The Economics of Digital Forensics

Tyler Moore

Computer Laboratory, University of Cambridge
15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom
Tyler.Moore@cl.cam.ac.uk

Abstract

The collection of electronic data as evidence of crime is an important responsibility given to law enforcement. The technical constraints of this task are arguably far less significant than usability and economic ones, since police officers are non-specialists and police departments face significant budgetary limitations. In this position paper, we consider the economics of digital evidence recovery. We argue that the incentives of technology companies, law enforcement agencies and society do not always align, and furthermore that by studying these incentives in different applications we can better understand the efficiency and extent to which digital evidence is gathered.

1 Introduction

We consider the problem of analyzing and processing digital evidence for investigating physical crimes. Digital devices keep precise records of incriminating activity, even more than is typically realized. Perhaps the most compelling reason driving the growing use of digital forensic techniques is a side-effect of how digital information is stored: remnants of deleted data often remain on devices, unbeknownst to users [4]. Many police officers have received special training to use standard tools which examine a suspect's hard drive for incriminating evidence, from 'deleted' correspondence about an insider stock trade to illicit images of child pornography. But analysis is not limited to PCs: thumb drives, PDAs and mobile phones are now routinely examined. Law enforcement agencies have long operated under limited budgets and finite resources. Thus far, their capabilities have been put under significant strain, though they have managed to largely keep pace. However, technological changes threaten to undermine their capacity for complete data analysis. A thorough analysis of the challenges to the efficiency and viability of existing techniques is warranted.

It turns out that many of the important constraints on digital forensic practices are not technical, but economic. The incentives of technology companies do not always align with law enforcement, and both often conflict with those of

consumers. While it may be the case that government agencies at the highest level can influence industry behavior to a certain extent, most law enforcement agencies are left to react to the challenges posed by technological advances. PC recovery and analysis is a good example of law enforcement responding to a change in behavior: people began to store records of potentially incriminating records on computers without properly deleting them, spawning the discipline of computer forensics.

Our contributions in this position paper follow. We cast the problem of recovering digital evidence in economic terms. We are unaware of any prior academic attempts to do so. In Section 2, we study how technology choices can impact the costs imposed on law enforcement. In particular, we compare the costs of examination for devices adhering to open, standard formats to those using closed, proprietary ones, using recovery from PCs and mobile phones as an example. We present a simple model and find that the social cost is greater in the latter case, both in terms of development costs and the reduction in the number of devices available for analysis. In Section 3, we characterize network communication as generating an externality of data, whose use carries differing values for users and service providers. We argue that the business case to store increasing amounts of user data may impose a significant burden on law enforcement in the near future.

2 Open versus Closed Systems

The debate over the merits of open and closed systems has centered on questions of reliability and security [6], yet the choice also greatly impacts the cost of extracting digital information. Forensic examination techniques have high fixed costs for each particular technology in use, requiring tailored software to extract data according to a specification. When many applications adhere to common formats, high fixed costs are justified since the extraction software can be reused many times. If instead each device manufacturer relies on its own proprietary formats, the overall cost of extraction rises significantly. Furthermore, many devices storing data less likely to be relevant an investigation may not justify the high fixed costs of software development.

2.1 PCs and mobile phones

The differing approaches to extracting information from PCs and mobile phones provides a telling case study of how a technological decision to use proprietary formats can hinder law enforcement's forensic capacity by raising the cost of recovery. For PCs, the standard procedure is to make a bitwise copy of a seized hard drive, examining it directly using one of the many available tools which bypass the operating system altogether [9, 5]. This works because the file structure for hard drives is standardized to a only a few types (e.g., FAT, NTFS, ext). Open file system formats encourages wider adoption, and therefore, fewer storage types emerge. As a result, law enforcement can readily access deleted

files in standard formats (e.g., ASCII, MS Word) hidden to the operating system.

In contrast, for mobile phones, information is stored in the phone's internal memory not according to any particular standard. Relevant data like call histories are stored in proprietary formats in locations that change with the phone model. Even the cable used to access the handset's memory varies by model. So extracting data directly from the phone's memory is much costlier for mobile phones than PCs since there are no standard storage or document formats as for PCs. Imagine if Dell, Gateway and IBM, along with every other PC maker, used their own proprietary drive storage and document formats.

