ЦИФРОВА ФОРЕНЗИКА В АРХИВИТЕ

проф. д-р Юри Тодоров, СУ

Цифровата форензика е относително нова област, която се занимава с възстановяване и анализиране на данни от цифрови формати с намерението да разкрие съдържание и автентичност извън частните приложения в областта на компютърното хакерство и киберкриминалистиката. Понастоящем цифровата форензика се развива в най-разнообразни направления, включително в сферата на електронните документи и архиви. Архивната цифрова форензика има пряко отношение към проблемите на документалното историческо наследство. Основната ѝ задача се състои в пресъздаване и възстановяване на съдържанието от файлове, записани върху различни компютърни памети в най-разнообразни формати. Първостепенно форензиката се ориентира към документи, които оригинално съществуват в електронен вид. Преобладаващата част от електронните документи днес са създадени електронно и попадат в т. нар. категория "born digital"[1]. Цифровата форензика предлага подходящи средства, методи и инструменти за възстановяване на съдържанието на електронните документи в цялост, след като са били извлечени от цифровия източник. По-точно, методите на цифровата форензика позволяват възстановяването на изтрити файлове, особено интересни са възможностите за реконструкция на заличени версии от електронно създадените ръкописи. Възникват определени морално-етични проблеми както и херменевтични тълкования около документалното наследство с оглед на ролята на модерната архивистика при определяне на границите между личното и публичното във веригата фондообразувател – архивист – общество.

Още през 1985 г. Маккензи определя обсега на електронно генерираното съдържание като текст, който включва вербална, визуална,

звукова и цифрова информация, под формата на изображения, печатни произведения и музика, звукозаписи, на филми, видео и всякакви компютърно съхранени данни[2]. Съществото на тази дефиниция, наред със споменаването на понятието цифрови данни, е пълното освобождаване от намерението за библиографско копиране на мисълта и заместването му с материалните подробности около начините на съхранение, което стои в основата на форензиката.

Изследването на Рос и Гоу[3] от 1999 г. е друга отправна точка за съвременната цифрова форензика. По-новите публикации в областта се отнасят до актуални носители и системи, а в споменатото съобщение се прави обзор на методите с редица приложни примери, които се използват за възстановяване на цифрови данни от исторически носители. И наистина, по-вероятно е архивистът да работи върху проблем, който е свързан с остарели формати и системи отколкото с възстановяването на данни от актуални носители. Интересен пример се цитира от времето на обединението на Германия, когато във Федералните архиви постъпват електронни носители от бившата ГДР. Данните са били недостъпни, не поради физическо разрушаване, а заради различията в хардуера, софтуера, форма́та и кодирането на документите. Компютърните системи в Източна Европа, които са използвани през 80-те години на миналия век, са били деривати на западни компютри от 70-те. Отбелязва се заниженият качествен контрол при производството на цифровите носители, особено на твърдите дискове, при които неравностите в повърхността на магнитния слой повишават вероятността от загуба на информацията. Съществените изводи, които са били направени тук са, че надеждното съхранение на данните трябва да се прави заедно с подходящата документация. Става ясно, че архивирането на дигитални материали има свои частни проблеми, отнасящи се до форма́та на данните, програмите и документацията за тях, които се решават с помощта на нетрадиционни методи (например с привличане на предишни системни администратори).

Практиката на цифровата форензика се разпростира между абстракцията и индивидуализацията, между изменчивостта и стабилността. Според Фармер и Венема изменчивостта е артефакт на абстракцията, който прави компютърните системи използваеми. Според правилата на изменчивост вероятността за съхранение на информацията расте с увеличаването на мимолетните подробности в противовес на деструктивните процеси[4]. Колкото и да се стремим да съхраняваме всички сведения

наведнъж, невъзможно е да записваме промените в процеси и файлове в реално време, защото докато запазваме данни в една част от паметта, данните в другата и част се променят. От известния тест на Тюринг[5] знаем как да установим интелигентността на една машина. Форензиката е своеобразен тест на Тюринг, защото проверката на информацията, съхранена в компютърна система се свежда до изводи относно достоверността на данните: дали информацията отразява действително състояние или вървим по следите на нещо, което е било недобросъвестно променено.

