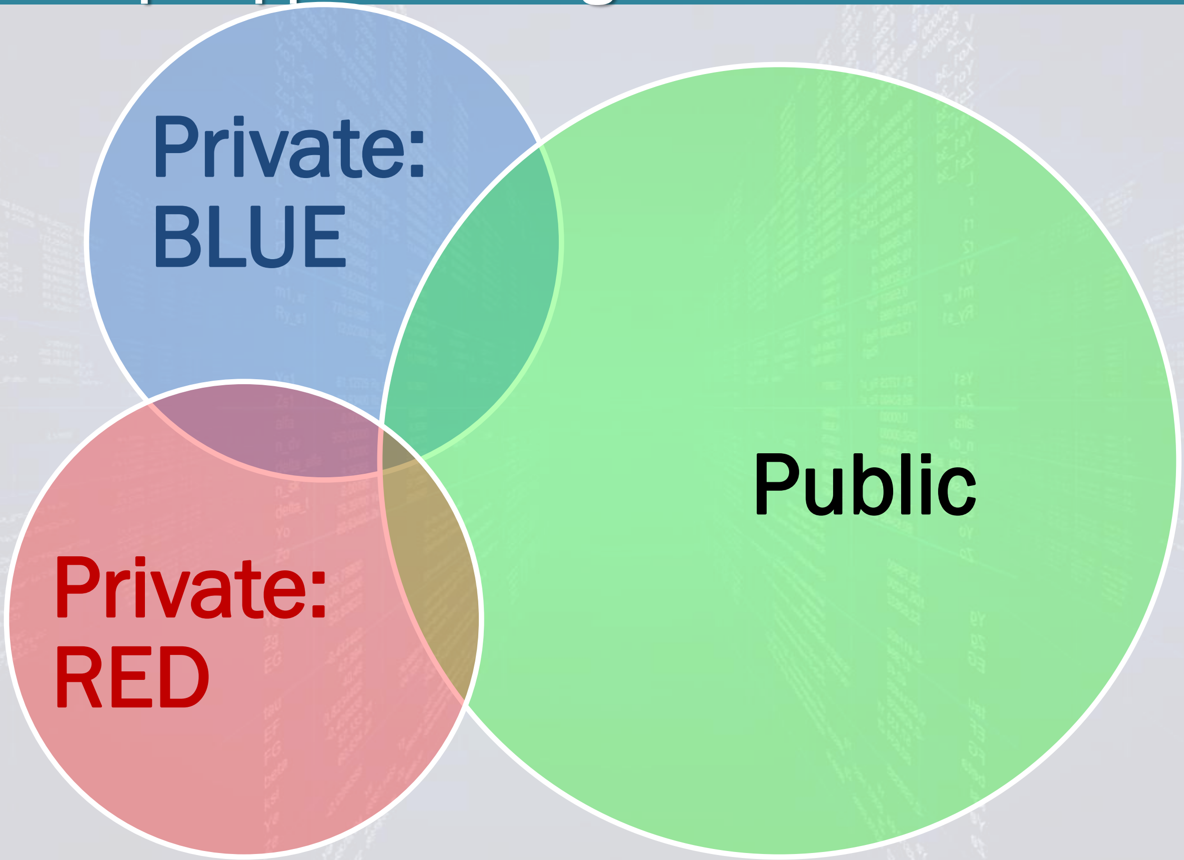


disclosure by BLUE may hinder BLUE's offensive posture

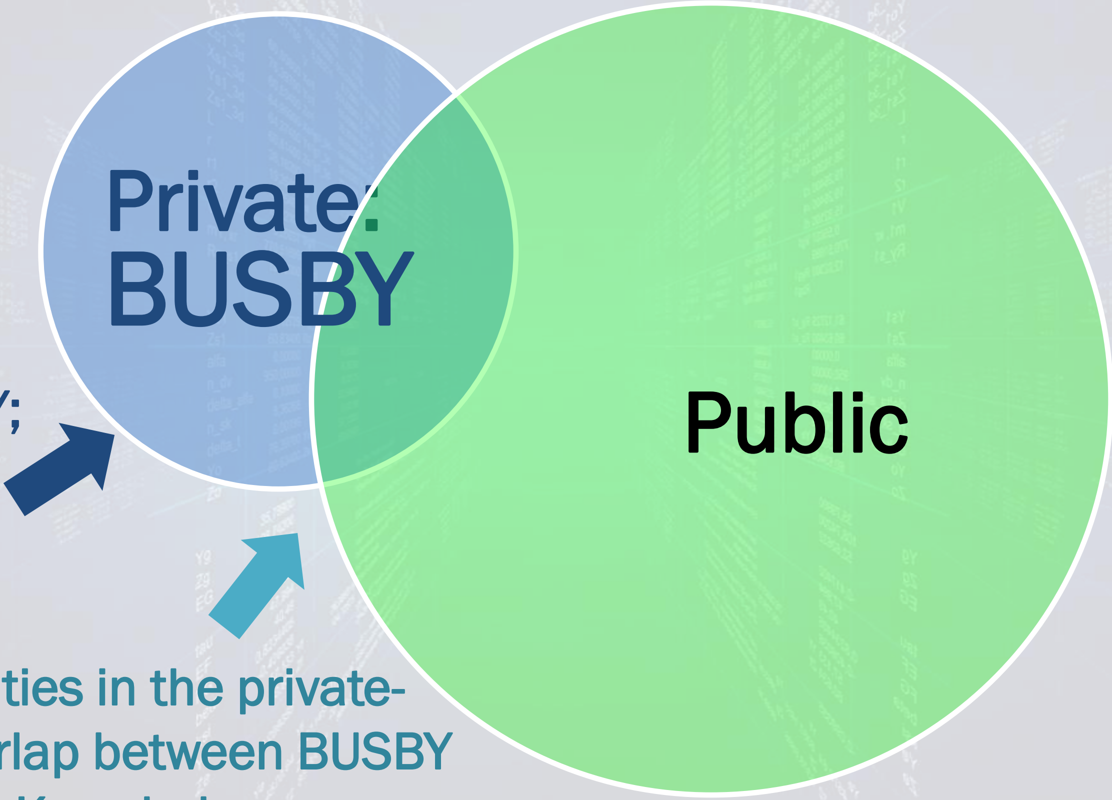
A large overlap supports an argument to disclose vulnerabilities



A small overlap supports an argument to retain vulnerabilities



We focus on zero-day characteristics in the public/private overlap



Vulnerabilities known to BUSBY; not in Public Knowledge

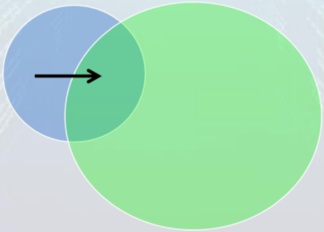
Vulnerabilities in the private-public overlap between BUSBY and Public Knowledge