As a result, the standard procedure for extracting data from phones is quite different to that of PCs. Many examiners first look to SIM cards for information, because it is stored in a standard format. However, a SIM's limited storage capacity forces most of the relevant data to be stored on the phone itself. Since reading the phone's memory directly is so expensive, most investigators query the phone using interface commands [11]. But this approach creates several significant problems. First, it requires dependence on phone manufacturers to provide software for retrieving data. By serving as intermediary, handset manufacturers can withhold access to relevant information; law enforcement often misses all deleted data (e.g., call lists, contacts and SIM usage histories) that the phone cannot obtain via its interface software. Such an approach is also susceptible to compromised phone software, since the phone's operating system is not bypassed. So simple deletion, easily detectable on PCs, is still a reasonable evasion strategy for mobile phones.

What is more, existing forensic software is only designed for the most common system types. Paraben, a leading provider of mobile phone forensic software, sells packages for examining the top six US handset manufacturer's phones along with SIM analysis software to partially cover the remaining GSM phones [13]. This software accesses the phone memory via interface commands, so it is susceptible to the limitations mentioned above. They do offer a more sophisticated application to analyze the phone memory directly for two manufacturers. The top six handset makers account for 84% of the phones sold in Q4 2005 globally [10]; however, the remaining handsets taken together represent a significant fraction of phones that cannot be analyzed forensically by this popular software. This is a natural consequence of the closed system model on law enforcement. At present, the strategy adopted by most governments is to develop a close relationship with each of the leading manufacturers to ensure access to the largest number of devices [8]. Clearly, such an outcome is suboptimal: a criminal can greatly enhance her stealth by using an obscure phone, for instance by importing a European tri-band GSM phone to the United States.

2.2 Cost model

We now present a simple cost model to more precisely weigh these implications.

Suppose there is a set of M incompatible platforms and each of the i technologies has an expected number of m_i devices to be examined. Developing a method for extracting data from a particular technology bears a high fixed cost

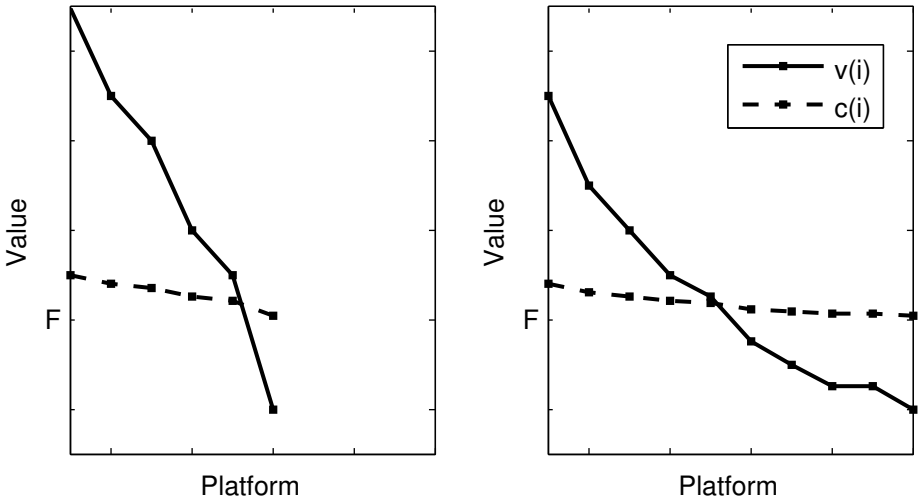

Figure 1: Hypothetical value and cost distributions for open, standardised (left) and closed, proprietary (right) platforms.

F. There may also be a marginal cost MC for performing the examination (e.g., officer time for acquisition and analysis). For simplicity we assume these costs are the same across technologies. (In practice, the costs of reverse-engineering a proprietary format are certainly higher than the developing software to process an open standard. However, here we only weigh the costs brought about by the rise in incompatible formats.) We also anticipate $F >> MC$. Thus the overall cost for analysis all devices is:

$$c(M) = F * |M| + MC * \sum_{i=1}^{|M|} m_i$$

In a largely closed environment with many incompatible formats, $|M|$ is large and so is the cost incurred. Whereas, if systems are based on only a few open standards, then the cost is much lower.

