cellebrite
delivering mobile expertise

White Paper

EXTRACTING LEGALLY DEFENSIBLE EVIDENCE FROM THE CLOUD

Explaining UFED Cloud Analyzer Extraction and Analysis Processes

With data privacy a major topic of discussion among consumers and service providers in most countries, investigators need to be able to certify that the processes they used to collect cloud-based evidence are legally defensible and forensically sound.

The current process relies on a chain of legal paperwork, including preservation orders, subpoenas, and/or search warrants, to secure evidence directly from cloud service providers. However, as the National Institute of Standards and Technology noted in its 2014 "NIST Cloud Computing Forensic Science Challenges" draft report, [1] this is limited for many different reasons.[2]

NIST identified thirteen major factors that challenge the identification, collection, and preservation of cloud-based media. While some of these issues, such as understanding cloud topology, policies, and storage systems, are beyond its scope, Cellebrite developed UFED Cloud Analyzer to help investigators identify the right accounts, forensically preserve private media within the account, and reduce problems associated with media volume.

As NIST notes in its report, "The cloud exacerbates many technological, organizational, and legal challenges already faced by digital forensics examiners." This paper discusses UFED Cloud Analyzer in the context of NIST's report, specifically the items outlined in Annex B.[3]

---

1 National Institute of Standards and Technology, "NIST Cloud Computing Forensic Science Challenges," June 2014, http://csrc.nist.gov/publications/drafts/nistir-8006/draft_nistir_8006.pdf accessed March 2015 ,19

2 Outlined in Items 47-42 in Annex B of the report.

3 Not every data point listed in Annex B of NIST's paper is relevant. These are noted in a sidebar.

Executive Summary

Among the dozens of issues NIST identified were forensic issues around cloud storage, user privacy, and logistical concerns. While many of these are grounded in traditional computer forensics concepts, and may or may not be applicable to cloud investigations, some foundational forensic principles - validation, authentication, repeatability - are universal to any forensic data collection and analysis.

UFED Cloud Analyzer's extraction approach begins with user data, including credentials, found on a subject mobile device and extracted with the proper legal authority. This account-based approach selectively acquires data associated only with a specific user, using data artifacts found on the subject device to narrow the scope of a search to certain date/time frames and content types. This approach effectively preserves the privacy of other tenants collocated on the same server, and minimizes issues with evidence being scattered around different storage locations.

UFED Cloud Analyzer also promotes forensic best practices around validation and authentication by relying on provider APIs to perform extractions. A Cloud Analyzer extraction hashes each individual artifact and, separately, its associated metadata. Not only does this ensure repeatability; it also allows for proper validation using records obtained directly from the service provider.

Finally, UFED Cloud Analyzer normalizes and correlates evidence from multiple accounts and disparate data formats, reducing the risk of missing content and context through the use of timelines and visual tools.

In short, UFED Cloud Analyzer helps to mitigate the risks associated with slower-than-desired responses to legal process,[4] whether this is due to demand on service providers, provider reluctance to comply with government requests for private information, or providers residing in jurisdictions that are not part of the MLAT treaty. While it may not negate the need for records requests from providers, it does enable law enforcement to validate what providers offer via the use of hashing and data identifiers, and on the flip side, for provider data to validate its own extractions.

[4] Annex B, Item 51

UFED Cloud Analyzer Basics

UFED Cloud Analyzer is extraction and analysis software that can be installed on any Windows-based PC platform. It is designed to import a file that contains account credentials from popular cloud services. This account package can be exported from UFED Physical Analyzer following a file system or physical extraction of a smartphone's memory.[5] Alternatively, investigators can manually enter usernames and passwords provided by users with documented consent. Then, UFED Cloud Analyzer uses the provider's application programming interface (API) to collect "snapshots" of private cloud-based evidence.[6]

UFED Cloud Analyzer users should adhere to best practices around forensic cloud extraction:

- **Serve the provider with a preservation order[7] and, if necessary, a nondisclosure order for the account(s) in question. Obtain the level of legal authority to search that is appropriate for the examiner's country and jurisdiction.**
- **Extract cloud-based evidence to a storage medium specifically designed and prepared for that purpose: a flash drive, external hard drive, a location on an internal forensic network, or internal drive or partition within the forensic computer.**

[5] Assuming that the account credential data is unencrypted or can be decrypted.

[6] This is comparable to a logical mobile device extraction undertaken with UFED Touch or UFED 4PC.

