

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/259164846>

# Triage template pipelines in digital forensic investigations

Article in *Digital Investigation* · September 2013

DOI: 10.1016/j.diin.2013.03.001

---

CITATIONS

4

---

READS

98

3 authors, including:



[Richard Overill](#)

King's College London

108 PUBLICATIONS 1,222 CITATIONS

[SEE PROFILE](#)



[Keith A. Roscoe](#)

Metropolitan Police Service, London, UK

1 PUBLICATION 4 CITATIONS

[SEE PROFILE](#)

All content following this page was uploaded by [Richard Overill](#) on 25 July 2014.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.



Contents lists available at SciVerse ScienceDirect

## Digital Investigation

journal homepage: [www.elsevier.com/locate/diin](http://www.elsevier.com/locate/diin)

## Triage template pipelines in digital forensic investigations

Richard E. Overill<sup>a,\*</sup>, Jantje A.M. Silomon<sup>a</sup>, Keith A. Roscoe<sup>b</sup><sup>a</sup> Department of Informatics, King's College London, Strand, London WC2R 2LS, UK<sup>b</sup> Metropolitan Police Service, Digital Electronics & Forensics Service, New Scotland Yard, 8–10 Broadway, London SW1H 0BG, UK

## ARTICLE INFO

## Article history:

Received 6 November 2012

Received in revised form 28 February 2013

Accepted 6 March 2013

## Keywords:

Digital forensics

Triage template pipelines

Cost-effective prioritisation

Investigative process

Digital crimes

## ABSTRACT

This paper addresses the increasing resources overload being experienced by law enforcement digital forensics units with the proposal to introduce triage template pipelines into the investigative process, enabling devices and the data they contain to be examined according to a number of prioritised criteria.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

As society becomes increasingly digitalised the proportion of criminal investigations which involve the forensic examination of digital devices also increases. These devices may be peripheral to the main investigation, e.g. establishing the geotemporal location of individuals of interest to the investigation from sat-navs or mobile phones, subject to attribution of the devices in question. Alternatively, they may be linked to the main investigation in a generic way, e.g. browsing the internet for particular information or using social media. Finally, they may be central to the investigation if the particular use of the device itself constitutes a criminal act (e.g. downloading and viewing child pornography, or committing online fraud).

The very reliance on digital devices for the conduct of most people's daily professional and personal lives has led to an overload on digital forensic examination resources. As an example, the London Metropolitan Police Service Digital Electronics and Forensics Service (MPS DEFS) currently

receives over 38,000 digital devices *per annum* to be examined by a team of approximately 80. Not untypically these devices will have a storage capacity of gigabytes or terabytes. This all adds up to a demand – supply equation that is severely skewed towards the demand side, and with public sector budget cuts necessitated by the global recession there is no realistic possibility of increasing the level of resourcing to match the ever-increasing demand.

Under these circumstances, it has become necessary to attempt to optimise the forensic examination process. The notion of *triage* (literally, sifting or screening) is borrowed from historic medical practice on battlefields and more recently at the Accident & Emergency Departments of busy hospitals. Thus hopeless cases can be swiftly dismissed while urgent or straightforward examinations can be given priority. Again, in analogy with battlefield medical evacuation strategy where a serious casualty can be stabilised in the field before being 'medevaced' to a field hospital, the performance of simple digital forensic procedures *in situ* at the suspected crime scene by first responders from law enforcement is currently being trialled by MPS DEFS, before the devices are disconnected, bagged, tagged and removed to the digital forensic laboratory (DFL). Except in the most straightforward case of seizing already powered-off devices, decisions regarding the potential loss of evidence

\* Corresponding author. Tel.: +44 20 7848 2833.

