# List of Articles

CompSci 725 S2 2019

Every student will make an oral presentation on one of the articles listed in this document.

The themes for this year's articles are IoT Forensics, Forensic Tools and Processes, and Multimedia Forensics.

**Assignment.** By 5pm Friday 26 July, you should send an email to Clark listing three articles you are willing to present, in rank order (#1 being your most preferred article). You should identify an article by the surname of its first author and the year of its publication. For example, your list might be "1. Chung 2017; 2. Domingues 2016; 3. Li 2019".

## IoT Forensics

(Chung 2017). Hyunji Chung, Jungheum Park, and Sangjin Lee. 2017. "Digital forensic approaches for Amazon Alexa ecosystem." *Digital Investigation* 22: S15-S25. DOI: 10.1016/j.diin.2017.06.010.

**Abstract.** Internet of Things (IoT) devices such as the Amazon Echo – a smart speaker developed by Amazon – are undoubtedly great sources of potential digital evidence due to their ubiquitous use and their always-on mode of operation, constituting a human-life's black box. The Amazon Echo in particular plays a centric role for the cloud-based intelligent virtual assistant (IVA) Alexa developed by Amazon Lab126. The Alexa-enabled wireless smart speaker is the gateway for all voice commands submitted to Alexa. Moreover, the IVA interacts with a plethora of compatible IoT devices and third-party applications that leverage cloud resources. Understanding the complex cloud ecosystem that allows ubiquitous use of Alexa is paramount on supporting digital investigations when need raises. This paper discusses methods for digital forensics pertaining to the IVA Alexa's ecosystem. The primary contribution of this paper consists of a new efficient approach of combining cloud-native forensics with client-side forensics (forensics for companion devices), to support practical digital investigations. Based on a deep understanding of the targeted ecosystem, we propose a proof-of-concept tool, CIFT, that supports identification, acquisition and analysis of both native artifacts from the cloud and client-centric artifacts from local devices (mobile applications and web browsers).

(Domingues 2016). Patricio Domingues and Miguel Frade, "Digital Forensic Artifacts of the Cortana Device Search Cache on Windows 10 Desktop." 2016. *11th International Conference on Availability, Reliability and Security (ARES)*: 338-344. DOI: 10.1109/ARES.2016.44.

**Abstract.** Microsoft Windows 10 Desktop edition has brought some new features and updated other ones that are of special interest to digital forensics analysis. The search box available on the taskbar, next to the Windows start button is one of these novelties. Although the primary usage of this search box is to act as an interface to the intelligent personal digital assistant Cortana, in this paper, we study the digital forensic artifacts of the search box on machines when Cortana is explicitly disabled. Specifically, we locate, characterize and analyze the content and dynamics of the JSON-based files that are periodically generated by the Cortana device search cache system. Forensically important data from these JSON files include the number of times each installed application has been run, the date of the last execution and the content of the custom jump list of the applications. Since these data are collected per user and saved in a resilient text format, they can help in digital forensics, mostly in assisting the validation of other sources of

information.

(Li 2019). Shancang Li, Kim-Kwang Raymond Choo, IEEE, Qindong Sun, William J. Buchanan, and Jiuxin Cao. 2019. “IoT Forensics: Amazon Echo as a Use Case.” *IEEE Internet of Things Journal* 14(8): 1-17. Download

**Abstract:** Internet of Things (IoT) are increasingly common in our society, and can be found in civilian settings as well as sensitive applications such as battlefields and national security. Given the potential of these devices to be targeted by attackers, they are a valuable source in digital forensic investigations. In addition, incriminating evidence may be stored on an IoT device (e.g. Amazon Echo in a home environment and Fitbit worn by the victim or an accused person). In comparison to IoT security and privacy literature, IoT forensics is relatively under-studied. IoT forensics is also challenging in practice, particularly due to the complexity, diversity, and heterogeneity of IoT devices and ecosystems. In this paper, we present an IoT based forensic model that supports the identification, acquisition, analysis, and presentation of potential artifacts of forensic interest from IoT devices and the underpinning infrastructure. Specifically, we use the popular Amazon Echo as a use case to demonstrate how our proposed model can be used to guide forensics analysis of IoT devices.

