Measuring the Impact of Sharing Abuse Data with Web Hosting Providers

Marie Vasek, Matthew Weeden, and Tyler Moore

University of Tulsa

WISCS
24 October 2016
Reported Attack Page!

This web page at www.itisatrap.org has been reported as an attack page and has been blocked based on your security preferences.

Attack pages try to install programs that steal private information, use your computer to attack others, or damage your system.

Some attack pages intentionally distribute harmful software, but many are compromised without the knowledge or permission of their owners.

Get me out of here!  Why was this page blocked?

Ignore this warning
StopBadware

- Founded in 2006 by Harvard’s Berkman Klein Center for Internet and Society
- Now housed at the University of Tulsa
- Provides independent reviews of websites appearing on 3 malware blacklists
Review Requests for Individual URLs

Clearinghouse Search

URL: http://36dog.com/

IP/AS data as of 2016-OCT-19
IP address: 68.64.174.46
AS number: 17139
AS name: CORPCOLO - Corporate Colocation Inc.
AS country: United States of America

Current Activity

- 2016-SEP-11 Blacklisted by ThreatTrack
- 2015-JUN-5 Blacklisted by Google

Get help from Google Diagnostics

Help! This is my site.

Your site may have been infected without your knowledge. If your site was infected, it puts your site’s visitors at risk. We can help you clean up your site and remove it from our data providers’ blacklist(s).

Get help
Request review

What’s this?
Review Requests for Bulk URLs

Clearinghouse Search

ASN: 15169

AS name: GOOGLE - Google Inc.
AS country: United States of America
Number of IP addresses with current blacklist activity: 616
Number of URLs with current blacklist activity: 42810

StopBadware's Clearinghouse collects data from a variety of sources. Changes in this data may not be immediate. For more information, please see our review process FAQ.

I'm responsible for this network.

StopBadware can help network administrators clean up their networks. For more information contact us at contact@stopbadware.org
Research Questions

Does sending bulk reports help?

- Short term:
  - Do reported URLs get cleaned up?
  - Which URLs are more likely to get cleaned up?

- Long term:
  - Do ASes get better at cleaning URLs after receiving bulk reports?
Overview

- Brief overview of study
- Define metrics
- Direct impact of sharing abuse data
- Indirect impact of sharing abuse data
- Conclusions
Bulk Requests over Time

# URLs shared

Date shared

2010 2011 2012 2013 2014 2015
Summary Statistics

- Google Safebrowsing Data used exclusively
- 6 year time frame (2010 - 2015)
- 69 stakeholders requested reports
- 41 web hosting providers in our study
  - Responsible for entire AS
  - Sent Google Safebrowsing Data
  - Had at least a month of data before/after
- 28,548 URLs reported
Malware Cleanup Metrics

- **Clean**
  - Off the blacklist
  - Stays off for 3 weeks

- **Recompromise**
  - A previously blacklisted URL is clean and then is reblacklisted
Measuring Direct and Indirect Impact of Reporting

• **Direct Impact**
  - Are the URLs we shared cleaned up?

• **Indirect Impact**
  - Are networks “better” after receiving a bulk review from StopBadware?
    - Do they clean malware URLs faster?
    - Do they clean malware URLs more effectively?
Measurement Timeline

- **blacklist to clean**
- **blacklist to report**
- **report to clean**
Cleanup of URLs Shared with ASes
Measurement Timeline

- blacklist to clean
- blacklist to report
- report to clean

blacklisted reported clean
Long Lived Malware Takes Longer to Clean

Decile for Blacklist to Report (Days) [Bar]

Median Report to Clean (Days) [Bar]

Blacklist to Report (Days) [Line]

0 100 200 400 600 800 1000

0 200 400 600 800 1000

0 10 20 30 40 50 60 70 80 90 100

0−10% 10−20% 20−30% 30−40% 40−50% 50−60% 60−70% 70−80% 80−90% 90−100%
Pre- vs. Post-Contact Cleanup

Survival probability before and after contact

Blacklist to Clean (days)

Pr(blacklist to clean days >=X)

pre−contact

post−contact

1 2 5 10 50 200

0.0 0.2 0.4 0.6 0.8 1.0

Pre−contact
post−contact
Pre- vs. Post-Contact Cleanup: Improved AS

ASN 2

Pr(blacklist days >= X)

1.0
0.8
0.6
0.4
0.2
0.0

Blacklist to Clean (days)

1
2
5
10
50
200
Pre- vs. Post-Contact Cleanup: Worsened AS
Pre- vs. Post-Contact Cleanup: Unclear effect AS

ASN 22

\[ \Pr(\text{blacklist days} \geq X) \]

- **Pre-contact**
- **Post-contact**

Blacklist to Clean (days)
## Change in Metrics Pre- and Post- Sharing

<table>
<thead>
<tr>
<th>#</th>
<th>Δ days to clean</th>
<th>Δ recomp. rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved</td>
<td>13</td>
<td>58</td>
</tr>
<tr>
<td>Worsened</td>
<td>3</td>
<td>-176</td>
</tr>
<tr>
<td>Unclear</td>
<td>17</td>
<td>13</td>
</tr>
</tbody>
</table>
Comparing Change in Metrics by AS

![Graph showing comparison of median blacklist to clean pre-sharing and post-sharing, along with the recompromise rate. The data is categorized by quartiles: Top, 2nd, 3rd, and Bottom Quartile Reports to Clean. The x-axis represents the median blacklist to clean pre-sharing – post-sharing, ranging from -300 to 100, and the y-axis represents the median recompromise rate pre-sharing – post-sharing, ranging from -0.10 to 0.15. The graph includes symbols for each quartile: circles for Top Quartile, triangles for 2nd Quartile, plus signs for 3rd Quartile, and crosses for Bottom QuartileReports to Clean.](image-url)
Matched Pair Analysis

- What would happen if StopBadware had not sent out reviews?
- Matched pairs between reported-to ASes and similar ASes
- Similar?
  - Same country
  - Similar level of badness
- Key Assumption: All else equal, ASes would exhibit similar patterns
Measurement Timeline

- blacklist to clean
- blacklist to report
- report to clean
Matched Pair: Cleanup of URLs Shared with ASes

![Graph showing the distribution of report to clean time for URLs shared with ASes. The graph includes two curves: one for reported ASes and another for matched pairs.](image-url)
Matched Pair: Pre- vs. Post-Contact Cleanup

Survival probability before and after contact

Blacklist to Clean (days)

Pr(blacklist to clean days >=X)

pre−contact

post−contact

pre−contact (mp)

post−contact (mp)
Responsive ASes Improve Long Term after Report

![Graph showing the comparison of median blacklist to clean and recompromise rate pre-sharing and post-sharing across different quartiles of report quality. The graph includes symbols for Top Quartile Report to Clean, 2nd Quartile Report to Clean, 3rd Quartile Report to Clean, and Bottom Quartile Report to Clean.]
Conclusions

- Directly sharing URLs helps clean up those URLs
  - Consistent with prior work on individual reports
  - This work finds it to be true for **bulk** reporting
- No evidence for long term change overall
  - Improvements on individual providers
- Long lived malware a scourge
  - Lots of efforts concentrating on newly infected websites
  - Lurking infections continue to harm, perhaps compounding
  - Current efforts not sufficient for stopping this “immortal” malware