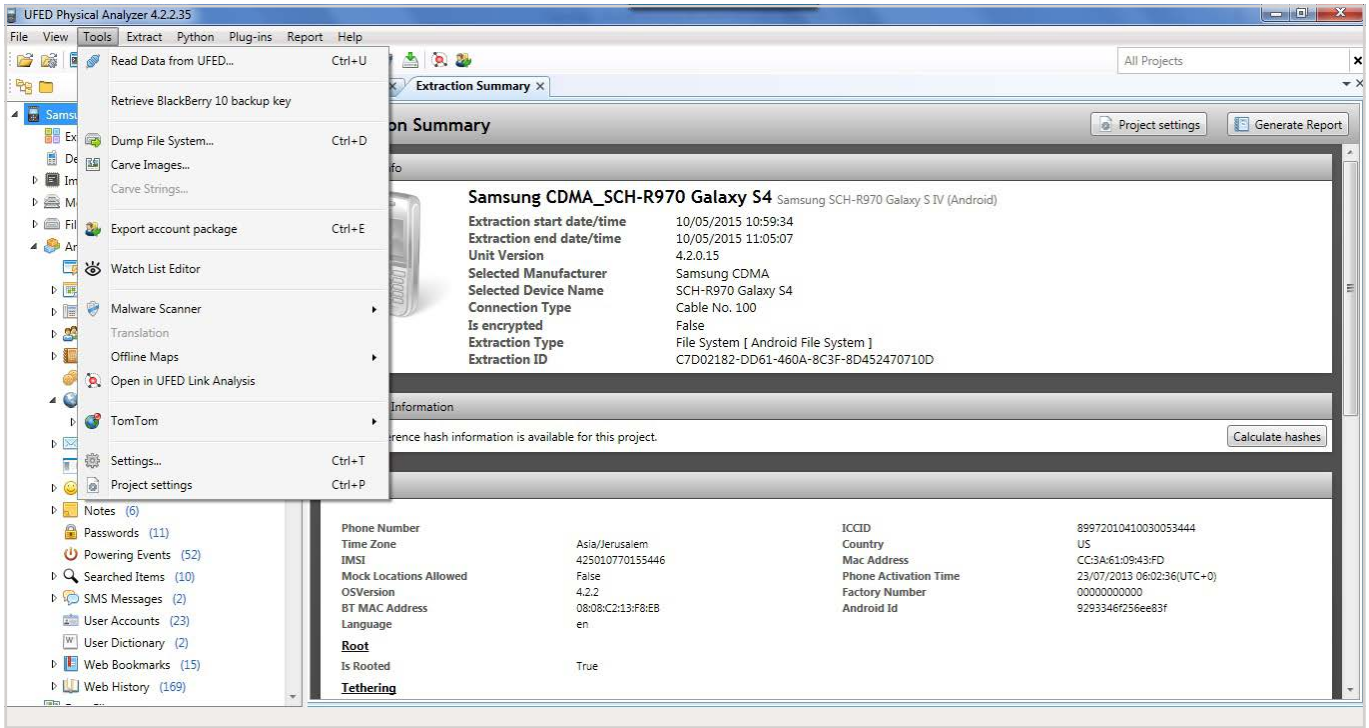
[7] In the United States, a preservation order is defined under 18 US Code § 2703(f), also known as the Stored Communications Act.

Solving cloud extraction problems with an account-based approach

Identifying evidence in the cloud is a challenge because of data volume.[8] Too much, and a search might be overbroad; too little, and investigators could miss important data.

Evidence extraction using UFED Cloud Analyzer starts with existing artifacts extracted and decoded from the user's mobile device. This helps to identify likely sources of evidence, reducing the chance of missing either inculpatory or exculpatory data. Timeline and data type particulars, which should be specified in a search warrant or other legal authority to search, prevent overbroad searches.

Assuming the investigator has the mobile device or can quickly identify a suspect, the mobile device can help confirm a suspect's true identity and account ownership.[9] In other words, using a suspect's mobile device to obtain login credentials means that investigators are in a better position to authenticate the evidence.



[8] Annex B, Item 32.

[9] Annex B, Item 58.

It's difficult for investigators to identify unauthorized third-party data access,[12] whether by persons in a suspect's or victim's life, or strangers. Although UFED Cloud Analyzer cannot differentiate between account hijacking and investigator access, the same process used to authenticate cloud-based evidence – comparing it to mobile device evidence – can identify anomalous activity.

In other words, content that doesn't read or sound like the device or account user, or that was created during a timeline when the device/account user could not have created it, can indicate unauthorized access.

Once an investigator has identified evidence, "imaging" the cloud[13] is, in NIST's word, "impractical" - even though it's recommended because of the difficulty providers have in responding to subpoenas.[14] High volumes of evidence and location issues make it difficult to image an entire cloud server.

Furthermore, data may be scattered on several servers, mixed among data pieces belonging to other accounts. This makes the computer forensics concept of "imaging" a hard drive impractical, and probably not applicable for the realm of cloud-stored data. As NIST states, however, "Partial imaging may have legal implication in the presentation to the court."

