The New York Times

January 27, 2009

A Tool to Verify Digital Records, Even as Technology Shifts

http://www.nytimes.com/2009/01/27/science/27arch.html?scp=2&sq=hash mark&st=cse

By JOHN MARKOFF

Simple-to-use digital technology will make it more difficult to distort history in the future.

On Tuesday a group of researchers at the University of Washington are releasing the initial component of a public system to provide authentication for an archive of video interviews with the prosecutors and other members of the International Criminal Tribunal for the Rwandan genocide. The group will also release the first portion of the Rwandan archive.

This system is intended to be available for future use in digitally preserving and authenticating first-hand accounts of war crimes, atrocities and genocide.

Such tools are of vital importance because it has become possible to alter digital text, video and audio in ways that are virtually undetectable to the unaided human eye and ear.

The researchers said history was filled with incidents of doctoring, deleting or denying written records. Now, they say, the authenticity of digital documents like videos, transcripts of personal accounts and court records can be indisputably proved for the first time.

“The closest analogy are the revisionist histories of the Holocaust, where there are assertions that people weren’t put in camps and put in ovens,” said Batya Friedman, a professor of computer science at the Information School at the University of Washington. “It doesn’t take a rocket scientist to say that in a period of time some people will say there really weren’t 800,000 people who were massacred with machetes.”

Designing digital systems that can preserve information for many generations is one of the most vexing engineering challenges. The researchers’ solution is to create a publicly available digital fingerprint, known as a cryptographic hash

mark, that will make it possible for anyone to determine that the documents are authentic and have not been tampered with. The concept of a digital hash was pioneered at I.B.M. by Hans Peter Luhn in the early 1950s. The University of Washington researchers are the first to try to simplify the application for nontechnical users and to try to offer a complete system that would preserve information across generations.

Both because of the rapid pace of innovation and the tendency of computers to wear out in months or years, the likelihood that digital files will be readable over long periods of time is far less certain even than the survival of paper documents. Computer processors are quickly replaced by incompatible models, software programs are developed with new data formats, and digital storage media, whether digital tape, magnetic disk or solid state memory chips, are all too ephemeral.

Several technologists are already grappling with the evanescent nature of digital records.

Danny Hillis, a computer scientist, helped found the Long Now project in 1996, warning about the possibility of a "digital dark age." The group is now designing a clock that will "tick" annually and that is designed to have a life span of 10,000 years. It is intended as a counterpoint to the "faster/cheaper" ethos of today's increasingly computerized world.

Mr. Hillis has argued that before the rise of digital information people valued paper documents and cared for them. Since then, there has been progressively less attention paid to the preservation of information. Now information is routinely stored on media that may last for only several years.

To that end, another computer scientist, Brewster Kahle, founded the Internet Archive in 1996 in an effort to preserve a complete record of the World Wide Web and other digital documents. Similarly, in 2000 librarians at Stanford University created LOCKSS, or Lots of Copies Keep Stuff Safe, to preserve journals in the digital age, by spreading digital copies of documents through an international community of libraries via the Internet.

However, Ms. Friedman distinguishes her design work from those who have focused on the simple preservation of digitized materials. Instead, she said she was trying to design complete digital systems that would play a role in strengthening social institutions over time by creating a digital historical record that offered continuity across multiple life spans.

"Building a clock is iconic," she said. "What is really different is that we are trying to solve socially significant, real-world problems."

Because problems like genocide, H.I.V. and AIDS, famine, deforestation and

global warming will not be solved in a single human lifetime, she argues that information systems designed to ensure continuity across many generations are a necessity.

To ground the group's research in a real-world situation, the researchers began by building an archive of video interviews with the judges, prosecutors and other members of the International Criminal Tribunal for Rwanda. The goal was to design a system that would ensure that the information was secure for more than a century.

Last fall Ms. Friedman traveled with a group of legal experts and cinematographers to Arusha, Tanzania, where the tribunal is based, and to Kigali, Rwanda, to conduct video interviews.

After capturing five gigabytes of video in 49 interviews, the group began to work on a system that would make it possible for viewers to prove for themselves that the videos had not been tampered with or altered even if they did not have access to powerful computing equipment or a high-speed Internet connection.

Despite the fact that there are commercial applications that make it possible to prove the time at which a document was created and verify that it has not been altered, the researchers wanted to develop a system that was freely available and would stand a chance of surviving repeated technology shifts.

At the heart of the system is an algorithm that is used to compute a 128-character number known as a cryptographic hash from the digital information in a particular document. Even the smallest change in the original document will result in a new hash value.

In recent years researchers have begun to find weaknesses in current hash algorithms, and so last November the National Institute of Standards and Technology began a competition to create stronger hashing technologies. The University of Washington researchers now use a modern hash algorithm called SHA-2, but they have designed the system so that it can be easily replaced with a more advanced algorithm.

Their system will be distributed as part of a CD known as a "live CD," making it possible to compute or verify the hash just by inserting the disk in a computer. The disk will also include software components that will make it possible to view documents and videos that may not be accessible by future software.

The problem is complex, said Michael Lesk, a professor in the department of library and information science at Rutgers University, because not only must you be able to prove that the information has not changed in its original format, but you must also be able to prove that once the format is altered, the original digital hash is still valid.

The Long Now Foundation is developing a software tool to easily convert documents between digital formats, said Stewart Brand, a co-founder of the project.

“The idea is to be able to change anything into anything else,” he said.