We focus on zero-day characteristics in the public/private overlap

Life Status

Longevity

- Survival Rate
- Life Expectancy

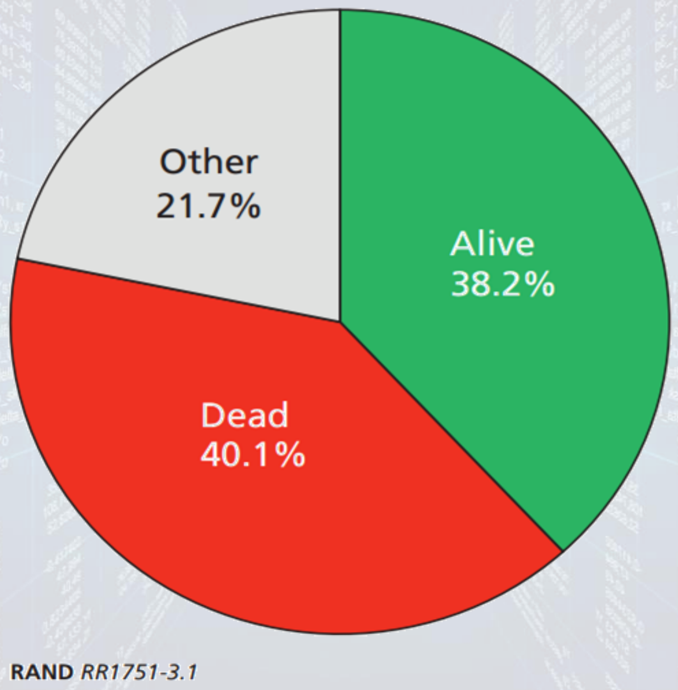


Collision Rate

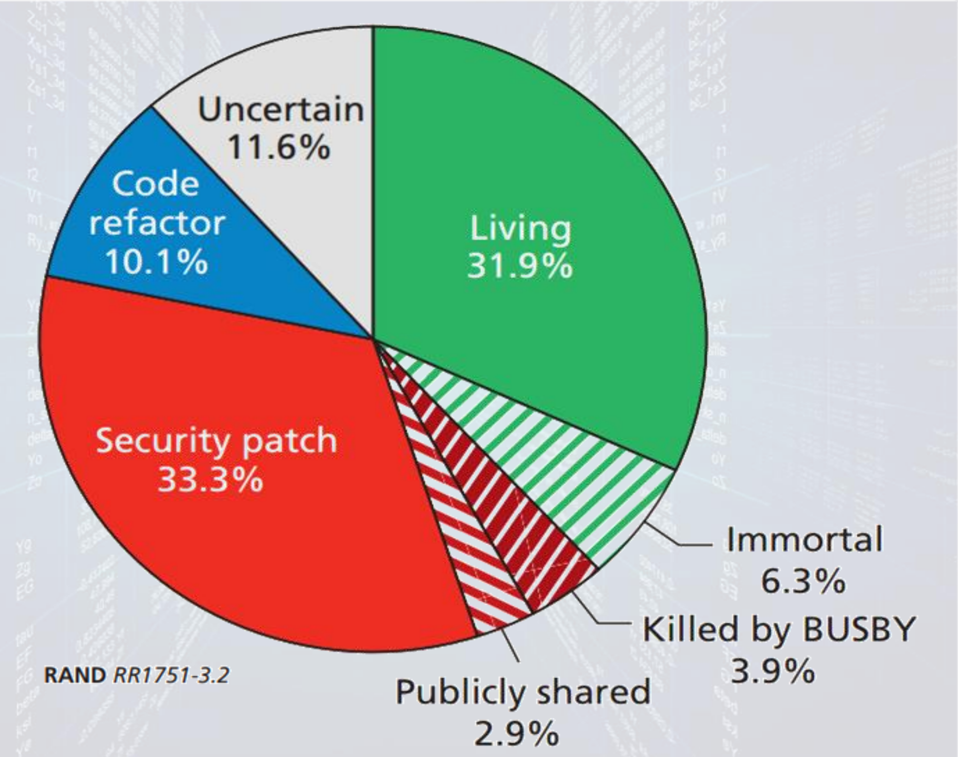
Research Question: What are various "life stages" a zero-day vulnerability can be in?

Metric: What proportion of zero-day vulnerabilities are:

- Alive (publicly unknown / blue)
- Dead (publicly known / teal & green)
- Somewhere in between



There is more granularity to a vulnerability being either alive or dead

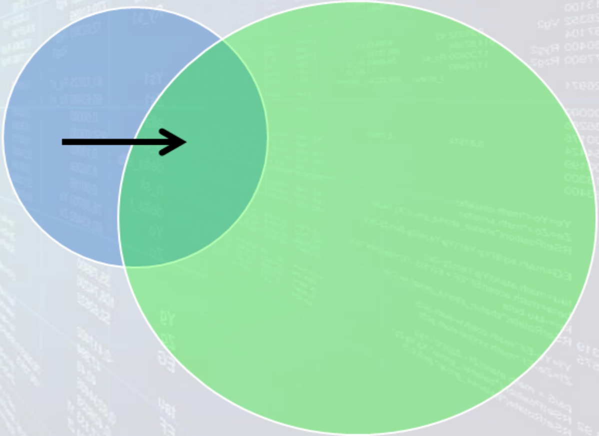


Labeling a vulnerability as either alive or dead is misleading and too simplistic

Research Question: How long will a zero-day vulnerability remain undiscovered and undisclosed to the public?