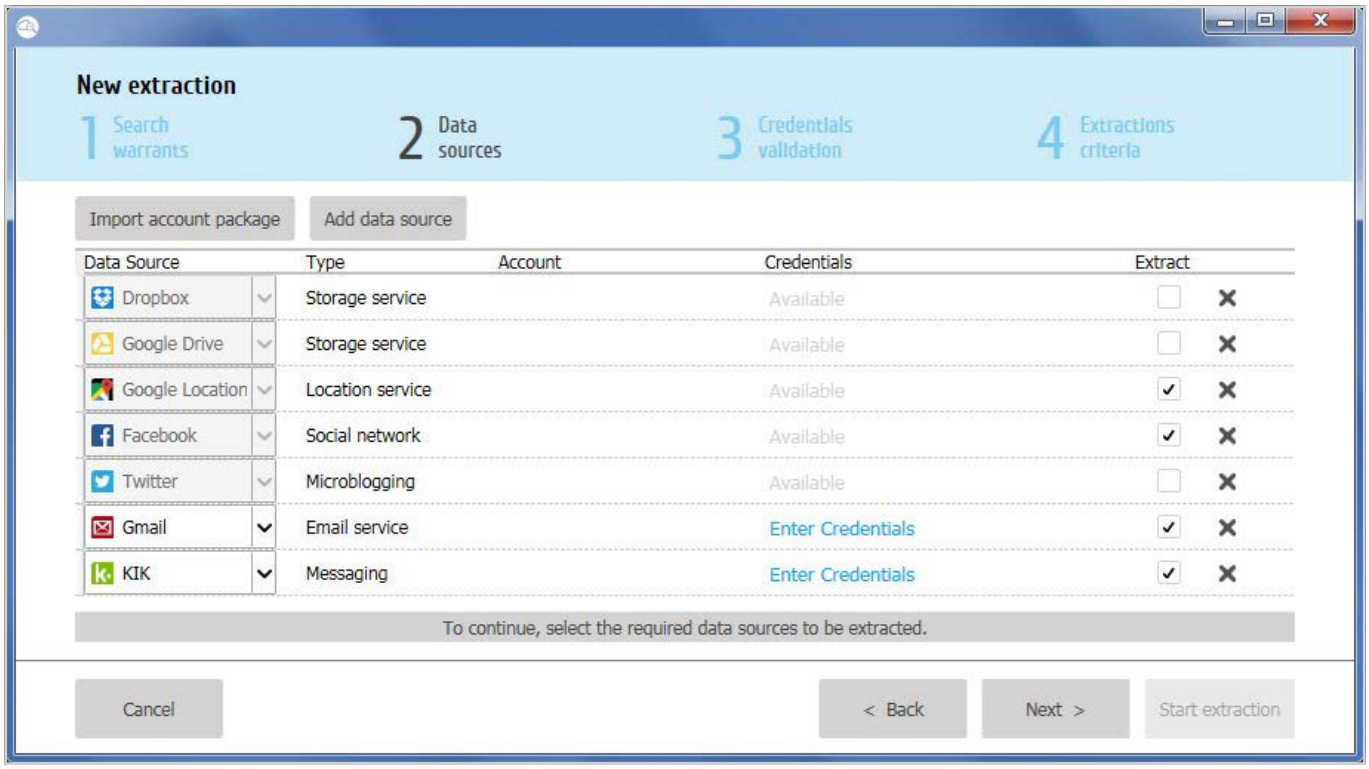
DEALING WITH FICTITIOUS CLOUD ACCOUNTS

The process of identifying and authenticating evidence becomes more complicated when the user's identity is fictitious.[10]

"A criminal can trivially obtain credit card numbers, and then create fake profiles on existing legitimate social media websites to make his/her cloud identity appear to have a corresponding equivalent in the 'real world,'" NIST's report states. "A forensic investigator is then faced with the daunting challenge of obtaining data on the criminal identity from multiple online entities, many of which are geographically spread around the world."

However, even in cases where the suspect has used a "burner phone" to further conceal their identity, commonalities will likely exist between device(s) and cloud account(s) such that investigators will be able to tie devices and accounts to an actual person.[11]

[10] Annex B, Item 59.
[11] Annex B, Item 60.
[12] Annex B, Item 61.
[13] Annex B, Item 38.
[14] Annex B, Item 48.



UFED Cloud Analyzer's account-based approach reduces these risks by focusing on a specific user's data, no matter where it's stored. This selective data acquisition, NIST acknowledges, "is a challenge because prior knowledge about relevant data sources is often difficult to obtain in a cloud environment."[15]

With that prior knowledge coming from the account holder's mobile device, an investigator can reduce the scope of search. That's because UFED Cloud Analyzer relies upon not only user credentials, but also existing artifacts extracted and decoded from the user's mobile device. These help to narrow the field of cloud data, including a relevant time/date range and specific cloud services involved.