E-mail addresses: [richard.overill@kcl.ac.uk](mailto:richard.overill@kcl.ac.uk) (R.E. Overill), [jantje.a.silomon@kcl.ac.uk](mailto:jantje.a.silomon@kcl.ac.uk) (J.A.M. Silomon), [keith.roscoe@met.police.uk](mailto:keith.roscoe@met.police.uk) (K.A. Roscoe).

that may be overwritten by shutting down or powering-down digital devices must be taken. In cases where it is believed that the suspect may be technically savvy, an attended or booby-trapped device may be at threat of hot-key erasure of potentially valuable evidence, and a more sophisticated approach to seizure, such as using social engineering based distraction techniques, may be required. Fortunately the majority of seizure scenarios will not involve such extreme anti-forensics. However, it is not the aim of this paper to enter the somewhat contentious debate about *in situ* forensics (see Rogers et al., 2006, for a detailed account of field triage), and we therefore assume here that all examinations are performed in a DFL.

## 2. Modelling digital forensic triage

The underlying initial hypothesis as to what type of crime has been committed can be used to provide an initial direction for the search, seizure and screening of digital devices in preparation for subsequent forensic examinations of any relevant evidential traces they may contain. Previous hypothesis based approaches to digital forensic investigation include the use Bayesian networks as pioneered in Kwan et al. (2008).

Following Casey et al. (2009) we base our discussion upon a three-level investigative hierarchy comprising triage/survey forensic inspection, preliminary forensic examination, and in-depth forensic examination. Within this context, we take the survey/triage level to encompass the following sequence of activities: (i) pre-seizure – generating a list of anticipated digital devices of potential relevance to the investigation based on the initial hypothesis and itemised on the search warrant; (ii) search for and seizure of such relevant digital devices and logs as can be located; (iii) post-seizure – screening of the seized devices for the likely existence of relevant evidence, in a prioritised manner. It should be mentioned that (ii) has become less straightforward now that USB keys are frequently disguised

as everyday non-digital items such as sunglasses or safety pins and thereby ‘hidden in plain sight’. We note the ongoing debate as to whether (iii) constitutes survey/triage or preliminary forensic examination, but we opt to follow Casey et al. here. Taken together, these three activities permit the construction of triage template pipelines for specific classes of crime involving a digital element. Examples are shown in Figs. 2 and 3, and will be discussed in more detail below.

The preliminary forensic examination level, in addition to interpreting relevant evidence and making it available to others involved in the case (Casey et al., 2009), also affords the possibility of creating a feedback loop into the survey/triage level through the discovery of evidence that opens up new avenues of investigation by refocusing subsequent searches for evidence, or potentially necessitates a realignment of the investigation. An example of this is given in the P2P template as described below and illustrated in Fig. 3.

In addition to feedback, the concept of parallel tracking will be described briefly. Since certain tasks possess a higher latency than others, these should be initiated as soon as there is a reasonable prospect that they will yield material evidence of sufficient probative value. For example, requests for information from ISPs or CSPs under the UK Regulation of Investigatory Powers Act (RIPA) normally require several working days to be fulfilled, while overseas requests for Web 2.0 data harvesting under a mutual legal assistance treaty (MLAT) may take several weeks.

The urgency and/or severity of the hypothesised crime may also necessitate the use of parallel tracking in order to expedite the investigation, for example, where digital evidence relating to an ongoing abduction is encountered.

From an organisational perspective the availability of human and technical resources is likely to be a key factor. Parallel tracking becomes relevant if the size of a digital forensic triage team permits it to be subdivided into groups which can focus their attention on different sub-tasks involving evidence from the same case. For example, separate groups could be tasked with examining the evidence