(Oriwoh 2013). Edewede Oriwoh, David Jazani, Gregory Epiphaniou, and Paul Sant. 2013. “Internet of Things Forensics: Challenges and approaches”, *9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing:* 1-8. Download

**Abstract:** The scope of this paper is two-fold: firstly it proposes the application of a 1-2-3 Zones approach to Internet of Things (IoT)-related Digital Forensics (DF) investigations. Secondly, it introduces a Next-Best-Thing Triage (NBT) Model for use in conjunction with the 1-2-3 Zones approach where necessary and vice versa. These two ‘approaches’ are essential for the DF process from an IoT perspective: the atypical nature of IoT sources of evidence (i.e. Objects of Forensic Interest - OOFI), the pervasiveness of the IoT environment and its other unique attributes - and the combination of these attributes - dictate the necessity for a systematic DF approach to incidents. The two approaches proposed are designed to serve as a beacon to incident responders, increasing the efficiency and effectiveness of their IoT-related investigations by maximizing the use of the available time and ensuring relevant evidence identification and acquisition. The approaches can also be applied in conjunction with existing, recognised DF models, methodologies and frameworks.

## Forensic Tools and Processes

(Ahmad 2018). Ijaz Ahmad, Haider Abbas, Asad Raza, Kim-Kwang Raymond Choo, Anam Sajid, Maruf Pasha, and Farrukh Aslam Khan (2018). “Electronic crime investigations in a virtualised environment: A forensic process and prototype for evidence collection and analysis.” *Australian Journal of Forensic Sciences* 50(2): 183-208, DOI: 10.1080/00450618.2016.1229814.

**Abstract.** The constant evolution of virtualisation technologies and the availability of anti-forensic techniques and tools complicate efforts by forensic investigators to investigate a crime or a cyber security incident. Forensic collection can be complicated and requires significant efforts to investigate incidents involving contemporary technologies (*e.g.* crime launched from a virtual machine and there had been attempts to erase evidence after the incident). This paper presents a forensic process to collect and analyse traces of a virtual

machine and its corresponding manager, recorded across multiple sources including the file system, Windows registry, history, and log files from a forensic viewpoint. To demonstrate utility of the forensic mechanism, the Virtual Machine Forensic Artefact Collector (VMFAC) prototype is developed and presented in this paper.

(Quick 2014). Darren Quick, and Kim-Kwang Raymond Choo. 2014. “Google Drive: Forensic analysis of data remnants”, *Journal of Network and Computer Applications* 40: 179-193. DOI: 10.1016/j.jnca.2013.09.016.

**Abstract.** Cloud storage is an emerging challenge to digital forensic examiners. The services are increasingly used by consumers, business, and government, and can potentially store large amounts of data. The retrieval of digital evidence from cloud storage services (particularly from offshore providers) can be a challenge in a digital forensic investigation, due to virtualisation, lack of knowledge on location of digital evidence, privacy issues, and legal or jurisdictional boundaries. Google Drive is a popular service, providing users a cost-effective, and in some cases free, ability to access, store, collaborate, and disseminate data. Using Google Drive as a case study, artefacts were identified that are likely to remain after the use of cloud storage, in the context of the experiments, on a computer hard drive and Apple iPhone3G, and the potential access point(s) for digital forensics examiners to secure evidence.

(Roussev 2013). Vassil Roussev, Candice Quates, and Robert Martell. 2013. “Real-time digital forensics and triage.” *Digital Investigation* 10(2): 158-167. Download

**Abstract:** There are two main reasons the processing speed of current generation digital forensic tools is inadequate for the average case: a) users have failed to formulate explicit performance requirements; and b) developers have failed to put performance, specifically latency, as a top-level concern in line with reliability and correctness.

In this work, we formulate forensic triage as a real-time computation problem with specific technical requirements, and we use these requirements to evaluate the suitability of different forensic methods for triage purposes. Further, we generalize our discussion to show that the complete digital forensics process should be viewed as a (soft) real-time computation with well-defined performance requirements.