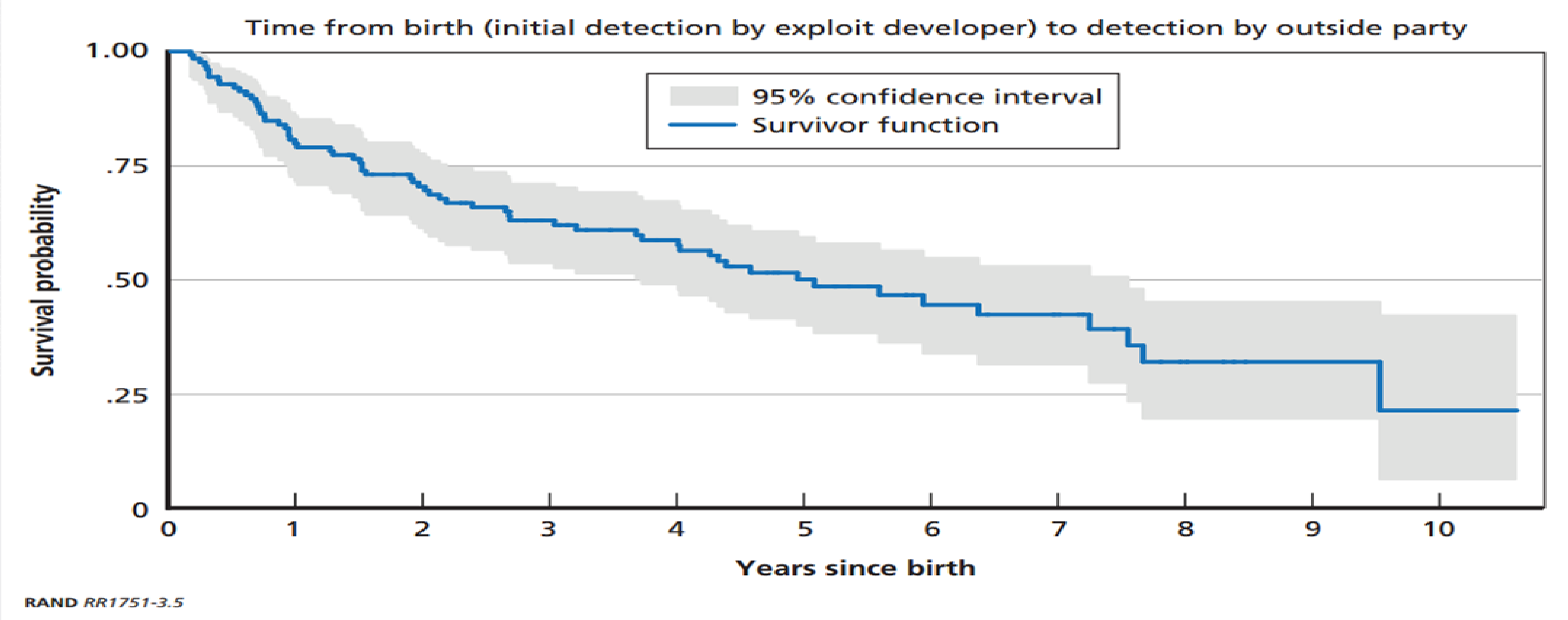
Metrics:

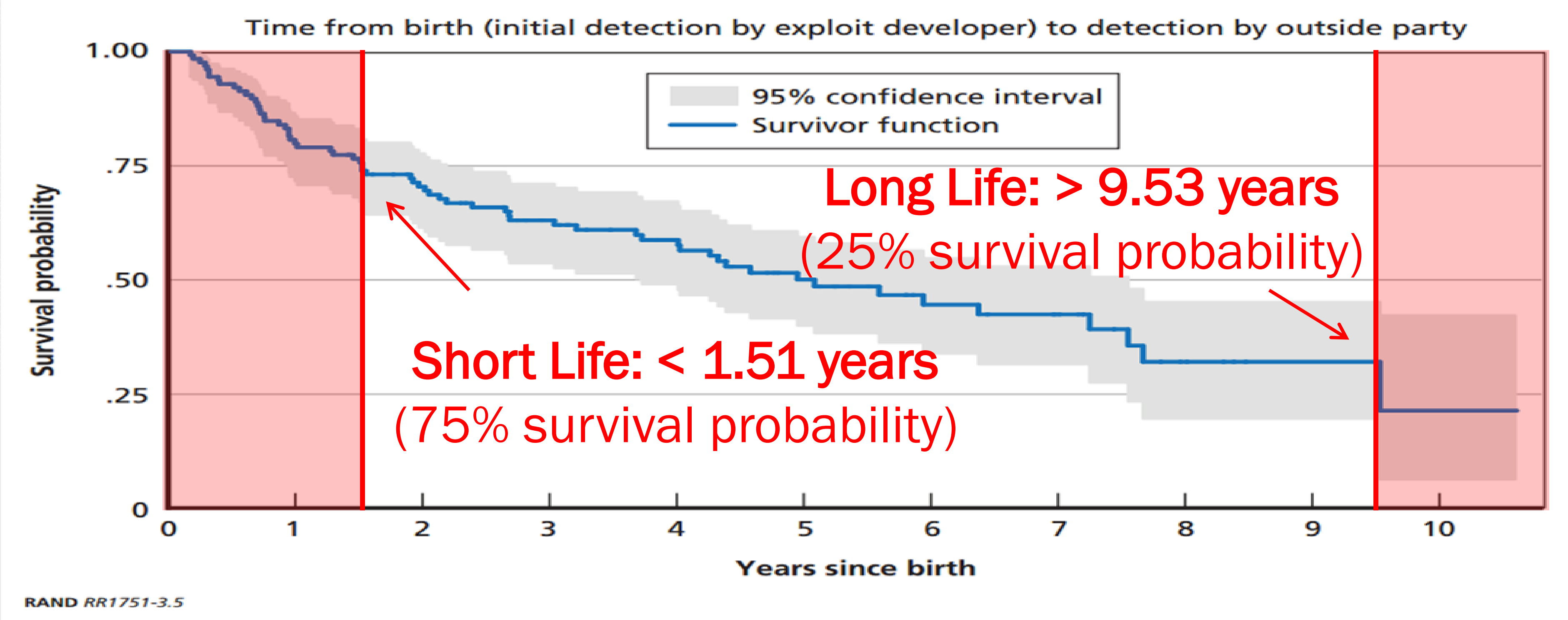
- What is a short and long life for a zero-day vulnerability?
- What is the average life expectancy of a zero-day vulnerability and its exploit?