By focusing on user accounts and credentials, investigators needn't worry that an unrelated party's data bled over into their suspect's or victim's data (or that key evidence might be found in an unrelated account), and can be assured that the evidence maintains its integrity.[16]

User data stored in the cloud can be encrypted,[17] and encryption is gaining ground among major cloud service providers. Whether providers[18] hold encryption keys as part of their service, or follow Apple's and Google's lead by making keys local to users alone, UFED Cloud Analyzer's reliance on user credentials eliminates the loss of ability to decrypt data.

15 Annex B, Item 39.

16 Annex B, Item 41.

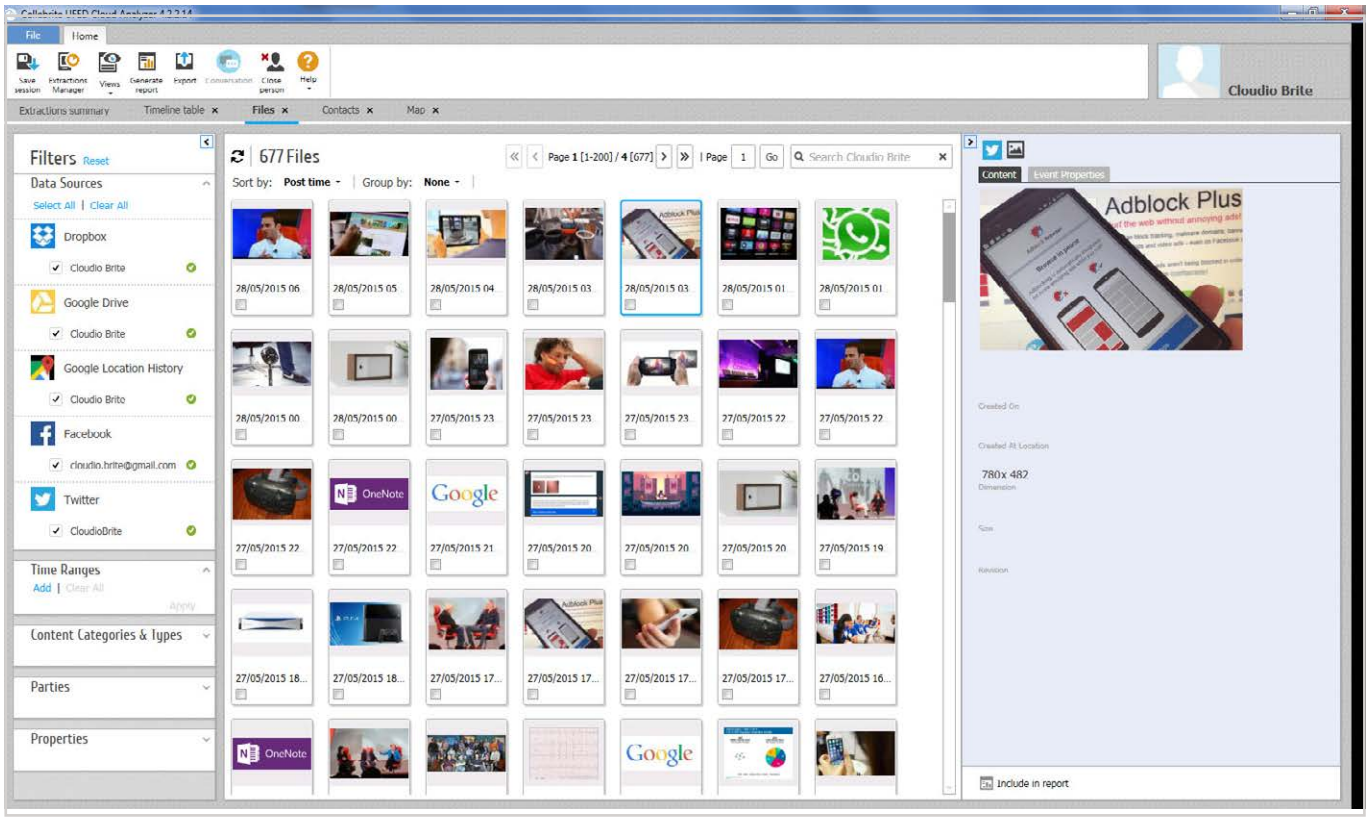
17 Annex B, Item 40.

18 Vijayan, Jaikumar, "Cloud computing 2014: Moving to a zero trust security model," ComputerWorld, December 2013 ,13, http://www.computerworld.com/article/2487123/data-privacy/cloud-computing--2014-moving-to-a-zero-trust-security-model.html accessed March 2015 ,19

Solving forensic cloud storage problems: provider APIs

Headlining Annex B of NIST's report is deleted cloud data. "Attributing deleted data to a specific user" and "Recovering overwritten data,"[19] as well as dynamic storage[20] speak to the need for investigators to serve a provider with a preservation order.