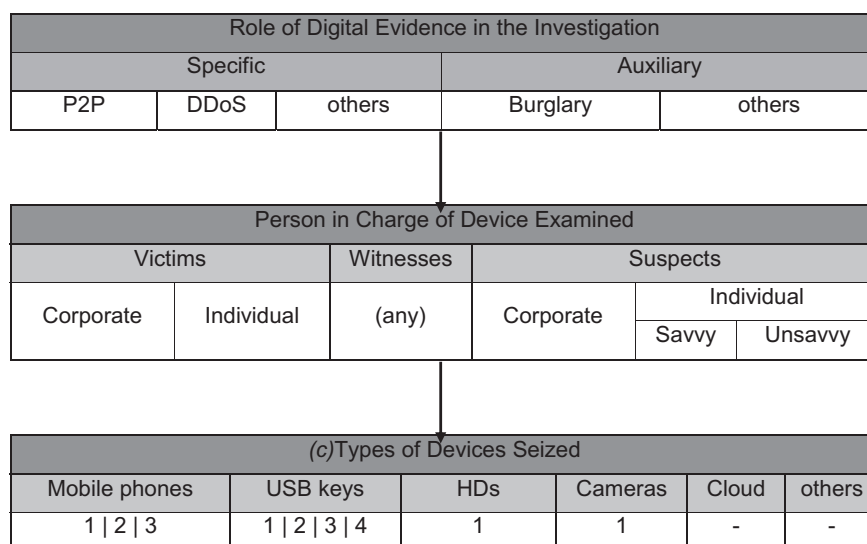


Fig. 1. Triage elements.

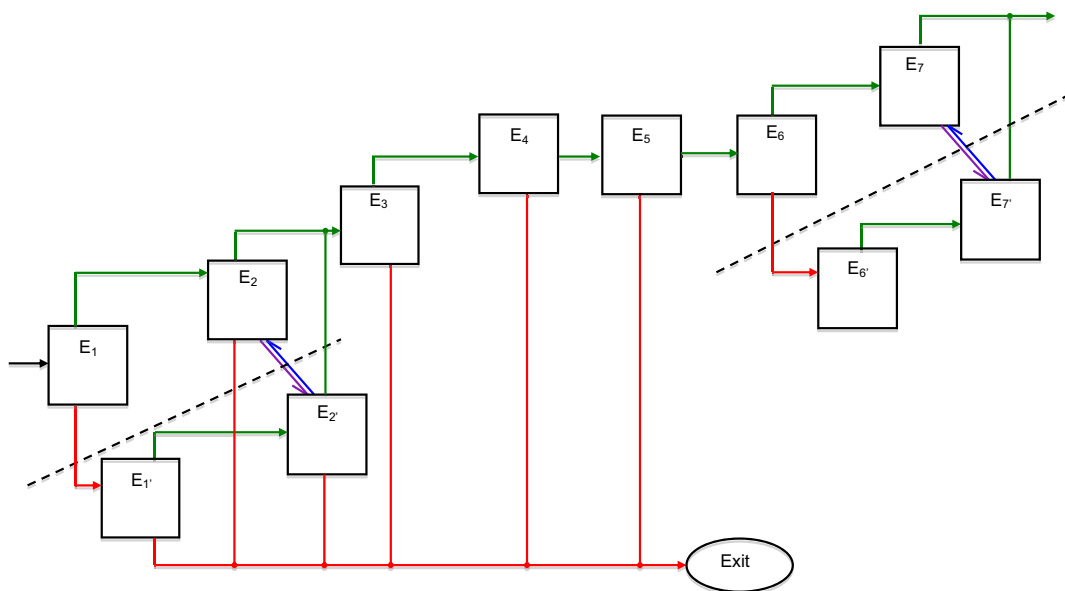


Fig. 2. DDoS template diagram.

from seized HDs, USB keys and mobile phones respectively. This is indicated by means of a diagonal dashed line on the template diagrams for individual digital crime scenarios shown in Figs. 2 and 3, which will be discussed later.

Since parallel tracking treats several aspects of the overall evidence associated with a case, the findings of each group must be made available to all the other groups in order to either corroborate or refute the current hypothesis. This may potentially lead to a refocusing and/or reprioritisation of the investigation strategy, including restarting a previously suspended thread of the investigation. This can be viewed as introducing an element of feedback, as described above, into the investigative process. Of course, even when only a single group is involved,

it is still possible that their interim findings can have the effect of altering the overall direction of the investigation. An example of an investigative strategy, based on a concept suggested by MPS DEFS, makes use of an enhanced hybrid of Gantt charts and Critical Path diagrams, to provide a graphical representation of both parallel tracking and feedback (Fig. 4).