Електронните документи изправят архивите пред разнообразни предизвикателства, от начините за съхранение на документалното наследство в електронен вид, а това са етични и методологически проблеми, до техническите въпроси, които касаят целостта, достъпността, цената на цифровото съхранение и възстановяването на информацията. Цифровите носители, които постъпват в електронния архив за съхранение, са както в исторически така и в съвременни формати. Стратегиите за съхранение на данните, които архивът преследва, отделят внимание на редица фактори. Започва се от намирането на подходящ хардуер за достъп до съответните цифрови носители и се стига до разпознаването на файлови системи и формати. Възможна ситуация е в архива да постъпят документи на цифрови носители (флопи дискове, оптически дискове, външни и вътрешни харддискове и т. н.). Или цяла компютърна системи, която трябва да се изследва и анализира. Начинът, по който информацията се съхранява на носителя, се определя от файловата система. Съществуват разнообразни файлови параметри в зависимост от използваната при тяхното създаване и жизнен цикъл файлова система. Като примери можем да споменем дължината на имената на файловете както и знаците, които се използват в тях. От гледна точка на архивното съхранение имената на файловете трябва да останат непроменени заради принципите, които се спазват при създаването им във фондообразувателите. Технологията, която тук предлага компютърната форензика, е произвеждането на огледален побитов образ на данните от оригиналния носител. С това се редуцират проблемите, произтичащи от употребата на различни файлови системи. Но когато този огледален образ се превърне в съставна част от цифровия репозиториум в архива, той се включва в друга сложна файлова структура. На лице е едно външно ниво на огледалния файл, във вътрешността на който са запазени индивидуални файлове в оригиналната файлова система. Когато в архива постъпват пълни

компютърни конфигурации като физическа среда заедно с файловете с данни, софтуера, който е използван за създаването им, и с контекстни сведения на системно ниво, на реконструкция подлежат всички подробности за създателя на документите и неговия начин на работа. Възстановяват се следите от възникването на документа до неговия окончателен вид. Компютърът е работна среда, взаимодействаща с други компютри и виртуални мрежи, които не са постъпили в архива. Приложение намират неинвазивните технологии за регистрация, при които информацията се извлича при оптимални за оригиналните носители условия, без опасност за техния интегритет.

Проблемите около историческите носители и архивирането на електронни документи са в унисон с развитието на компютърните технологии. Цифровата форензика предлага средства и методи за по-ефективна и надеждна работа в електронния архив като се започне от регистрацията на цифрово създадени документи, определянето на оригиналния им формат и условия на създаване и се стигне до взаимодействията на документите във виртуална среда. Миграцията на файлове в актуални формати е една възможна стратегия за работа. Паралелно обаче трябва да се запазва оригиналният побитов образ на цифровия носител, който да съпровожда преобразувания в актуален формат файл. В образа на оригиналния носител са включени всички скрити и изтрити файлове и данните запазват тяхната пълна функционалност. Чрез емулация в електронния архив се пресъздадат условията на работа от исторически компютърни системи върху съвременен хардуер. Така документите стават достъпни за изследване в оригинални условия, отговарящи на тези при тяхното създаване. Файловете на цифрови носители варират от елементарни документи с текст на една страница, без специално форматиране, до по-сложни конструкции като хипертекстове, цели уебсайтове или йерархични бази данни. В същото време един елементарен наглед документ може да включва персонализирани елементи или скрити сведения, които се отразяват върху начините на дългосрочно архивиране и достъп. Някои автори използват променливи кодове в Уърд за поставяне на допълнителни сведения в колонтитулите на документа. Тези променливи кодове се актуализират при всяко отваряне или отпечатване на документа. Това е още един пример в подкрепа на методите на цифровата форензика чрез побитово регистриране на образа от оригиналния носител. Създаването на образ

от оригинала в един единствен файл, който съдържа 1000 оригинални файла, организирани в някаква сложна йерархия, е много по-лесно отколкото копирането им поотделно и пресъздаването на логическата структура. По този начин се улеснява и извличането на метаданни към архивираните файлове. Ясно е, че образът на цифровия носител представлява абстракция или своеобразна интерпретация на физическите особености, които съществуват в реалната медия и че образът е само едно съответствие на реалния артефакт.

В известна степен историческото наследство е резултат от взаимодействието на правителства, организации, обществени групи, на семейството и отделните личности, според начините по които те общуват помежду си. Компютърните файлове също са част от някаква сложна система, която определя вътрешните отношения. Съхранението на електронни документи не означава да запазим само единичния обект. Интересни са най-вече неговите отношения към други обекти, значението на цифровия обект като част от сложно организирания процес в пълния му жизнен цикъл. Именно тази обвързаност на отделния файл в конкретна система, независимо дали това е файлова система или нарочно създадена стратегия в рамките на сложна виртуална среда или мрежа, прави отделния файл незаменим и уникален. От друга страна сме длъжни да отчитаме възможностите, които Интернет предлага за разпространение и достъп на електронните документи, независимо от техния формат – текстове, изображения и звуци, голяма част от които са пуснати в обръщение в мрежата непосредствено след тяхното създаване (напр. видеоклипове и снимки в YouTube и Twitter). Ако не сме в състояние да запазим подобни цифрови обекти, бихме загубили извори, историческата значимост на които в последствие би се оказала съществена. Множеството отделни елементи могат да се превърнат във важни фрагменти при осмислянето на сложната цялост.

