Analyzing 24 Years of CVD

Allen Householder
adh@cert.org
Agenda

Process & data overview
Cases & messages over time
Case duration distribution
Case size distribution
When does the work happen?
Observations on Case Complexity
The CVD Process

Legend

often same individual / organization

Role

relationship

optional relationship

https://resources.sei.cmu.edu/asset_files/SpecialReport/2017_003_001_503340.pdf
The Data

This is work in progress, all results are preliminary.

CERT/CC has been coordinating vulnerability disclosures since 1988.
  • Email-centered process, “hub and spoke” communication pattern
  • Messages sent/received as proxy measure of coordination effort

Database log of email sent/received by cert@cert.org about VU#nnnn and VR-nnn cases
  • Spans 1993-2017 (24 years)
  • 350k+ CVD-related email messages observed
  • 46k+ CVD cases observed
  • 2,300+ years of CVD embargo*

*sum across all domain-level participants
Cases, Messages per Quarter 1994-2017
Cases, Messages per Month 2009-2017
Cases, Messages per Month 2009-2017

(Same chart, but remove VU#582497)
How long do cases last?

- Most cases are 95% done around 50 days from start.
- 75% of cases are halfway done around 30 days from start.
- Half of all messages occur within 1 month of the case start.
- 80% of all messages occur within 100 days of the case start.
Case Sizes

Case Counts by Class (log count)

Message Counts by Class

"Case class" = $\log_4(n_{\text{messages}})$
Workload Distribution By Case Size

½ of messages are in cases with >64 msgs

>70% of messages are in cases with >3 msgs

CDF of messages across case classes

N_msg: 0-3, 4-15, 16-63, 64-255, 256-1023, 1024+
Case Size-Frequency

46k total cases observed
11k with >3 messages (top quartile)

Case size limited by timespan? Human capacity?
Larger cases possible when tech assisted disclosures
Case Start Relative to Date Public

- 9,600 cases with >3 messages and public base date
- 1,400 cases with ≥1 msg before public base date
- 8,200 cases with 100% of msgs after base date

Public Base Date (PBD) = min(date_public, date_first_published)
Case Midpoint Relative to Date Public

Relative date on which cases reached 50% of their total messages

Public Base Date (PBD) = \( \min(\text{date}_{\text{public}}, \text{date}_{\text{first published}}) \)

- 400 cases with \( \geq 50\% \) of msgs before public base date
- 9,200 cases with \( > 50\% \) of msgs after base date
- 8,200 cases with 100% of msgs after base date
- 1,000 cases with \( \geq 1 \) msg pre-PBD and \( > 50\% \) of msgs post-PBD
Relative date on which cases reached 95% of their total messages

Coordination does not end with disclosure!

Public Base Date (PBD) = min(date_public, date_first_published)
Case End (Effective) Relative to Date Public

Coordination does not end with disclosure!

A few cases are just wrapping up at >3 Years post-disclosure

Public Base Date (PBD) = min(date_public, date_first_published)
Case Breakdown

Most of the coordination doesn’t even happen pre-disclosure
Underlying Mechanism for Case Size: Multiparty

Most cases involve 1 or a few parties. Large multiparty are where all the complexity happens.

For each of the largest cases, the daily traffic spikes have a characteristic size, appears to correlate to the # of parties involved.
### Limiting Factors for Case Size

<table>
<thead>
<tr>
<th>Factor</th>
<th>Potential causes of limit</th>
</tr>
</thead>
</table>
| Timespan                      | • Vendor responsiveness to creating patches (*This is a good thing!*)  
                                | • Attention span of orgs before other work takes precedent  
                                | • Reporters sometimes under inflexible timelines |
| Identifying affected vendors  | • What products contains libfoo?  
                                | • What vendors are affected by a vul in libfoo?  
                                | • What vendors implemented this protocol? |
| Number of vendors involved    | • Contact management (acquisition & maintenance)  
                                | • Communication channel efficiency (hub & spoke, tools)  
                                | • How many people can keep a secret for how long? |

These are all about *efficiency* and *efficacy* of vulnerability response processes, driven by information *availability* and *utility*. 
Parting Thoughts

CVD doesn’t end with public disclosure.
  - Most of the coordination work actually happens after public disclosure

“Average case” is not a useful concept for capacity planning
  - Large cases are rare, but dominate the day-to-day work

Case complexity is driven by the number of participants involved

Case sizes appear to be limited by organizational factors
  - There might be an upper limit to how big a coordination can be before it’s better to just go public

Got Data?

- This is ongoing research work at CERT
- We are looking for CVD metadata from other orgs
- Minimum required: (Case ID, Message Timestamp)
- Contact us if you have data you can share.
Contact Info

Allen Householder
Team Lead, Threat Ecosystem Analysis
CERT Division

Email: adh@cert.org
Twitter: @adh