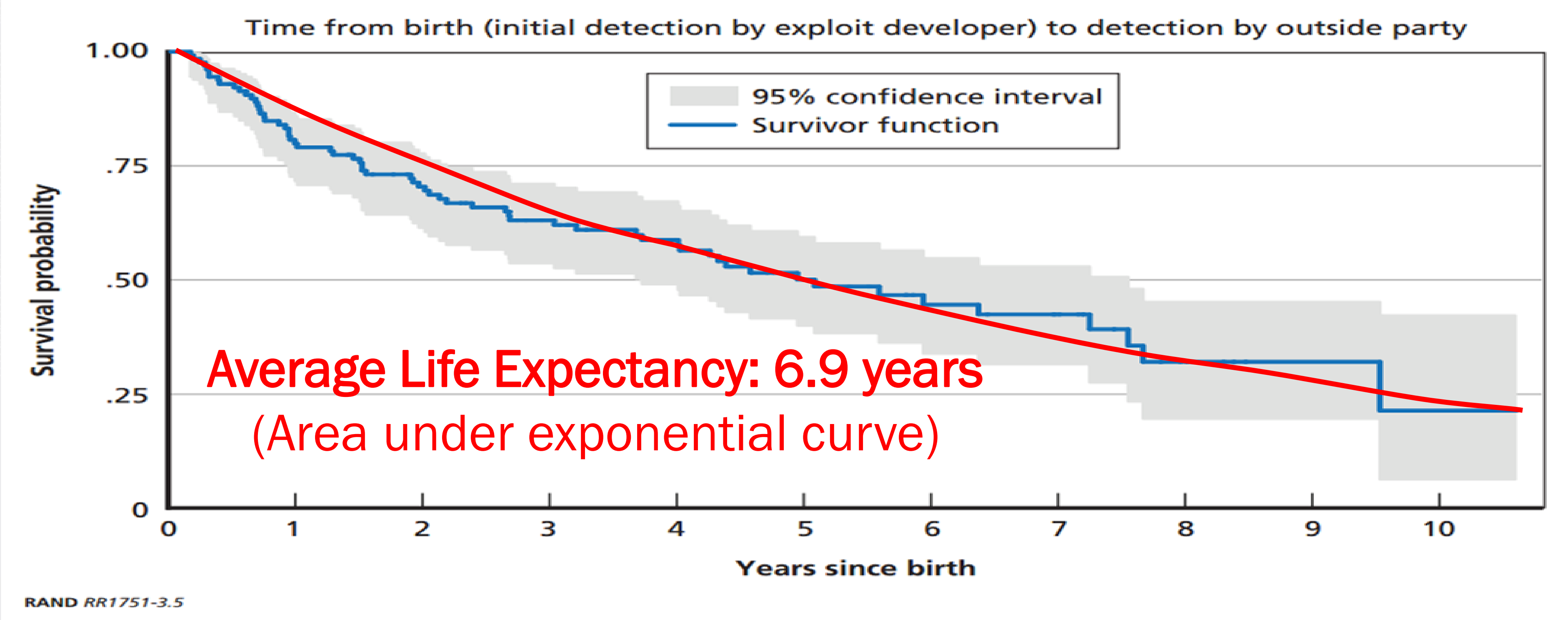


- We do not know what is going to happen to those vulnerabilities that are still currently alive
  - Calculating short life, long life, and average lifetimes requires taking into account alive vulnerabilities

