The Increasing Challenge of DIGITAL FORENSICS

By Prof. William J. Buchanan
Edinburgh Napier University, UK

Over the past decade or so, we have seen a massive increase in the usage of digital devices for evidence gathering in law enforcement, and the methods that we have used are typically based on "dead" forensics, where investigators examined a powered-off computer which had an image taken of it. In itself the focus has always been to preserve data and comply with the ACPO 3 regulations for the handling of the seizure of data. Unfortunately digital forensics is a discipline that has many increasing challenges, including:

- **Range and number of devices.** Investigators are becoming overwhelmed with the range and number of the devices that they have to investigate. A seizer now might involve a number of laptops, USB sticks, game units, cameras, mobile phones, and so on.
- **Increasing move toward live and cloud forensics.** There are increasing numbers of investigations that involve the signs of malware on system, and the investigation of these things often require that the systems are running and for the investigator to observe the dynamic nature of the system, rather than based on a static image of the system at one point in time. With malware it is often difficult to examine the propagation and operation of the malware, without actually running it on a live system.
- **Encryption-by-default.** There is now a move from the major IT companies to move toward both storing data with an encryption and creating network connections that use secure tunnels. Again forensics tools and networking monitoring/logging often struggle to pick any signs of evidence. Networks such as Tor will completely anonymize any sort of trace either on the host or through network logs (such as from the ISP).
- **Operating in the Cloud.** We are increasingly running programs and storing data within the Cloud, and there is an ever-decreasing trace of evidence on computer disk system. Many of the traditional digital forensics was built upon the investigation of dead forensics.
- **Time to archive.** As disk drives become larger, often over 1TB, it takes longer to archive them for scientific analysis. A 1TB Hard Disk (HDD) takes around 20 hours to take a snapshot (while it is around 20 minutes for a Solid-State Disk Drive (SSD)).
- **Crackers are moving faster than the tools and the skills base.** Increasingly malware and cracking methods are moving so fast that it is difficult for investigators to keep up-to-date on these new methods, which leaves a massive skills gap.
- **Move from Windows...to Mac OS, Android and iOS.** Many of the tools have been built around Microsoft Windows, but more home computers and laptops are moving toward Apple systems, and to Android and Apple iOS operating systems for mobile devices. With many of the laws related to mobile phone forensics, there are many unknowns related to investigations on them.
- **From hardware to software.** Computer systems are becoming visualized rather than running a single operating system on a single machine. This includes creating whole networks with interconnected systems within a software-defined infrastructure.

THE SKILLS GAP AND DIGITAL SHADOWS

There is a massive skills gap evolving where, at one time, an investigator needed to understand the NTFS/FAT file system, and associated file format, and that would help them to find contraband content. With modern systems, investigators now need to understand network protocols; different operating systems (such as iOS, Android, Mac, and so on); memory analysis (live forensics); cyber-attacks (e.g. DDoS, SQL Injection, and so on); penetration testing; malware operation; traces of digital artifacts on the Internet; and so on. It is thus difficult...
for investigators to keep up-to-date with the complete picture of the increasingly complex nature of our digital footprint. A key challenge of this is understanding the creditability of information and the ethics of actually gathering it.

**IT’S ALL GOING ELECTRONIC…**

In the UK, the standard procedure used in digital forensics is defined by ACPO’s Good Practice Guide for Computer-based Evidence (Figure 1)[3], and defines the key stages. The guide, though, is really more focused on computer systems rather than investigating modern malware, cloud-based systems and networked devices. See chart top right.

**TUNNELS AND TOR**

One method that has been used in investigations has been to examine network logs. With the increasing usage of tunnels, such as using SSL/TLS, the IP address of the remote site can be logged, but there will be no logs of the actual pages that were visited (as this is protected in through the SSL layer). With the Tor network, the complete trace of the accesses cannot be picked up, as the network packets are routed through peer devices rather than taking the normal routes through the network. Again, a secure tunnel is used to route the data, so that the contents of the packets cannot be viewed. For investigators, especially for live investigations, the “wire taps” thus need to happen on the host, and it is almost impossible to crack the encryption keys used within the tunnels. With current cryptography, the current state-of-art cracking systems can break 72-bit encryption, which is still well short of the 128-bit or 256-bit keys which are used within most tunnels.