### 3. Triage template pipelines

Pipelines are not new in digital forensics. For example, Garfinkel (2006) has described an eight-stage pipeline process for cross-drive analysis. In addition, a three-stage triage pipeline for performing digital forensics using

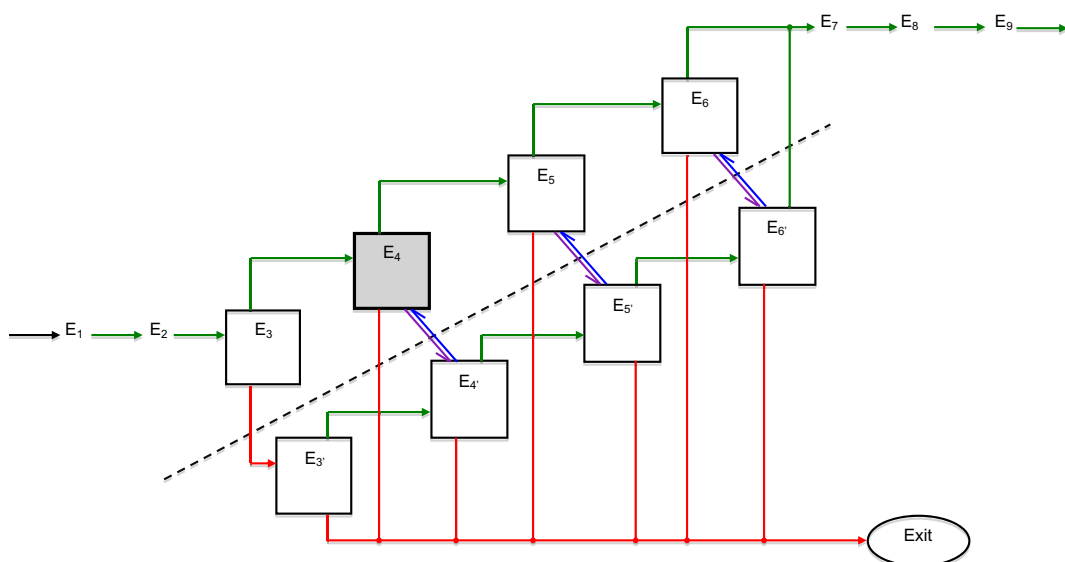


Fig. 3. P2P template diagram.

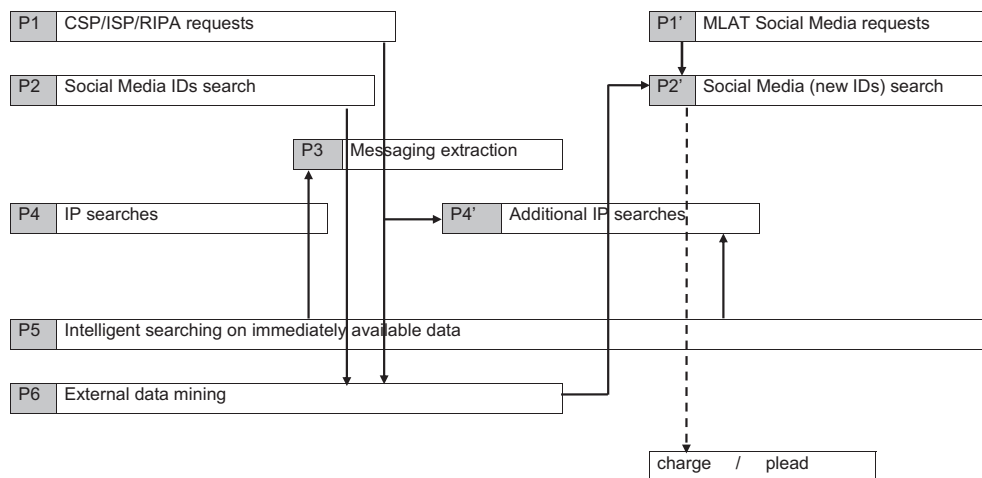


Fig. 4. Example investigative strategy involving parallel tracking and feedback.