Not all technologies are equally popular; in fact, we anticipate the distribution of technologies to be skewed in favor of a few technologies, as is the case for mobile phones. We expect to see different distributions according to whether open standards or proprietary formats prevail. These distributions may greatly impact the extent to which viable extraction techniques will be developed.

Suppose each of the i platforms are ordered by their popularity, so that $m_1 > m_2 > \ldots m_{|M|}$. Now define a value function $v(i) = m_i * u$, where u is the expected utility of extracting data.

Figure 1 plots $v(i)$ against the cost $c(i) = F + MC * m_i$ for two different value distributions. The plot on the left side corresponds to a concentrated distribution into a few standardized platforms, whereas the plot on the right

matches the case where proprietary formats dominate. Note that the point at which costs exceed benefits is met much sooner when proprietary systems are used. So not only are the overall costs of extraction higher in the proprietary case, but the social cost is magnified further by the large portion of platforms that do not warrant developing tailored recovery software.

2.3 Discussion

So what are the implications for computer systems as the number of intelligent devices and networks expand? What happens if the pervasive computing vision of smart air conditioners and refrigerators comes true? Certainly the prospects for inexpensive data extraction rest on whether air conditioners settle on a common operating platform. If so, then the cost of forensic examination falls accordingly since standardized tools can be reused on many devices. As more devices become networked, the tendency towards proprietary storage formats could greatly increase the number $|M|$ of incompatible platforms, many of which will not be popular enough (in terms of criminal investigations) to justify the fixed costs of development. Devices explicitly designed for communication–like mobile phones–will consistently be of higher value to investigators than other devices. In contrast, few forensic examiners can justify the time and resources to extracting the memory of a single washing machine, even when it offers crucial evidence (like providing an alibi).

Note that network externalities only force standardization on communications. So while every phone adheres to a standard (GSM, CDMA, TDMA) for communication, individual phone architectures remain proprietary. The combination of high fixed and low marginal costs with positive network externalities typically force de facto standardization in software industries [14], yet here the benefits of file system standardization only help law enforcement, not the handset manufacturers who determine formats. Since much of the valuable forensic information is particular only to the phone and not communications between devices, forensic examiners are left to analyze proprietary handsets. The same may very well happen for ubiquitous computing: devices may be identified with IPv6 addresses using RFID tags and communicate via standard network protocols, while their internal storage could yet remain proprietary. The resulting cost to law enforcement would be significant.

The policy implications are surprising. It is clear that governments should push harder for open standards to lower the cost for law enforcement to recover data. But what about privacy advocates? Often supporters of open formats, perhaps they should instead advocate closed systems to make it more costly for governments to snoop.

3 Externalities of Networked Communication

When two entities communicate over a network, be it via email, telephone or banking systems, a trail of evidence is generated as a byproduct. Yet users

cannot control the type or extent of the data that is stored. In the phone system, for instance, EU laws have until recently required the storage of call-related data only as long as a business case can justify keeping it; for example, many phone companies use calling records as input to fraud-detection systems. Once a business case has been established, however, these records may be turned over to law enforcement when potentially relevant to particular investigations. Recent legislation is extending data retention limits to as much as two years in the name of terrorism prevention [3]. But it is not only governments that are interested in retaining information: Google has shown there is great monetary value in tying user behavior to targeted advertising. It is no surprise then that the question of how such user-generated data is exploited is an economic consideration; viewing network data retention and examination in terms of consumption externalities can help to explain the existence of different equilibrium outcomes and guide policy.

An externality of data is generated when a user communicates over a network. This externality can take the form of metadata (attributes regarding correspondence) or the communication itself. Unlike PCs, users do not own and manage networks, so they have little choice but to allow the service provider to keep a record of activity. Critically, only service providers can control how and whether data is stored; users may be assigned property rights but remain at the liberty of service providers to respect them. Regulations have been put in place to ensure providers do not abuse this position of authority. In this section, we study the incentives facing different types of network providers for retaining and exploiting user-generated data, the resulting effect on forensic examination, and several options for addressing the externality's impact.