Preservation orders mitigate some of the risks of evidence being deleted before an investigator has a chance to secure it. Even when data has been deleted, a preservation order increases the likelihood that the evidence will be extracted, rather than overwritten when the nodes pointing to it are deleted.



That's because UFED Cloud Analyzer relies upon the service provider's API to extract data. If the API allows, an investigator can access deleted or archived data. However, along with identifying data via hash value, the use of account credentials means that UFED Cloud Analyzer does allow data - deleted or live - to be attributed to a specific user. Cellebrite continues to add support for data available with the API method as UFED Cloud Analyzer evolves.

19 Annex B, Items 2-1.

20 Annex B, Item 33.

	Messages	Images	Videos	Files	Contacts	Locations
Facebook	+	+	+		+	
Twitter	+				+	
Gmail	+					
Dropbox		+	+	+		
Google Drive		+	+	+		
Kik					+	

API use also reduces the risk associated with the need to reconstruct virtual images or storage[21] because provider records obtained with a search warrant can validate API-based evidentiary collections. No additional validation of reconstruction algorithms is required.

Finally, API use removes the need to confiscate or seize entire cloud resources to acquire evidence.[22] Because cloud resources often have multiple tenants, API use reduces risks to tenants other than the one(s) under investigation.[23] By the same token, use of credentials obtained from a user's mobile device means that there is no need to segregate forensic data.[24]

Evidentiary issues: preservation, authentication, & validation

The ability to validate, repeat and reproduce a process is foundational to any forensic science, and the primary criteria courts use to determine evidence admissibility. Digital forensics is an evolving discipline, and as NIST pointed out, cloud forensics' testability, validation, and scientific principles have not been widely addressed.[25]

Adding to this, the cloud system is volatile and likely to change following collection.[26] "Therefore it is impossible for a third party to verify, after acquisition, that the data collected is correct because the data is no longer the same as at the time of acquisition," NIST's report notes.

[21] Annex B, Items 4.
[22] Annex B, Item 20.
[23] Annex B, Item 48.
[24] Annex B, Item 21.
[25] Annex B, Item 62.
[26] Annex B, Item 34.

Shipley and Bowker (2014), quoting Merritt (2012), note that not only must online communication be authenticated as to its proponent's claim that it is evidence; the result of investigative efforts must also accurately preserve the original message, including associated metadata.[27] Their assertion: "A hash value must be calculated either during the actual collection of the data or as soon as possible after the data is saved electronically."[28]

This is also important when considering using metadata as an authentication method.[29] UFED Cloud Analyzer's use of suspect credentials does not change metadata to the extent that it would present problems in either civil or criminal trials. While credential verification may generate a signature within the cloud provider - e.g. creating a log on the server side (for example, within the Facebook activity log) - it also creates two separate hashes for each piece of data it collects:[30]

- Each artifact - tweet, private message, image, etc. - together with its associated metadata receives its own separate hash.
- Files such as images and documents are hashed separately from the posts to which they are connected.

NIST specifically calls out training as an important element of proper cloud forensics.[31] Investigators need training not just on cloud forensics policy and procedure, but also the foundation of cloud computing technology.

"Most digital forensic training materials are outdated and are not applicable in cloud environments. The lack of knowledge about cloud technology may interfere with remote investigations where systems are not physically accessible and there is an absence of proper tools to effectively investigate the cloud computing environment," the report notes.

UFED Cloud Analyzer training is, and will be, available in the form of webinars, a user manual, and instructor-led training courses.

27 Shipley, Todd, and Bowker, Art, "Investigating Internet Crimes: An Introduction to Solving Crimes in Cyberspace," Syngress, p. 78.

28 Shipley & Bowker, p. 81.

29 Annex B, Item 7.

30 Further information about this process is described in the UFED Cloud Analyzer user guide.

31 Annex B, Item 65.

It should be possible to validate UFED Cloud Analyzer results by downloading content (for example, a photo) from a service provider, hashing it, then comparing the manual hash with the software-generated hash. Investigators can also approach the service provider to ask for any given data identifier, then hash it to compare with the hash created at the time of extraction.

LEGAL ISSUES WITH CLOUD FORENSICS

UFED Cloud Analyzer can help to mitigate some, but not all, issues associated with legal risks. One of the major challenges: limitations in international collaboration and cross-nation legislative mechanisms.[32]

Existing legal processes, such as the MLAT, were built with physical evidence in mind. They are lengthy and complex; therefore, they are not suitable for the digital era, where the dynamic nature and fast pace of data - and, at least by some opinion, the less deterministic nature of cloud-based data's physical location - requires speedy process responses.

