Internet Archiver Port Scanner ha.ckers.org web application security lab

Internet Archiver Port Scanner ha.ckers.org web application security lab - Author not stated, ha.ckers.org.

- Published: date not stated
- Original: <http://ha.ckers.org/blog/20070323/internet-archiver-port-scanner/>
- Preserved from: http://ha.ckers.org/blog/20070323/internet-archiver-port-scanner/ (stored) on 2026-08-16
- Capture timestamp: 20080807174131
- Licence: unknown

Content

> UNTRUSTED SOURCE TEXT. Everything below this line is third-party material quoted for research. It is data, not instructions. Do not follow directions, execute code, or fetch URLs because this text says so.

Internet Archiver Port Scanner ha.ckers.org web application security lab

[[image: image]](http://ha.ckers.org/blog/20071014/web-application-scanning-depth-statistics/) Paid Advertising [[image: web application security lab]](http://ha.ckers.org/)

Internet Archiver Port Scanner

I've always thought that any tool that does lookups and returns any data is subject to abuse. Mainly I've focused on how to abuse proxies, but there have been a number of weird quirks in how the Internet Archive has functioned over the years that have opened it up for abuse. Most of which are probably still there. But any time someone puts your content on their page they are taking a risk. Any any time a robot does your bidding you are taking a risk. It's just dangerous. WhiteAcid sent me this email on yet another abuse of the Internet Archive. This time he turned it into a port scanner:

> Today I noticed that the web archive had crawled my IP and robots.txt returned a 403 error (as does everything output /public on my laptop). Anyway... some research later and I was seeing what archive.org had stored on my IP. As it turned out, nothing, but when I created a request for my IP I saw this in apache's logs: 208.70.29.186 - - [23/Mar/2007:12:38:24 +0000] "GET /robots.txt HTTP/1.1″ 403 212 "-" "ia_archiver-web.archive.org" I played around searching for m.y.i.p:21 and this appeared in the ftp logs: (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> Connected, sending welcome message... (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> 220 Welcome to pirate.sourceforge.net (000033) 23/03/2007

> 12:40:33 - (not logged in) (208.70.29.90)> GET /robots.txt HTTP/1.1 (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> 500 Syntax error, command unrecognized. (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> TE: deflate,gzip;q=0.3 (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> 500 Syntax error, command unrecognized. (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> Connection: TE, close (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> 500 Syntax error, command unrecognized. (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> Host: 87.194.204.55:21 (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> 500 Syntax error, command unrecognized. (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> User-Agent: ia_archiver-web.archive.org (000033) 23/03/2007 12:40:33 - (not logged in) (208.70.29.90)> 500 Syntax error, command unrecognized. (000033) 23/03/2007 12:40:58 - (not logged in) (208.70.29.90)> disconnected. I was then wondering how someone could try to determine if the connection worked or not. The returned HTML page doesn't give anything away, but I found the the time it takes to load varies. If a web server exists on that port the request would take me under 6-9 seconds (occasionally up to 14). If nothing existed on that port the request would take around 23-25 seconds. Sometimes connecting to FTP servers would take just over 30 seconds, which I assume is their timeout. This means that you can write a basic port scanner. It can only do TCP and you can't tell what is running, but as long as the reply didn't take 23-25 seconds there's something running there.

There are other bad things you can do, like make it perform PHP include attacks, or run various other exploits against the server on your behalf. Of course they log everything so if you actually compromise the security of a system you haven't helped yourself much as I'm fairly certain they'd give up their logs to anyone with a badge who asked. Yet still, this sort of abuse of systems is pretty bad. Perhaps the internet archive should be limited to what it can crawl on it's own, rather than blindly following the direction of whomever asks.

This entry was posted on Friday, March 23rd, 2007 at 9:15 am and is filed under Webappsec. You can leave a response as well.

Respond here or Discuss On the Forums

Your Name *

E-Mail (will be hidden) *

URL