Bayesian Belief Networks (BBN) was recently proposed by Overill and Silomon (2011). It is worth emphasising here that the value in developing triage template pipelines relies upon the observation that a form of the Pareto principle or 80–20 rule (Newman, 2005) commonly operates for digital crimes. This rule-of-thumb, first reported by the economist Pareto in 1906, maintains that for a multitude of phenomena around 80% of the effects are generated by just 20% of the causes. For example, in Hong Kong about 80% of all digital crimes are accounted for by just five types of crime (Kwan and Law, 2009–2010). It therefore makes sense to develop templates for the efficient investigation of each of these specific classes of digital crime. Naturally, the set of prevalent digital crime classes will vary in different parts of the world, depending on cultural, economic and political factors, so a different set of templates may be required for different geopolitical regions, e.g., Europe, North America, East Asia, and the Middle East.

The role of the digital evidence in the overall investigation forms a basis for the survey/triage level and can be classified as either *specific* or *auxiliary*. *Specific* digital evidence is considered to be central to the commission of a particular, prevalent digital crime, resulting in the creation of distinctive digital evidential traces, for example launching DDoS attacks or disseminating copyright material. *Auxiliary* digital evidence, on the other hand, encompasses the digital element of an otherwise non-digital crime, for example a mobile phone providing geotemporal location data, text messages or emails, in a burglary case.

In a similar way, the persons in charge of the seized digital devices can be categorised as *Suspects S*, *Victims V* or *Witnesses W*, and the devices recovered from them should be processed differently in each case in line with the role of the digital evidence in the overall investigation (*vide supra*).

Finally, the type of each digital device, for example USB keys, mobile/smart phones, HDs, needs to be considered in relation to both the overall investigation and the individual in charge of the device. These can influence the order in which the digital evidential traces should be sought (*vide infra*). Note however that it may not be necessary to examine every seized device if evidence of sufficient

probative value is recovered from the examination of higher priority devices.

The three triage elements described above are shown diagrammatically in Fig. 1, which shows examples of the devices sought in a particular case based on the roles of the digital evidence in the investigation and the persons in charge of those devices at the material time. We now consider the cases of specific evidence and auxiliary evidence separately since their roles are quite distinct.

### 3.1. Specific evidence

When investigating a case involving specific digital evidence, the persons in charge of the devices can be subdivided into two types of victim *V*, corporate and individual. Corporations and institutions usually maintain firewall, network, system and anti-malware logs to capture evidence of, for example, intrusion, denial of service, data manipulation and ex-filtration, which must be examined in addition to servers, hard disks, USB keys and other mobile devices. However, cases involving digital piracy will not in general leave digital evidence at the corporate *V*. Individuals, on the other hand, rarely maintain such logs and the majority of the digital evidence, for example the presence of backdoors, rootkits, remote access Trojans, etc. will be found on the local hard disc, mobile devices, or in the cloud. Nevertheless, the volume of data to be examined is likely to be much smaller for the individual.

Similarly, the suspects *S* may be classified as corporate or individual. The latter may be further subdivided into those that are savvy, implying that they are relatively IT literate, and individuals that are unsavvy. The evidential traces on devices belonging to a savvy individual *S* are likely to be more varied, complex and better hidden than those of an unsavvy individual *S*. In addition, the prioritisation for recovery of the traces may be different due to the different *modus operandi* of the savvy and the unsavvy individual *S*. For example, a savvy individual *S* may use a different web browser or operating system from an unsavvy individual *S*, and may attempt to use anti-forensic measures to cover their tracks.

While it is rare for there to be a witness  $W$  in a case involving specific digital evidence, a third party may be capable of either exonerating or incriminating a suspect  $S$  by means of, for example, geotemporal location data from mobile devices. (In the first instance we do not consider it necessary to subdivide the witness category.)

Each particular, prevalent digital crime results in the creation of a distinctive set of digital evidential traces. These can be analysed to form templates which determine an order for the recovery of the associated traces, prioritised on the cost-effectiveness criteria of front-loading probative value and back-loading resource utilisation.