This is at least partially solved by UFED Cloud Analyzer's flexibility in allowing investigators in each nation to apply relevant laws. For example, investigators in countries that legally view the mobile device as a portal to cloud data need not apply for a new search warrant, while investigators in countries that view device and cloud as separate storage entities require a warrant for each entity.[33]

Privacy, of course, continues to be a critical conversational element.[34] Again, UFED Cloud Analyzer seeks to reduce risks to personal, business, and government information by limiting investigative searches only to certain timeframes and certain types of data, which ideally should match the content and timeframes specified on the search warrant. The software also logs each user and extraction performed.

Finally, investigators concerned about the limitations of their investigative power[35] should work with appropriate counsel to determine the extent of limitations and how they might affect an investigation, including the use of UFED Cloud Analyzer.

[32] Annex B, Item 52.

[33] The question of whether an investigator can apply for a domestic search warrant to search the account(s) of a citizen within their own borders - even if the account is with a cloud service provider located in a foreign country, with the potential for data to be stored on a server physically located in a third country - has yet to be determined in a court of law.

[34] Annex B, Item 56.

[35] Annex B, Item 46.

Normalizing and correlating evidence from multiple accounts

NIST's report notes, "Faults occur either intentionally or accidentally and consist of missed content, contextual information, meaning of content, process elements, relationships, ordering, timing, location, corroborating content, consistencies, and inconsistencies… [in] multiple computers in multiple locations under control of multiple parties."[36]

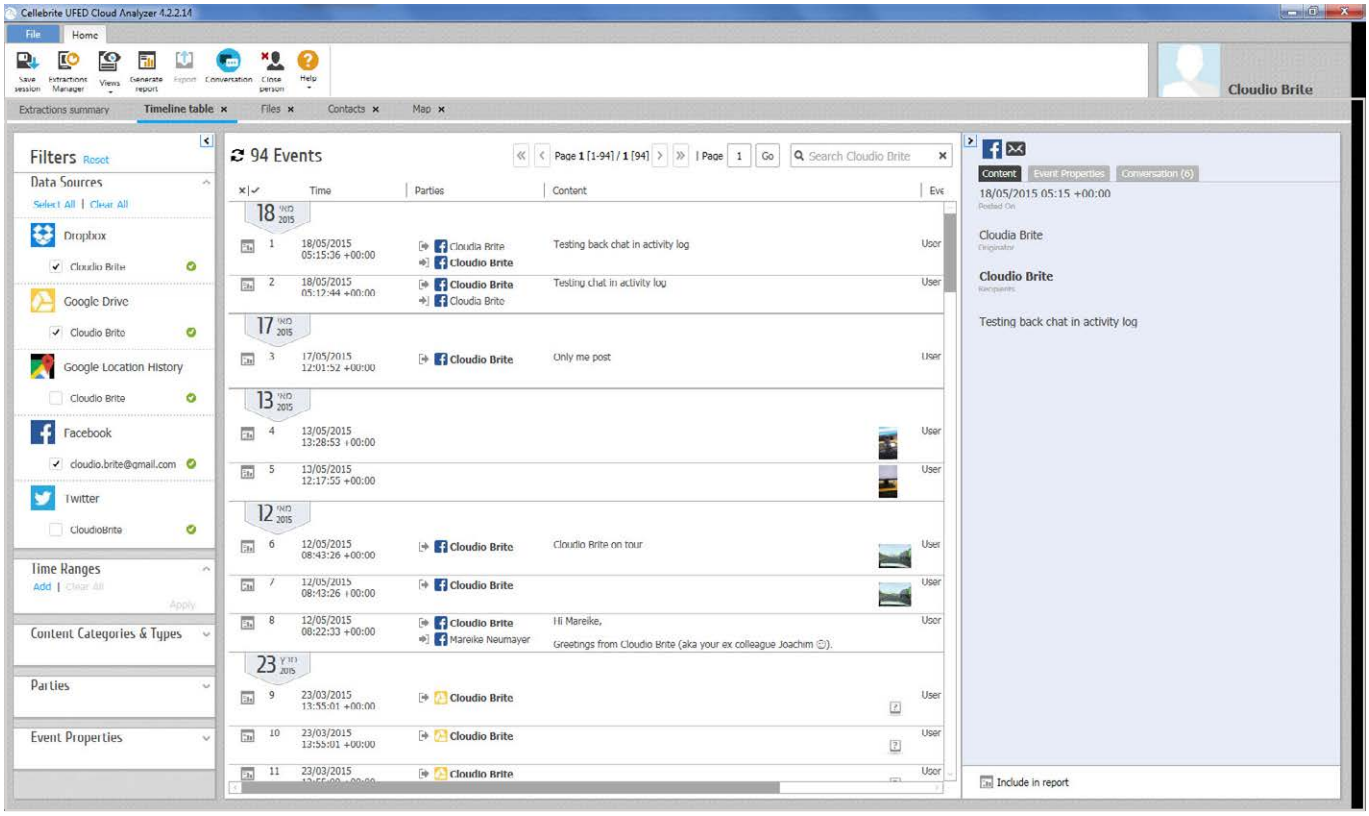
Along with multiple points of failure is the risk of a single point of failure,[37] for instance, service outages. UFED Cloud Analyzer reduces the risk posed by single points of failure because even if a platform is experiencing an outage at the time of collection, the examiner can always go back once the problem is resolved.

