Malware in Archives, Museums, & Libraries

Jonathan Farbowitz (@jfarbowitz) / New York University , Moving Image Archiving & Preservation

Why Collect Malware?

Hex dump of the Blaster Worm with a message for Bill Gates

Part of the texture of digital life and "a pervasive feature of the internet" (Pennock)

Enhances histories of software development, Net Art, hacking, and political activism—all of the social (and anti-social) uses of computing

Can be evidence of state repression

May provide information about the habits or lives of individual computer users

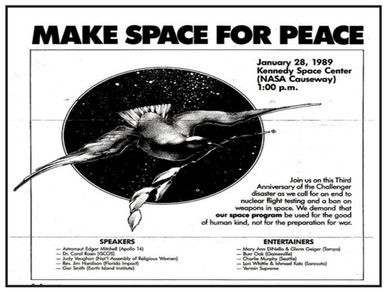
<Attila_Nagy> Computer viruses are almost as old as personal computers themselves...Within each code is a story about its author, about the time it was written, and about the state of computing when it wrought havoc upon our hard drives.

To date, no cultural heritage institution is committed to collecting and preserving malware.

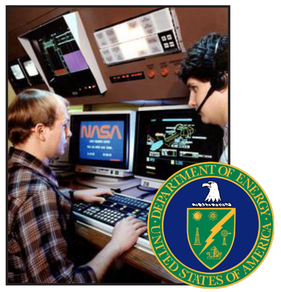
The WANK Worm

Released in 1989, the WANK worm spread first through NASA computers and then throughout the world. WANK is considered the first instance of hacktivism and would be an important piece of malware to collect. The identity of the worm's creator(s) are unknown, although its coding was traced to Melbourne, Australia. (Dreyfus and Assange)

Payload screen of a computer infected with the WANK worm. Despite the message, no files were deleted.



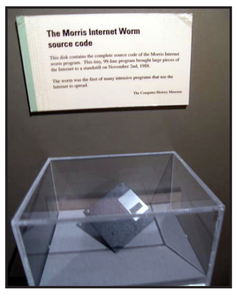
Flyer from Florida Coalition for Peace and Justice. In 1989 the group protested the launch of the Galileo space probe. WANK's creator(s) were likely sympathetic to this group.



Nearly all of NASA's computers on the SPAN network were infected. The US DOE was affected as well. The worm spread as far as Switzerland and Japan over DECNet.

Preservation Strategies

Preservation strategies will depend on research goals. ***Do researchers want to analyze the code, see malware demonstrated, or do they want to see malware infections "in the wild"?***



Source code of the Morris Worm at the Computer History Museum. Saving physical objects like disks is an extremely limited preservation strategy.

> Save source code, compiled code, or 'snapshots' of code

> Take screen captures and recording video demonstrations

> Record oral histories of those affected or of malware creators

> Collect ancillary materials, including posts on security message boards, emails, articles, & websites

> Save snapshots of command-and-control servers of a botnet

VIRUS BULLETIN

Preserving malware will likely require an **"electronic art approach," (Besser)** acknowledging its variability and saving related contextual items. Archivists and librarians will need to:

Employ various preservation strategies simultaneously

Save extensive metadata and documentation, enough to emulate period hardware/software in the future

Make decisions on a case-by-case basis

Sources

Besser, Howard. "Longevity of Electronic Art." Accessed May 7, 2015. http://besser.tsoa.nyu.edu/howard/Papers/elect-art-longevity.html.

Dreyfus, Suelette, and Julian Assange. Underground: Tales of Hacking, Madness, and Obsession on the Electronic Frontier. Kew, Australia: Mandarin, 1997.

Gruning, Jane. "Rethinking Viruses in the Archives." Poster presented at the Archival Education and Research Institute, 2012. https://www.ischool.utexas.edu/~janegru/images/Gruning_AERI2012.pdf.

Pennock, Maureen. "Web Archiving." Digital Preservation Coalition, March 2013. http://dx.doi.org/10.7207/twr13-01.

Nagy, Atilla. "14 Infamous Computer Virus Snippets That Trace A History Of Havoc | Gizmodo Australia." Accessed April 25, 2015. http://www.gizmodo.com.au/2013/07/14-infamous-computer-virus-snippets-that-trace-a-history-of-havoc/.

Acknowledgments

Chris Avram, Jefferson Bailey, Snowden Becker, Howard Besser, Jeff Chiu, Suelette Dreyfus, Simson Garfinkel, Julia Kim, Steve Lamb, Dave Riordan, Jason Scott, & Doug White

Collecting Institutions Encountering Malware

Archives, museums, and libraries encounter various types of malware when accessioning born-digital material (such as hard drives, disks, and email) into their collections.

"Current digital archival practice often treats virus checking and quarantine as an unproblematic aspect of ingesting digital objects... often before any formal appraisal is done." (Gruning) Currently, there is no standard procedure for documenting the infection or removal of malware so that this information is available to researchers.

Are "Cleaned" Files Authentic?

Collecting institutions strive to preserve bit-for-bit copies of hard drives and disks when they process them using digital forensics tools. However, by removing malware, removing viral code from files, or quarantining infected files, bits are being altered. ***Are these actions altering the authenticity of the items accessioned?***

If malware must be removed, how can it be documented in a standardized way that is accessible to researchers?

Issues with Anti-Virus Software

False positives:
Archivists could be altering files that are not even infected

Incorrect identification of malware or variants:
Institutions don't know what they actually have

Different classification systems between AV companies:
Metadata is not standardized

<Jane_Gruning> Archives are creating a gap in the history of computers and their use in our society—a gap that we could potentially avoid... Archivists need to rethink how we, as a profession, are addressing this issue.

Risk Assessment

Saving malware should not be taken lightly. Conducting a risk assessment is necessary. However, the full capabilities of a particular piece of malware are often unknown. Some considerations:

> Malware can inhibit access to disks or files

> It can damage the integrity of files

> It can allow unauthorized access to computers or networks

> Sophisticated malware ≈ weapons

> Preserving contemporary malware may inadvertently aid law enforcement

Institutions can start by collecting malware that is historically significant and well-understood and whose effects are innocuous then move on to harder cases as they gain more experience.

A Viral Dark Archive?

If disk images of infected items are saved, where will they be stored safely? Jane Gruning and others have suggested creating a dark archive of non-network-connected storage and keeping malware there for a period of years until it becomes less of a threat due to operating system and hardware obsolescence.