We propose and validate a new approach to target acquisition that enables file-centric processing without disrupting optimal data throughput from the raw device. We evaluate core forensic processing functions with respect to processing rates and show their intrinsic limitations in both desktop and server scenarios. Our results suggest that, with current software, keeping up with a commodity SATA HDD at 120 MB/s requires 120–200 cores.

(Singh 2016). Bhupendra Singh, and Upasna Singh. 2016. “A forensic insight into Windows 10 Jump Lists.” *Digital Investigation* 17: 1-13. DOI: 10.1016/j.diin.2016.02.001.

**Abstract.** The records maintained by Jump Lists have the potential to provide a rich source of evidence about users’ historic activity to the forensic investigator. The structure and artifacts recorded by Jump Lists have been widely discussed in various forensic communities since its debut in Microsoft Windows 7. However, this feature has more capabilities to reveal evidence in Windows 10, due to its modified structure. There is no literature published on the structure of Jump Lists in Windows 10 and the tools that can successfully parse the Jump Lists in Windows 7/8, do not work properly for Windows 10. In this paper, we have identified the structure of Jump Lists in Windows 10 and compared it with Windows 7/8. Further, a proof-of-concept tool called JumpListExt (Jump List Extractor) is developed on the basis of identified structure that can parse Jump Lists in Windows 10, individually as well as collectively. Several experiments were conducted to

detect anti-forensic attempts like evidence destruction, evidence modification and evidence forging carried out on the records of Jump Lists. Furthermore, we demonstrated the type of artifacts recorded by Jump Lists of four popular web browsers with normal and private browsing mode. Finally, the forensic capability of Jump Lists in Windows 10 is demonstrated in terms of activity timeline constructed over a period of time using Jump Lists.

## Multimedia Forensics

(Al-Mohair 2015). Hani K. Al-Mohair, Junita Mohamad Saleh, and Shahrel Azmin Suandi. 2015. "Hybrid Human Skin Detection Using Neural Network and K-Means Clustering Technique." *Applied Soft Computing* 33(C): 337-347. Download

**Abstract:** Human skin detection is an essential step in most human detection applications, such as face detection. The performance of any skin detection system depends on assessment of two components: feature extraction and detection method. Skin color is a robust cue used for human skin detection. However, the performance of color-based detection methods is constrained by the overlapping color spaces of skin and non-skin pixels. To increase the accuracy of skin detection, texture features can be exploited as additional cues. In this paper, we propose a hybrid skin detection method based on YIQ color space and the statistical features of skin. A Multilayer Perceptron artificial neural network, which is a universal classifier, is combined with the k-means clustering method to accurately detect skin. The experimental results show that the proposed method can achieve high accuracy with an F1-measure of 87.82% based on images from the ECU database.

(De Bock 2016). Johan De Bock, Patrick De Smet. 2016. "JPGcarve: An advanced tool for automated recovery of fragmented JPEG files." *IEEE Transactions on Information Forensics and Security* 11(1): 19-34. Download

**Abstract:** In this paper, we present a new tool for forensic recovery of single and multi-fragment JPEG/JFIF data files. First, we discuss the basic design and the technical methods composing our proposed data carving algorithm. Next, we compare the performance of our method with the well-known Adroit Photo Forensics (APF) state- of-the art tool. This comparison is centered on both the carving results as well as the obtained data processing speed, and is evaluated in terms of the results that can be obtained for several well-known reference data sets. Important to note is that we specifically focus on the fundamental recovery and fragment matching performance of the tools by forcing them to use various assumed cluster sizes. We show that on all accounts our new tool can significantly outperform APF. This improvement in data processing speed and carving results can be mostly attributed to novel methods to iterate and reduce the data search space and to a novel parameterless method to determine the end of a fragment based on the pixel data. Finally, we discuss several options for future research.

(Hosseini 2019). Morteza Darvish Morshedi Hosseini, and Miroslav Goljan. 2019. "Camera Identification from HDR Images." *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*: 69-76. DOI: 10.1145/3335203.3335717.

**Abstract:** Performance of camera identification methods based on PRNU is very sensitive to geometric operations applied to images during acquisition and processing. Handling images that have been geometrically transformed, such as rotated, downsampled, and/or cropped requires overcoming pixel desynchronization problem. This work expands applicability of camera identification methods based on PRNU to the class of HDR images. Geometric transformations in HDR images revealed in this work are reversed in a series of

steps involving block-wise PRNU matching. Efficiency of this method is then tested on HDR images from publicly available UNIFI dataset spanning 26 cameras of mobile devices.