3.1 Differing values for data externalities

The user and network provider have different preferences to the extent user-generated data should be leveraged by the network operator. Just exactly how could an operator exploit communication data? He can choose to retain data, analyze it or even share it with other third parties (like companies or governments). We focus on the decision to retain data since this fundamentally impacts law enforcement's subsequent capacity to analyze user behavior.

Most *consumer* service providers (e.g., Google) would much prefer unfettered access to communication records whenever it benefits them. This is reflected by a trend for network providers to attempt to store records of all communications. Several factors are driving service providers to retain more information from network interactions. First is the falling cost of storage; this lowers the threshold for making customer information profitable. Second is the recognition that user participation is a valuable asset that can be leveraged: email correspondence reveals interests for targeted advertising; search histories can be used to improve the relevance of results. Odlyzko has argued that most value from a communications network stems from its capacity to connect people [12]; that companies see enough value to keep records of communications reinforces his claim. So while privacy laws limit disclosing information to third parties, ser-

vice providers can create value in other ways. Finally, end users find value in storing information on networks they do not own: it provides a convenient way to access files on different devices and is an inexpensive way to keep a back-up copy of communications.

There is an exception to this trend, however: employers operating corporate networks. In fact, many companies routinely delete all correspondence after a time as a matter of policy. This is because old records of correspondence may one day prove incriminating. One key difference here is that the employer, in effect, owns both the network and the employee's official correspondence. So the externality is internalized. Since the employer has an interest in protecting its employees' actions from outside scrutiny, it typically retains as little data as legally possible.

3.2 New opportunities create new responsibilities for law enforcement

Forensic examiners are acutely interested in collecting these data side-effects; yet they have a limited mandate to compel data retention. In fact, they are largely dependent upon the operating strategies and preferences of service providers who choose to retain user information and users choosing to store data on a network. So while law enforcement can expect limited user data from corporate networks, the business models of many consumer-oriented service providers are creating a wealth of information available for analysis. While a privacy law may compel telephone companies to delete call records after a time, nothing can be done to prevent people from choosing to keep a permanent record of every VoIP conversation made using Google Talk. In fact, the sheer amount of data made available could overwhelm law enforcement.

New technology has repeatedly created new obligations for law enforcement. First PCs and now mobile phones have been targeted for examination because they often store information pertinent to investigations. The costs of officer training, investigation and analysis has been quite substantial, but so have been the benefits. However, analyzing network data could prove much costlier than PC and even mobile phone recovery for two reasons. Unlike PCs and mobile phones, people do not own the devices storing their information on networks. So any examination must be coordinated with the network operators themselves. Second, the amount of data relevant to a suspect has the potential to grow very large. Many different networks may collect information, while the amount of information that may be relevant to an investigation is a decreasing fraction of the overall data available. Having perfect digital records of all transactions throughout a person's life is at best onerous [15], while at worst the noise of extraneous data can provide cover for incriminating activity.

Already, the levels of digital evidence is overwhelming officers; as users begin to store important data on multiple networks, the burden will multiply. Acquisition and much of the analysis are labor-intensive; the only viable long-term strategies are to increase the number of trained officers and improve software automation and cross-correlation techniques [7].

3.3 Options for addressing data externalities

The explicit assignment of property rights is key to reaching efficient outcomes. At one extreme, all rights may be transferred to the service provider. In this case, users accept that any communications may be examined by outside parties. Many users may choose to limit online communications or adopt end-to-end encryption to limit outside exposure. At the other extreme, users could be granted complete control over communication records. Here, any data externality can be efficiently resolved, viewing the network as an extension of personal storage.