- Kaplan-Meier analysis estimates the probability of surviving from some event of interest over time
  - Ex: For humans, the probability of someone having a heart attack
  - For vulnerabilities, the probability of dying and becoming publicly known

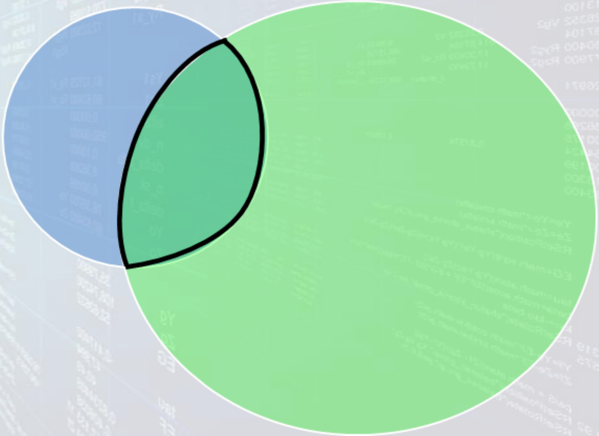

Time from birth (initial detection by exploit developer) to detection by outside party
95% confidence interval
Survivor function
Survival probability
1.00
.75
.50
.25
0
0
1
2
3
4
5
6
7
8
9
10
Years since birth
RAND RR1751-3.5


Time from birth (initial detection by exploit developer) to detection by outside party
95% confidence interval
Survivor function
Survival probability
1.00
.75
.50
.25
0
0
1
2
3
4
5
6
7
8
9
10
Years since birth
Long Life: > 9.53 years
(25% survival probability)
Short Life: < 1.51 years
(75% survival probability)
RAND RR1751-3.5

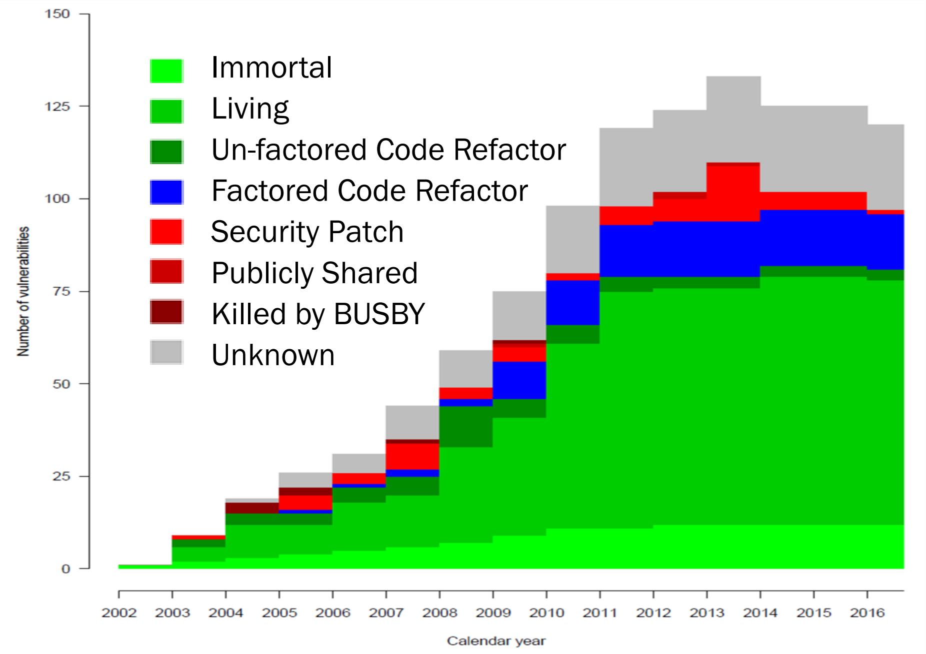

Time from birth (initial detection by exploit developer) to detection by outside party
95% confidence interval
Survivor function
Survival probability
1.00
.75
.50
.25
0
0
1
2
3
4
5
6
7
8
9
10
Years since birth
Average Life Expectancy: 6.9 years
(Area under exponential curve)
RAND RR1751-3.5