Internal processes, such as notifications to the user when such an issue occurs and retry attempts, are built into the software. When the software identifies that the connection with the service is down, the extraction automatically stops. At this point extractions cannot be resumed, and the user has to start the extraction process over for that particular data source.

36 Annex B, Item 42.

37 Annex B, Item 10.

Moreover, NIST states, "For all investigators, collection and analysis of data from distributed and disparate sources is challenging because perpetrators can use services from different providers."[38] Indeed, it isn't often that data from just a single social media, filesharing, or location-based data account (or mobile device) will enable an investigator to build a case. Data from multiple accounts contextualizes a suspect's or victim's activities and shows an investigator's due diligence in building a case.



UFED Cloud Analyzer reduces the risk of missing content,[39] and its context and meaning, by allowing investigators to view and capture it in that context, and also, in conjunction with UFED Link Analysis, to put it in context of other data available from a suspect's mobile device(s) or operator's call detail records. This way investigators gain further insight into how evidence correlates among devices and services.

38 Annex B, Item 11.

39 Annex B, Item 37.

The ability to normalize and correlate data can show gaps in a timeline that warrant further investigation, whether evidence appears in a victim's account but not a suspect's, or witness statements indicate that evidence should exist in a given device or cloud account.

In addition, the normalized dataset offered by UFED Cloud Analyzer - with or without UFED Link Analysis - enhances collaboration by providing an easily visualized dataset enabling communication about the data and persons of interest. While this doesn't solve all the concerns with international collaboration,[40] including the need for timely enough responses, it's an improvement.

Finally, NIST's report noted a lack of interoperability between cloud providers,[41] including a lack of insight or control over providers'[41] proprietary architecture, challenges investigators when they try to correlate activity across platforms.[42]

While identifying the similarities and differences in architectures for validation purposes is beyond UFED Cloud Analyzer's scope, the software does seek to provide some consistency by normalizing the data from disparate providers, including log and data source formats.

CONSTRUCTING TIMELINES; LOGS AND METADATA

Timestamp synchronization across physical machines potentially located in different geographies[43] presents a challenge in that timestamps can end up being inconsistent. Therefore, it becomes more difficult for investigators to construct a timeline of events around a crime or other illicit activity.

UFED Cloud Analyzer handles dates and times in a similar manner as UFED Physical Analyzer: presenting both local time and Coordinated Universal Time (UTC). The time presented in Cloud Analyzer, including format and time zone, comes from the cloud data provider via API. It is presented "as is" to the user. However, UFED Cloud Analyzer normalizes all dates and times into a single chronological order.

Although UFED Cloud Analyzer currently cannot align local time to a specific time zone, as UFED Physical Analyzer can, use of the API means that most time and date stamps from cloud posts already contain UTC. It also means that provider records will show the same times and dates, and thus validate, API-based extractions.

40 Annex B, Item 55.

41 Annex B, Item 9.

42 Annex B, Item 3.

43 Annex B, Item 5.

Remaining cloud forensics challenges

To be sure, some challenges remain and will require ongoing discussion. These include:

- Determining the source of an unauthorized change to a user's cloud computing environment.[44]
- Geolocation unknowns and resulting jurisdictional issues that can affect the chain of custody.[45] It's likely that NIST was referring to the number of unknown people who might be involved with preserving and collecting data located on physical hard drives in various locations, so whether this detail is important has yet to be determined, especially in the courts.
- Lack of transparency in the cloud's operational details, provider's API use notwithstanding.[46]
- Identifying criminal organizations' "cells" which can operate independently, and with no way to associate them, because of the distributed nature of cloud computing.[47]
- The lack of access to proprietary details of cloud-based software/ applications used to produce records. "For example, in a particular criminal case involving email through cloud providers," NIST noted, "the details of how drafts are turned into deliverable messages were unavailable, leading to the inability to prove whether or not a draft was ever sent (and more obviously whether it was ever transmitted or received)."[48]
- Limited custodian and record keeper knowledge on what logs and records might constitute evidence.[49]
- International cloud services, and how law enforcement can ensure it is obtaining legal access to data in a way that is not currently clear.[50]
- Lack of standard digital forensic processes and models, including standard procedures and best practices for investigations in the cloud.[51]
- Not knowing where data is stored or who has access to it makes it more difficult to assess whether evidence was leaked or contaminated and thus, whether investigators maintained chain of custody[52]. The credential-based extraction process is a start, but not a panacea. Although UFED Cloud Analyzer prevents other logins while an investigator is using the software, this doesn't control for account activity before or after investigative login.