**PUBLIC, PRIVATE AND HYBRID CLOUDS**

We are increasingly virtualizing systems and building up clustered systems, which pool together servers, memory and data storage. It is often extremely difficult to investigate these types of systems using traditional digital forensics methods, as data sizes might be measured in many hundreds or thousands of TBs. It is also difficult to shut these systems down and not affect other systems, thus the traditional guides around seizure are often not valid. So law enforcement has a major problem in investigative instances with cloud-based systems, especially where the cloud spans from a private to a public cloud, and also where instances can be deleted with a press of a button.

**THE INCREASING TIMES TO ARCHIVE**

With ever-increasing disk sizes, it has become difficult to archive the static snapshot of a disk system. Figure 2 outlines some recent tests on the time it takes to achieve a range of disk systems. A 3TB SATA HDD, for example, takes over 6 hours to archive, while and 512GB SSD takes around 17 minutes. It is becoming

---

**As we are increasingly more mobile**, we are now carrying around sensitive data that at one time was protected behind physical firewalls, and the risks to our data increases by the day.
difficult for investigators to archive the wide range of devices that are seized from a crime scene. See chart above.

**ENCRYPTION-BY-DEFAULT**

Google’s Lollipop has just been released, and security will be at the core of its changes. An important element of this is in encryption-by-default, where users will have to opt-out of encryption of their files. Apple, too, has taken the same route with iOS 8.

Our file attributes and content types have developed with little thought on keeping things truly private, and where systems are often still viewed as standalone machines. We also created an Internet that is full of the same protocols we used in the days of text terminals and mainframe computers, where users typed in commands to access data, and where there was little thought about protecting the data as it is stored, analyzed and transmitted. As we are increasingly more mobile, we are now carrying around sensitive data that at one time was protected behind physical firewalls, and the risks to our data increases by the day.

The major tension, though, is between law enforcement and the right to privacy. The FBI currently see the status quo as a way of investigating criminals and terrorists but can see this opportunity reducing with encryption-by-default, such as with the file encryption system used in Apple’s iOS 8. With iOS 8 and Google Lollipop, there will be no electronic methods to access encryption keys from existing digital forensics toolkits, and thus the encryption method breaches current laws, which force users to reveal their encryption keys when requested by law enforcement investigators. This would mean that users may be breaching current laws in both the U.S. and the U.K. The same battle too exists with Tor, where law enforcement members are scared that crime can go unnoticed, whereas privacy advocates promote the rights of privacy of using Tor.

**THE FOUNDATIONS ARE CRUMBLING**

Imagine if you were an electrical engineer, and you woke up one day, and they told you that ohms law was no longer relevant. Well this is happening with many of the protocols used on the Internet, especially the ones created over three decades ago.

Some of the original protocols used on the Internet, including HTTP, FTP and Telnet, are all fading fast for their credibility to use in an investigation. Many of the core security methods are also falling fast, with MD5, a standard method for creating a digital fingerprint of data, has been shown to create two images with the same hash signature. This week, though, Mat McHugh [2] showed that he could produce the same hash signature for different images, using HashClash, and for just 65 cents on the Amazon GPU Cloud, and took just 10 hours to process.

For 10 hours of computing on the Amazon GPU Cloud, Mat created these two images which generate the same hash signature (Figure 3). If we check the hash signatures we get:

C:\openssl> openssl md5 hash01.jpg
MD5(hash01.jpg)= e06723d4961a0a3f950e7786f3766338

C:\openssl> openssl md5 hash02.jpg
MD5(hash02.jpg)= e06723d4961a0a3f950e7786f3766338

See Figure 3 below

Continued on page 14