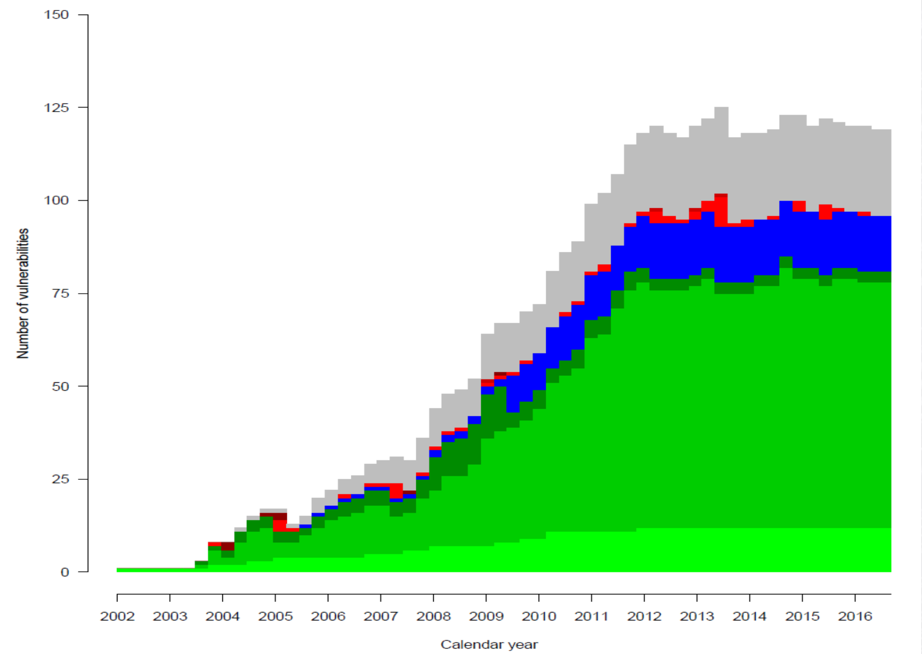
Research Question: What is the collision rate of zero-day vulnerabilities independently discovered and disclosed in a given time period?

Metric: What percentage of privately known vulnerabilities get independently rediscovered and publicly disclosed in a given time period?



- Choose a time interval (365 days, 90 days, 30 days, etc.)
- Over that time interval, new zero-day vulnerabilities are discovered and retained
- At the end of the time interval, examine how many have been found by others and publicly disclosed (i.e. died)
  - "Throw out" those that have died
  - Keep the ones that are still alive
  - Continue to discover and retain new ones until the end of the next time interval when re-evaluation begins again
- Collision rate: median percentage of those that died over all the time intervals





Time interval: 365-days
Collision rate: 5.7%

Time interval: 90-days
Collision rate: 0.87%

Time interval: All (14 years)	Time interval: 365-days	Time interval: 90-days
40%	5.7%	0.87%

Collision rates change significantly depending on the interval time

More research is needed to refine other analysis

- Characteristics of a vulnerability that indicate a long or short life*
- Average life expectancies based on vulnerability characteristic*
- Life expectancy variation based on birth year
- Collision rate variation based on vulnerability characteristic*
- Collision rate and timing for individual vulnerabilities
- Time to develop exploit based on vulnerability characteristic *
- Seasonality of vulnerability research
- Cost of developing an exploit

**No statistical significance found, likely due to limited data*

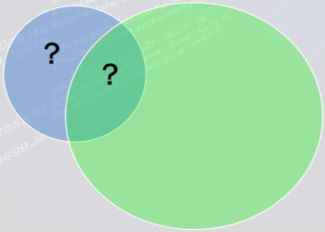
If you have data and would like to collaborate to refine this research, please contact me: lablon@rand.org or @lilyablon

Key findings (BlackHat Sound Bytes)

Life Status

7+ Categories

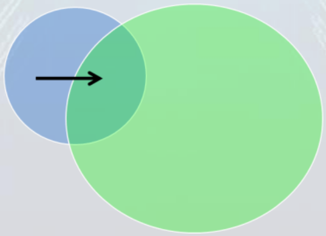
Labeling a zero-day vulnerability as either alive or dead can be misleading and too simplistic



Longevity

6.9 years

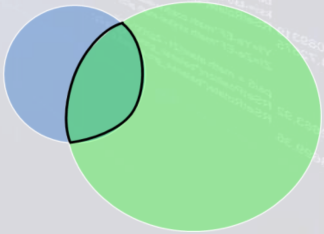
Zero-day vulnerabilities and their exploits have a rather long average life expectancy



Collision Rate

5.7% per year

Time interval examined can significantly change the percentage for likelihood of independent rediscovery



Report freely available at: http://www.rand.org/pubs/research_reports/RR1751.html