### 3.2. Auxiliary evidence

Auxiliary digital evidence obtained from devices in the charge of witnesses, for example a mobile phone providing geotemporal location data, can play an important role in the investigation of non-digital crimes. Examination of the witnesses' devices for digital evidence (where available and relevant to the investigation) should therefore be a priority in those cases where this is likely to provide the most rapid route to confirmation or refutation of the hypothesis.

The order of examining the devices in the charge of the suspect or of the victim may be influenced by the type of non-digital crime under investigation. The auxiliary digital evidence potentially available from a corporate  $V$  includes, for example, CCTV footage and logs from alarm systems, id card and biometric readers. An individual  $V$  on the other hand may be able to provide recordings of phone and/or live conversations, photos or video footage, messages (SMS, email, chat, etc.) and geotemporal location data.

In a similar manner to the case of specific evidence above, individual suspects  $S$  may be classified as either savvy or unsavvy. Both may use, for example, Google's Streetview, Earth or Maps data for reconnaissance prior to a burglary. However, an unsavvy individual  $S$  is unlikely to attempt to hide or delete evidential traces of their activity. Furthermore, digital devices, for example computers or smart phones, and cloud-based applications, such as Twitter or Facebook, may be used by a group of savvy individual  $S$  to co-ordinate their nefarious activities. The difficulty of locating and recovering such evidence is potentially increased by the possibility of it being deleted or encrypted.

Note that the model implies that different triage template pipelines are appropriate for use with devices acquired from savvy or unsavvy suspects; witnesses, and corporate or individual victims, and also that different priorities and evidential traces are used not only for each digital crime but also for each category of suspect.

## 4. Illustrative applications of triage template pipelines

The following three scenarios demonstrate some different applications of triage template pipelines. Each specific, prevalent digital crime has its own template of prioritised digital evidence, based on the cost-effectiveness criteria of front-loading probative value and back-loading resource utilisation. Quantitative weights for evidential probative value and resource implications can be obtained

from survey based consensus of experienced, expert examiners as in (Overill et al., 2009; Cohen, 2009), or by the prioritisation matrix approach as described in (Parsonage, 2009). Here, the former approach was employed in each case to quantify the potential probative value of the expected evidential traces and the resource implications associated with recovering them, and then prioritising the screening process based on these two criteria. A major advantage of the prioritisation of the evidential screening process is that if the evidence possessing high probative value is sought first, the search can be abandoned at an early stage if these traces are not located. Furthermore, it may be possible to terminate an evidential search once evidence of 'sufficient' probative value has been located, without the need to consume further resources searching for evidence of low probative value under the law of diminishing returns (Overill et al., 2009).

A number of automated triage software tools such as *Drive Prophet* and *Triage Examiner* enable the examiner to pre-specify lists of keywords, cryptographic hash values and other items to be searched for, and these have proved very useful in practice, as reported previously (e.g. Parsonage, 2009). An advantage that the triage template pipeline approach offers over such tools is that the evidential recovery process can be either terminated or abandoned as soon as it becomes apparent that the probative value criterion either has been or will not be fulfilled, respectively. In certain types of investigation this approach will result in a shorter triage turnaround time.

In the templates for specific digital crimes (Figs. 2 and 3), the upper (green) arrows from the  $E_s$  denote that they were found, whereas the lower (red) arrows denote that they were not. HDs are regarded as primary devices and USB keys as secondary devices. One reason for this is that HDs normally contain the operating system and possess much greater storage capacity than USB keys, so it makes sense to regard them as primary in an evidential search. However, if an expected evidential trace is not recovered from the HDs, the search is moved over to the USB keys where it continues. If a further evidential trace is not found on the USB keys, the search reverts to the HDs. The interchange between target devices is represented by the purple/blue half-headed double arrows, indicating whether or not a particular evidential trace has already been found on another device. The state information embodied in this way ensures the formation of acyclic templates. Note that  $E_4$  is grey-shaded in the peer-to-peer (P2P) template diagram (Fig. 3). This represents the potential outcome that during the search for the HD copy of the original copyright material, unrelated but criminally significant material (for example, child pornography) was discovered. This would be likely to lead to a separate investigative thread being spun-off from the P2P investigation with a different focus and alignment.