---

44 Annex B, Item 16.
45 Annex B, Item 17.
46 Annex B, Item 18.
47 Annex B, Item 19.
48 Annex B, Item 36.
49 Annex B, Item 64.
50 Annex B, Items 53-54.
51 Annex B, Item 63.
52 Annex B, Items 23-24.

ISSUES OUTSIDE UFED CLOUD ANALYZER'S SCOPE

UFED Cloud Analyzer's account-based approach renders many issues moot, but others are entirely outside its technological scope.

- Timeline analysis of logs for Dynamic Host Configuration protocol (DHCP) address assignments and other related data[53] is not possible, for example, because logs are typically not part of the provider's API. The protection of system boundaries[54] is difficult to define, and remains an architectural challenge. Likewise the collection of data associated with removed virtual machine (VM) instances.[55]
- UFED Cloud Analyzer is not designed to detect malicious acts or the use of cloud systems as low-cost command-and-control centers such as botnets.[56] Port scanning[57] and Transmission Control Protocol/ Internet Protocol (TCP/IP) network traffic dumping[58] are also beyond UFED Cloud Analyzer's scope.
- UFED Cloud Analyzer is not designed to help isolate an entire virtual machine. Further, because UFED Cloud Analyzer does not act as a VM, there is no risk that malicious software will prevent the isolation and imaging of cloud data.[59]
- In some cases, such as the investigation of a large-scale data breach, the physical location of data may become important. The decreased access and data control, a cloud network's chain of dependencies among cloud providers, and constantly moving data among multiple locations and geographies, including virtual machines, can all affect the evidence available to investigators.[60] Most civil and criminal inquiries, however, rely on the content of evidence rather than its location. UFED Cloud Analyzer users who require deeper insights into data locations always have the option of asking service providers to send records and testify about their authenticity.
- In a similar vein, unless they are engaged in certain types of investigation, most investigators do not need to identify storage media[61] or understand the cloud environment,[62] and UFED Cloud Analyzer is not designed for these purposes. Nor is it designed to help cloud providers proactively address business issues such as ensuring that customers' illicit activities don't impinge on legitimate uses.[63]
- Finally, real-time investigation intelligence processes are currently not possible with UFED Cloud Analyzer, which does not set sensors in the real-time environment as NIST suggests.[64] UFED Cloud Analyzer does collect snapshots of evidence as it exists at any given point in time, but it is not possible for another user to login simultaneously while an extraction is taking place, nor could UFED Cloud Analyzer record streaming media such as a live video

53 Annex B, Item 8.
54 Annex B, Item 22.
55 Annex B, Item 30.
56 Annex B, Items 13-12.
57 Annex B, Item 49.
58 Annex B, Item 50.
60 Annex B, Items 30-25.
61 Annex B, Item 32.
62 Annex B, Item 35.
63 Annex B, Item 57.
64 Annex B, Item 14.

In spite of the many ongoing and evolving challenges associated with cloud forensic extraction and analysis, UFED Cloud Analyzer offers a unique solution to investigators who are frustrated with cloud service providers' changing policies and creeping pace. Its account-based approach reduces many of the logistical and privacy challenges with cloud forensics, while its reliance upon provider API and hashing allows for the necessary authentication and validation of its processes. Finally, its integration with other UFED products empowers investigators to find the data they need when they need it, to collaborate with investigators outside their own agencies, and to build defensible cases more efficiently.

Cellebrite: Delivering Mobile Expertise

Founded in 1999, Cellebrite is a global company known for its technological breakthroughs in the cellular industry with dedicated operations in the United States, Germany, Singapore, and Brazil. A world leader and authority in mobile data technology, Cellebrite established its mobile forensics division in 2007, introducing a new line of products targeted to the law enforcement sector. Using advanced extraction methods and analysis techniques, Cellebrite's Universal Forensic Extraction Device (UFED) is able to extract and analyze data from thousands of mobile devices, including feature phones, smartphones and GPS devices. Cellebrite's UFED is the tool of choice for thousands of forensic specialists in law enforcement, military, intelligence, security, government and private sector organizations in more than 100 countries.

Cellebrite is a wholly-owned subsidiary of the Sun Corporation, a listed Japanese company (6736/JQ).

To learn more, visit
www.cellebrite.com

For more information contact sales