Implications of key findings and recommendations

For those **defensively** focused

- Refine tactical approaches:
  - Analyze previous versions of code that are still in heavy use (e.g., ICS)
  - Harness techniques of how offense finds vulnerabilities
  - Seek better options to detect vulns
- Consider strategic approaches: mitigation, containment, accountability, and a robust infrastructure of patching
  - Employ physical isolation
  - Account for software, devices, and removable media
  - Incentivize upgrading to new versions

For those **offensively** focused

- Retain a few vulnerabilities per particular software package
- Consider immortal or code-refactored vulnerabilities for operations
- Regularly revisit vulnerabilities thought to be unexploitable
- Plan for a specific vulnerability only for short-term planning operations; expand to *any* vulnerability may extend the timeline

Our findings can help inform retention v. disclosure discussions

Pro **retention**

- Long average lifetimes and relatively low collision rates may indicate that:

1. vulnerabilities are dense
    - The level of protection from disclosing a vulnerability may be modest
2. vulnerabilities are hard to find
    - There is a small probability of re-discovery by others

Pro **disclosure**

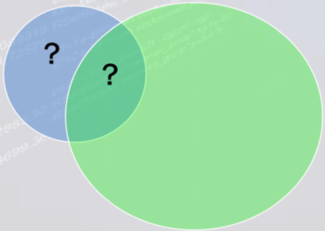
- Collision rates for zero-day vulnerabilities are non-zero
- A non-zero probability (no matter how small) that someone else will find the same zero-day vulnerability may be too risky

Key findings (BlackHat Sound Bytes)

Life Status

7+ Categories

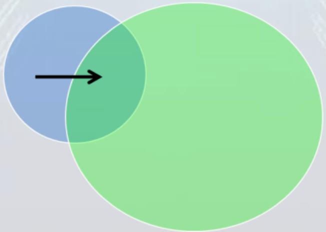
Labeling a zero-day vulnerability as either alive or dead can be misleading and too simplistic



Longevity

6.9 years

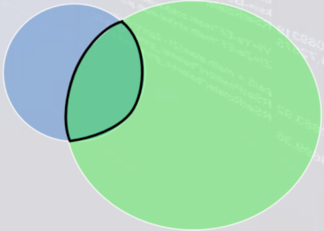
Zero-day vulnerabilities and their exploits have a rather long average life expectancy



Collision Rate

5.7% per year

Time interval examined can significantly change the percentage for likelihood of independent rediscovery

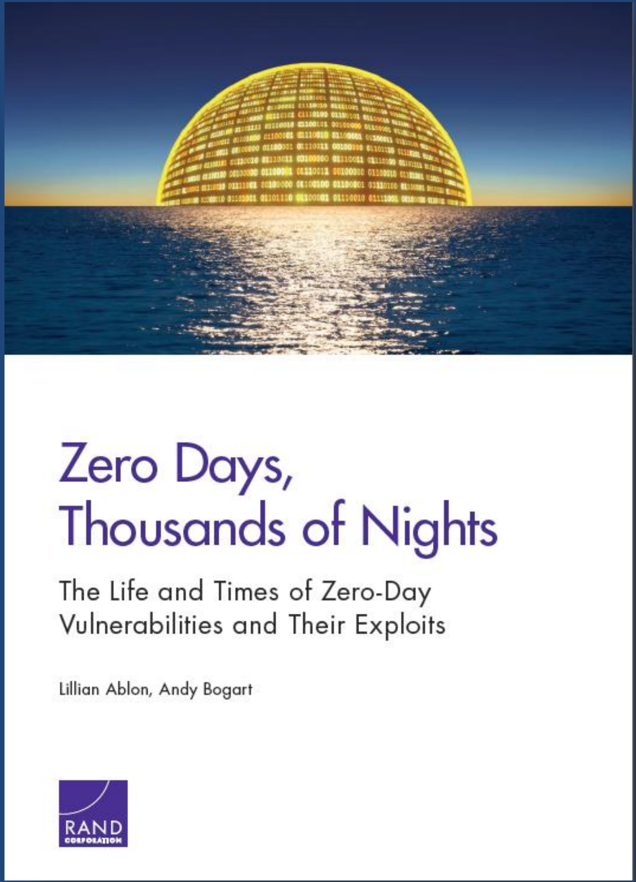


Report freely available at: http://www.rand.org/pubs/research_reports/RR1751.html

Thank you!

Lillian Ablon

lablon@rand.org
@LilyAblon



RAND CORPORATION

Report freely available at: http://www.rand.org/pubs/research_reports/RR1751.html