(Marra 2017). Francesco Marra, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. 2017. "Blind PRNU-based image clustering for source identification." *IEEE Transactions on Information Forensics and Security* 12(9): 2197-2211. Download

**Abstract:** We address the problem of clustering a set of images, according to their source device, in the absence of any prior information. Image similarity is computed based on noise residuals, regarded as single-image estimates of the camera's photo- response non-uniformity (PRNU) pattern. First, residuals are grouped by correlation clustering, and several alternative data partitions are computed as a function of a running decision boundary. Then, these partitions are processed jointly to extract a single, more reliable, consensus clustering and, with it, more reliable PRNU estimates. Finally, both clustering and PRNU estimates are progressively refined by merging pairs of the same-PRNU clusters, selected on the basis of a maximum-likelihood ratio statistic. Extensive experiments prove the proposed method to outperform the current state of the art both on pristine images and compressed images downloaded from social networks. A remarkable feature of the method is that it does not require the user to set any parameter, nor to provide a training set to estimate them. Moreover, through a suitable choice of basic tools, and efficient implementation, complexity remains always quite limited.

(Mohanty 2019). Manoranjan Mohanty, Ming Zhang, Muhammad Rizwan Asghar, and Giovanni Russello. 2019. "e-PRNU: Encrypted Domain PRNU-Based Camera Attribution for Preserving Privacy." *IEEE Transactions on Dependable and Secure Computing*: 1-12. Download

**Abstract:** Photo Response Non-Uniformity (PRNU) noise-based source camera attribution is a popular digital forensic method. In this method, a camera fingerprint computed from a set of known images of the camera is matched against the extracted noise of an anonymous questionable image to find out if the camera had taken the anonymous image. The possibility of privacy leak, however, is one of the main concerns of the PRNU-based method. Using the camera fingerprint (or the extracted noise), an adversary can identify the owner of the camera by matching the fingerprint with the noise of an image (or with the fingerprint computed from a set of images) crawled from a social media account. In this article, we address this privacy concern by encrypting both the fingerprint and the noise using the Boneh-Goh-Nissim (BGN) encryption scheme, and performing the matching in encrypted domain. To overcome leakage of privacy from the content of an image that is used in the fingerprint calculation, we compute the fingerprint within a trusted environment, such as ARM TrustZone. We present e-PRNU that aims at minimizing privacy loss and allows authorized forensic experts to perform camera attribution. The security analysis shows that the proposed approach is semantically secure.

(Perez 2017). Mauricio Perez, Sandra Avila, Daniel Moreira, Daniel Moraes, Vanessa Testoni, Eduardo Valle, Siome Goldenstein, and Anderson Rocha. 2017. "Video pornography detection through deep learning techniques and motion information." *Neurocomputing* 230(22): 279-293. DOI: 10.1016/j.neucom.2016.12.017.

**Abstract:** Recent literature has explored automated pornographic detection – a bold move to replace humans in the tedious task of moderating online content. Unfortunately, on scenes with high skin exposure, such as people sunbathing and wrestling, the state of the art can have many false alarms. This paper is based on the premise that incorporating motion information in the models can alleviate the problem of mapping skin exposure to

pornographic content, and advances the bar on automated pornography detection with the use of motion information and deep learning architectures. Deep Learning, especially in the form of Convolutional Neural Networks, have striking results on computer vision, but their potential for pornography detection is yet to be fully explored through the use of motion information. We propose novel ways for combining static (picture) and dynamic (motion) information using optical flow and MPEG motion vectors. We show that both methods provide equivalent accuracies, but that MPEG motion vectors allow a more efficient implementation. The best proposed method yields a classification accuracy of 97.9% – an error reduction of 64.4% when compared to the state of the art – on a dataset of 800 challenging test cases. Finally, we present and discuss results on a larger, and more challenging, dataset.