### Scenario 1 – DDoS

*Hypothesis:* the seized computer has been used to launch DDoS attacks on an organisation's website.

*Assumptions:* the computer believed to have been used for launching sophisticated DDoS attacks on against an organisation's website has been seized, together with USB keys, a laptop, and mobile phones.

*Role of digital evidence:* specific.



*Person I/C device:* suspect – quick IT assessment → savvy individual S.

*Types of devices seized:* 5 USB keys (1 connected), 2 mobile phones, 1 PC with an internal HD and 2 external HDs, 1 laptop.

*Overall order of examination of devices:* HDs (internal then external, then laptop), USB keys, mobile phones, other (ISP).

Prioritised search for evidential traces: (DDoS template – Fig. 2):

(HDs – first internal, then external, then laptop)

E1: DDoS tools are present

E2: BotNet C&C program is present

E3: connections to victim's machine(s) are found (IP address, URL, log file records, etc.)

E4: connections to a BotNet C&C program are found (IP address, log file records, etc.)

(USB keys – first connected, then disconnected keys)

E1': DDoS tools are present

E2' BotNet C&C program is present

(ISP)

E5: ISP confirms seized computer accessed victim's machine(s) (IP address, etc.)

(HDs – first internal, then external, then laptop)

E6: Extortion messages to the victim are found

E7: Bragging messages are found

(mobile phones)

E6': Extortion messages to the victim are found

E7': Bragging messages are found

## Scenario 2 – P2P

*Hypothesis:* the seized computer is suspected of being used in the illegal dissemination of films and/or albums protected by copyright via a Torrent based peer-to-peer (P2P) network.

*Assumptions:* the computer believed to have been used for uploading the material has been seized together with USB keys and copies of the originals (DVDs/CDs). A mobile phone which may contain messages announcing the upload has also been seized.

*Role of digital evidence:* specific.

*Person I/C device:* suspect – quick IT assessment → unsavvy individual S.

*Types of devices seized:* 9 DVDs/CDs, 3 USB keys (1 connected), 1 mobile phone, 1 PC with internal HD.

*Overall order of examination of devices:* DVD/CD, HD, USB keys, mobile phone.

Prioritised search for evidential traces: (P2P template – Fig. 3):

(DVD/CD)

E1: Check for presence of original copyrighted material (HD)

E2: Web browser software is present

E3: Torrent client software is present

E4: HD copy of original material found (time stamps; hash values, etc.)

(USB keys – first connected, then disconnected keys)

E3': Torrent client software is present

E4': USB copy of original material found (time stamps; hash values, etc.)

E5': Torrent file and link found on USB key (creation records, etc.)

E6': Torrent file activation record found (tracker server login, MAC time, etc.)

(HD)

E5: Torrent file and link found on HD (creation records, etc.)

E6: Torrent file activation record found (tracker server login, MAC time, etc.)

E7: Internet records found (cookies, cache, history, tracker server connection, etc.)

E8: Search for messages relating to investigation (e.g. upload announcement)

E9: Search for messages relating to investigation (e.g. upload announcement)

(Mobile phone)

## Scenario 3 – burglary

*Hypothesis:* A domestic burglary took place, and a suspect is in custody.

*Assumptions:* the burglary was opportunistic, without prior reconnaissance. Auxiliary digital evidence was captured in the form of CCTV footage, a witness' emergency call and mobile phone video footage, and the domestic alarm system.

*Role of digital evidence:* auxiliary.

*Persons I/C devices:* witness W, unsavvy individual suspect S.

*Types of devices seized:* CCTV footage, witness' mobile phone, suspect's mobile phone, logs from the domestic alarm system.

*Overall order of examination of devices:* witness' mobile phone, CCTV footage, logs from the domestic alarm system, suspect's mobile phone.