As is often the case with externalities, ambiguities over property rights dominate in practice, creating the potential for inefficient outcomes. Most users expect their communications to be kept private, yet current law and most privacy policies reserve the right to significant data collection and sharing. This well-documented gap between user expectations and actual provider behavior [1, 16] serves the interests of service providers and law enforcement, for it enables providers to gather more data on personal correspondence than might be possible if users were aware of these practices. Unfortunately, even if outstanding ambiguities are clarified, the damage may already be done: once users have adapted their behavior by storing correspondence via third parties, the high switching costs characteristic of the software industry may preclude an efficient equilibrium from ever being reached.

Another potential resolution is for users to internalize the data externality. Here, that would require users to own the network they communicate over. While this is feasible for home networks, the nature of communication essentially requires dependence on a third-party network for transmission.

4 Conclusions

The collection and analysis of digital evidence is an important task given to law enforcement with serious implications for society. We have argued that the incentives of companies, which develop technology, and law enforcement, which gathers data as evidence, are often in conflict. The proliferation of proprietary storage formats in mobile phones offers a compelling instance of this phenomenon: the rising cost of data extraction has forced governments to invest in more expensive acquisition techniques, even as less relevant information is ultimately collected.

Changes in business strategies may also challenge the viability of forensic examination. Many companies have recognized the value in collecting records of user behavior as the cost of storage has fallen, yet the subsequent cost of forensic analysis remains high. We have characterized the interaction between user and network as generating a data externality. The assignment of property rights remains uncertain, which could yield socially inefficient outcomes not easily reversed due to high switching costs.

It has been established that information security is usefully viewed in terms of economic constraints [2]. We feel that the same is naturally true for digital

forensics, arising from the budgetary limitations facing law enforcement. In this position paper, we have studied several compelling instances, demonstrating the promise of analyzing digital forensic practices economically. Much work remains to be done, however, perhaps beginning with a more thorough analysis of the costs and benefits of network topologies and technologies on gathering digital evidence.

Acknowledgments The author is supported by a scholarship from the Marshall Aid Commemoration Commission. The author thanks the anonymous reviewers for their comments.

References

[1] A. Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the ACM Electronic Commerce Conference (EC 04)*, p. 21–29. ACM Press, 2004.

[2] R. Anderson. Why information security is hard—an economic perspective. In *Proceedings of the 17th Annual Computer Security Applications Conference*, p. 358. IEEE Computer Society, 2001.

[3] EU approves data retention rules. *BBC News*, December 14, 2005. `http://news.bbc.co.uk/1/hi/world/europe/4527840.stm`

[4] M. Caloyannides. *Privacy Protection and Computer Forensics*. Artech House Publishers, 2004.

[5] B. Carrier. The Sleuth Kit and Autopsy: forensics tools for Linux and other Unixes, 2005. `http://www.sleuthkit.org`

[6] J. Feller, B. Fitzgerald, S. Hissam and K. Lakhani, editors. *Perspectives on Free and Open Source software*. MIT Press, 2005.

[7] S. Garfinkel. Cross-drive analysis and forensics. Submitted to the *15th USENIX Security Symposium*, 2006.

[8] V. Gratzner, D. Naccache and D. Znaty. Law enforcement, forensics and mobile communications. In *Proceedings of the Third IEEE International Workshop on Pervasive Computing and Communication Security*, p. 256–260. IEEE Press, 2006.

[9] R. Keightley. Encase version 3.0 manual revision 3.18, 2003. `http://www.guidancesoftware.com`

[10] K. Mackenzie. Nokia increases handset market share. *Financial Times*, February 28, 2006.

[11] P. McCarthy and J. Slay. Mobile phones: admissibility of current forensic procedures for acquiring data. In *Proceedings of the Second IFIP WG 11.9 International Conference on Digital Forensics*, 2006.

[12] A. Odlyzko. Content is not king. *First Monday*, vol. 6 num. 2, 2001.

[13] Paraben Cell Seizure, `http://www.paraben.com/`.

[14] C. Shapiro and H. Varian. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press, 1998.

[15] F. Stajano. Will your digital butlers betray you? In *Proceedings of ACM Workshop on Privacy in the Electronic Society (WPES)*, 2004.

[16] J. Turow. Americans and Online Privacy: The System is Broken. *Annenberg Public Policy Center Report*, 2003.