(Phillips 2018). Jonathon Phillips, Amy N. Yates, Ying Hu, Carina A. Hahn, Eilidh Noyes, Kelsey Jackson, Jacqueline G. Cavazos, Géraldine Jeckeln, Rajeev Ranjan, Swami Sankaranarayanan, Jun-Cheng Chen, Carlos D. Castillo, Rama Chellappa, David White, and Alice J. O'Toole. 2018. "Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms." *Proceedings of the National Academy of Sciences* 115 (24): 6171-6176. DOI: 10.1073/pnas.1721355115

**Abstract**. Achieving the upper limits of face identification accuracy in forensic applications can minimize errors that have profound social and personal consequences. Although forensic examiners identify faces in these applications, systematic tests of their accuracy are rare. How can we achieve the most accurate face identification: using people and/or machines working alone or in collaboration? In a comprehensive comparison of face identification by humans and computers, we found that forensic facial examiners, facial reviewers, and superrecognizers were more accurate than fingerprint examiners and students on a challenging face identification test. Individual performance on the test varied widely. On the same test, four deep convolutional neural networks (DCNNs), developed between 2015 and 2017, identified faces within the range of human accuracy. Accuracy of the algorithms increased steadily over time, with the most recent DCNN scoring above the median of the forensic facial examiners. Using crowd-sourcing methods, we fused the judgments of multiple forensic facial examiners by averaging their rating-based identity judgments. Accuracy was substantially better for fused judgments than for individuals working alone. Fusion also served to stabilize performance, boosting the scores of lower-performing individuals and decreasing variability. Single forensic facial examiners fused with the best algorithm were more accurate than the combination of two examiners. Therefore, collaboration among humans and between humans and machines offers tangible benefits to face identification accuracy in important applications. These results offer an evidence-based roadmap for achieving the most accurate face identification possible.

(Rossy 2012). Quentin Rossy, Sylvain Ioset, Damien Dessimoz, and Olivier Ribaux. 2013. "Integrating forensic information in a crime intelligence database". *Forensic Science International* 230(1-3), 137-146. DOI: 10.1016/j.forsciint.2012.10.010.

**Abstract.** Since 2008, intelligence units of six states of the western part of Switzerland have been sharing a common database for the analysis of high volume crimes. On a daily basis, events reported to the police are analysed, filtered and classified to detect crime repetitions and interpret the crime environment. Several forensic outcomes are integrated in the system such as matches of traces with persons, and links between scenes detected by the comparison of forensic case data. Systematic procedures have been settled to integrate links assumed mainly through DNA profiles, shoemarks patterns and images. A statistical outlook on a retrospective dataset of series from 2009 to 2011 of the database informs for instance on the number of repetition detected or confirmed and increased by

forensic case data. Time needed to obtain forensic intelligence in regard with the type of marks treated, is seen as a critical issue. Furthermore, the underlying integration process of forensic intelligence into the crime intelligence database raised several difficulties in regards of the acquisition of data and the models used in the forensic databases. Solutions found and adopted operational procedures are described and discussed. This process form the basis to many other researches aimed at developing forensic intelligence models.

(Uzun 2015). Erkam Uzun and Hüsrev Taha Sencar. 2015. “Carving orphaned JPEG file fragments.” *IEEE Transactions on Information Forensics and Security* 10(9): 1549-1563. Download

**Abstract:** File carving techniques allow for recovery of files from storage devices in the absence of any file system metadata. When data are encoded and compressed, the current paradigm of carving requires the knowledge of the compression and encoding settings to succeed. In this paper, we advance the state of the art in JPEG file carving by introducing the ability to recover fragments of a JPEG file when the associated file header is missing. To realize this, we examined JPEG file headers of a large number of images collected from Flickr photo sharing site to identify their structural characteristics. Our carving approach utilizes this information in a new technique that performs two tasks. First, it decompresses the incomplete file data to obtain a spatial domain representation. Second, it determines the spatial domain parameters to produce a perceptually meaningful image. Recovery results on a variety of JPEG file fragments show that given the knowledge of Huffman code tables, our technique can very reliably identify the remaining decoder settings for all fragments of size 4 KiB or above. Although errors due to detection of image width, placement of image blocks, and color and brightness adjustments can occur, these errors reduce significantly when fragment sizes are >32 KiB.