Основно свойство на един документ, независимо дали е отпечатан на хартия или представлява цифров файл е неговата достоверност и тя трябва да се запази през целия жизнен цикъл. Достоверността остава ненакърнена и във всички фази от движението на електронния документ, от първата му регистрация и извличане на метаданни в електронния архив до реализирането на стратегии за дългосрочно съхранение чрез миграция или емулация. Като основен компонент на достоверността надеждността отразява свойство, което касае фактологическата същност,

докато автентичността се отнася до предмета на неговите претенции. Автентичността на дигитални обекти понякога подлежи на субективна дефиниция според конкретния случай. Субективността на определението е особено важна, когато става дума за документалното историческо наследство, създадено от отделни личности, а не от служителите в действащи институции. Защото там документите преминават през много ръце и се проверяват периодично, преди да постъпят в архивохранилището.

Когато говорим за електронно създадени документи, има много повече възможности за нарушаване на тяхната достоверност, отколкото при традиционните документи, напр. чрез заличаване на метаданните от дигиталния обект, които са важен елемент на достоверността. Инструментите на цифровата форензика позволяват реконструкция на извършените промени. Така би могъл да се възстанови произходът на постъпилия в архива електронен документ. От друга страна електронни документи губят своята достоверност много лесно – не само физически, но и на логическо ниво (промяна на файловете и техните метаданни – датата на последната промяна или достъп до файла, напр.). Те са уязвими от момента на напускане на притежателя или автора до постъпването им в електронния архив. Налага се използването на неинвазивни методи за достъп до съдържанието на електронните документи. В момента на постъпване на архивохранилището задължително се регистрират всички сведения около достоверността, които да удостоверяват надеждност и автентичност. Методите на форензиката позволяват да се определи автентичността на файловете, да се разкрият техните логически и физически характеристики, без да се променя оригинала в цифровия запис. Наред с това цифровата форензика позволява възстановяването на изтрити и скрити файлове както и да се прави търсене на текстове и изображения с цел разкриване на конкретно съдържание. Аналитичните възможности на цифровата форензика разкриват критични данни, като например временни файлове или тези от историята на посещенията на браузъри, които авторът не е имал намерение да предава в архив. Използваните методи от една страна служат за определяне на автентичност, но от друга страна позволяват разкриването на неявни връзки в дейността на фондообразувателите.

В ерата на цифровата информация професията на архивиста е поставена пред нови изисквания относно професионалната етика. Той получава уникален достъп до електронно създадени документи на органи-

зации и отделни лица. Нов смисъл придобива и понятието реставрация, тъй като архивистът има задачата да управлява цифровите обекти през пълния им жизнен цикъл. Цифровият реставратор има пълната власт над цифровите обекти от постъпването им в архива до дългосрочното им съхранение в електронните репозиториуми. Достъпни стават личните цифрови истории на хората, които са предоставили своите фондове в архив. С цифровата форензика се възстановява съществена част от уеб историята на фондообразувателите, разкривайки поведение, което те не са имали намерението да предават за съхранение (напр. конфиденциална информация от финансово или медицинско естество).

Особено значение придобиват проблемите, които са свързани със сигурността в електронните архиви. Налага се разделянето на материали за свободно ползване от тези с чувствително съдържание, дори изключването им от справочните форми за публикуване. Ясно е, че съвременната култура на социално общуване налага едно по-свободно отношение към предлагането на информация. Понякога намерението за гарантиране на сигурност принуждава архивите да унищожават информация за да избегнат опасността от попадането и в недобросъвестни среди. Или да създават т. нар. „затворени“ архиви с физическа защита, която да гарантира необходимото ниво на сигурност.

Друг съществен въпрос около електронно създадените документи е тяхната ценност. Към настоящия момент цифровата информация не притежава някаква комерсиална стойност в смисъла на тази, която има една оригинална грамота или ръкопис. Експертизата на ценността на електронния документ трябва да става от гледна точка на неговата информационна стойност, а не според начина на формалното му представяне и използваните носители. Независимо от възможността за многократно копиране на файлове, съществуват обаче методи за разкриване на уникални файлове, особено със средствата на цифровата форензика, те са полезен инструмент при изготвяне на подобни експертизи.

Бележки

[1] Вж. напр. *Barata, K.* Archives in the Digital Age, Journal of the Society of Archivists 25, no. 1, 2004.

[2] *Mckenzie, D. F.* Bibliography and the Sociology of Texts: The Panizzi Lectures 1985. London, The British Library, 1986.

[3] *Ross, S. A. Gow.* Digital Archaeology: Rescuing Neglected and Damaged Data Resources. South Bank Univ. London, 1999.

[4] *Farmer, D. W. Venema.* Forensic Discovery. Addison Wesley, 2004.

[5] Тестът представлява своеобразно интервю между човек и машина, при което интервюиращият задава въпроси, без да знае, дали отговорите се получават от машина или от човек. Ако отговорите на машината не са различими от отговорите на човек, тогава същата притежава интелект. *Turing, A. M.* Computing Machinery and Intelligence. Mind, Vol. 59, No. 236, 1950, pp.433–460.