Prioritised search for evidential traces:

(W mobile)

E1: Emergency call log

E2: Video footage

(CCTV)

E3: Find footage for the material time

(V alarm system)

E4: Find logs for material time

(S mobile)

E5: Temporal geolocation data

E6: Bragging messages or pictures

## 5. Conclusions and further work

While it could reasonably be argued that the triage template pipelines proposed here are in essence formalised common sense, it appears that a common triage scheme based on the cost-effectiveness criteria of front-loading probative value and back-loading resource utilisation is nevertheless a well motivated ambition.

As mentioned above, it is envisaged that regional variants of the template set, reflecting the prevalence of different digital crimes in different regions of the world (e.g. Europe, North America, East Asia, and the Middle East) will require the development of region specific template sets for those regionally prevalent digital crimes. The work thus far has addressed the set of five digital crime

templates that account for around 80% of reported digital crime in the Hong Kong Special Administrative Region of the People's Republic of China. In addition, it is to be expected that the set of digital crime templates for a given geopolitical region will vary over time, as a consequence of both technological developments and socioeconomic changes; thus the sets of crime templates will require updating on a regular basis, perhaps every three years or so. These considerations suggest that automation of the template construction process would also be well motivated. So far, the templates have been constructed manually, but once the evidential weights and resource costs are assigned to each evidential trace for a digital crime, it is quite straightforward to automatically prioritise their recovery so that high probative value traces with low resource requirements are placed early in the screening process, while traces with lower probative value and higher resource requirements are positioned towards the end. Of course, high latency evidence (particularly any that involves an MLAT, as mentioned earlier) may need to be over-prioritised in order to avoid extending the critical path of the investigation; this can also be effectively automated by shifting the initiation of its acquisition backwards in time from its assigned prioritisation by the expected latency.

In certain situations where human life or safety is at stake, a conflict may arise between examining digital evidence in a forensically sound manner and obtaining vital information leading to the swift and successful rescue of the victim(s). In such a case the information recovered may be challenged and excluded as admissible evidence in a subsequent prosecution. While it not possible to be prescriptive about such a situation we believe it merits further consideration involving members of law enforcement.

## Acknowledgements

We thank the Editor and two anonymous referees for detailed and constructive comments on an earlier version of this paper. We also thank Dr M Kwan (Hong Kong Customs & Excise Dept.) and Dr F Law (Hong Kong Police Dept.) for information regarding the preponderance of the five most prevalent digital crimes in Hong Kong.

## References

- Casey E, Ferraro M, Nguyen L. Investigation delayed is justice denied: proposals for expediting forensic examinations of digital evidence. *Journal of Forensic Sciences* 2009;54(6):1353–64.
- Cohen F. Two models of digital forensic investigation. In: *Proceedings of the 4th IEEE workshop on systematic approaches to digital forensic engineering (SADFE)*, Berkeley, CA, USA; May 21 2009. p. 42–53. Drive Prophet. <http://driveprophet.guardianf.com>.
- Garfinkel S. Forensic feature extraction and cross-drive analysis. *Digital Investigation* 2006;3S:S71–81. (Proc. DFRWS 2006).
- Kwan M, Chow KP, Law F, Lai P. Reasoning about evidence using Bayesian network. *Advances in digital forensics IV*. Springer; 2008. p. 275–89 [chapter 22].
- Kwan M, Law F. personal communications (2009–2010).
- Newman MEJ. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 2005;46(5):323–51.
- Overill RE, Silomon JAM. Six simple schemata for approximating Bayesian belief networks 27–28 June 2011. p. 65–72.
- Overill RE, K Kwan Y, P Chow K, Y Lai K, W Law Y. A cost-effective digital forensics investigation model. *Advances in digital forensics V*. Springer; 2009. p. 193–202 [chapter 15].
- Parsonage H. Computer forensics case assessment and triage. Available online at: <http://computerforensics.parsonage.co.uk/triage/ComputerForensicsCaseAssessmentAndTriageDiscussionPaper.pdf>; November 2009.
- Rogers MK, Goldman J, Mislan R, Wedge T, Debrota S. Computer forensics field triage process model. In: *Proceedings of the 1st conference on digital forensics, security & law (DFSLS)*; 2006. p. 27–40. Triage Examiner: <http://www.adfsolutions.com/products/triage-examiner>.