A Survey of Security and Forensic Features In Popular eDiscovery Software Suites

Sundar Krishnan *skrishnan@shsu.edu*
Department of Computer Science
Sam Houston State University
Huntsville, TX, USA

Ashar Neyaz *axn026@shsu.edu*
Department of Computer Science
Sam Houston State University
Huntsville, TX, USA

Narasimha Shashidhar *karpoor@shsu.edu*
Department of Computer Science
Sam Houston State University
Huntsville, TX, USA

Abstract

Litigation these days involves Electronically Stored Information (ESI) for legal purposes. Electronic discovery, also known as eDiscovery, is a process involving legal parties on a case to preserve, collect, review, and exchange electronic information for the purpose of using it as evidence in the case. In the past two decades, the software industry has launched many products catering to eDiscovery. With the advent of cloud computing, storage of electronic data has become cheaper and attractive for eDiscovery needs. With the ever growing technological advances, access to such storage has become simplified for enterprises distributed across the globe. eDiscovery product vendors have embraced the cloud and often allow their products to store and retrieve electronic evidence from the cloud. In this paper, we survey and explore eDiscovery product features focusing on available product security features, security features for evidence protection, incident forensics readiness and cloud forensics. We strive to highlight the challenges in the eDiscovery field when handling vast volumes of electronic evidence and propose incorporating industry best practices in implementing effective security and incident forensics at the product level.

Keywords: Security, eDiscovery, Electronic Discovery Reference Model (EDRM), Electronic Stored Information (ESI), Digital Forensics, Cloud Security, Digital Evidence, Incident Forensics

1. INTRODUCTION

Litigation these days always involves Electronic Stored Information (ESI). Electronic discovery (also known as eDiscovery, e discovery, or eDiscovery) is a process involving legal parties on a case to preserve, collect, review, and exchange electronic information for the purpose of using it as evidence in the case. While most discovery during a legal process still comes in the form of testimony or recorded interrogations, discovery can also involve physical items, like device designs, medical exam results or a defective product. Increasingly, discovery is focused on ESI. Thus, the term eDiscovery is used in the legal industry to distinguish the discovery of electronic data (records) from other forms of discovery. Until two decades ago, eDiscovery was at its infancy with evidence largely in paper from computer printouts and few electronic files from tapes and computer hard-disks. With technological growth and the advent of cloud computing, electronic data is now stored in many formats in varying volumes across many types of devices. One of the reasons for data growth volume is the lower costs of storage media. Cloud service

providers have leveraged the declining costs of storage by offering abundant storage volumes for very low fees thereby making it attractive for the users. With growing technological advances in the internet and telecommunications world, access to such storage has also become simplified for enterprise offices distributed across the globe. The software industry and legal world has taken note of this and launched many products leveraging large storage choices for eDiscovery services. eDiscovery product vendors have embraced the cloud and these-days support mobile platform integrations with their products.

In the digital-legal field, three sets of rules govern the conduct of eDiscovery in U.S. federal court cases: Federal Rules of Civil Procedure (FRCP) [1], Federal Rules of Criminal Procedure (FRCrP) [2] and Federal Rules of Evidence (FRE) [3]. The Federal Rules of Civil Procedure (FRCP) is a set of rules established by the US Supreme Court for resolving civil cases at the federal level. Few of the FRCP rules cover pretrial conferences, duty to disclose, and interrogatories to parties. eDiscovery was not included in the original FRCP because it didn't yet exist; specific provisions related to ESI were only added in 2006 [4]. FRCP amendments were later introduced to protect and preserve data that might be involved in litigation. The Federal Rules of Criminal Procedure are rules set by the US Supreme Court regarding the rights of the individual(s) taking precedence in a criminal case. FRE details the rules applied to evidence that is presented in court for either civil or criminal cases. Rules and Policies Governing Digital Evidence are also found in the Sedona Principles [5]. The Sedona Principles are a set of best practices and guidelines that describe how electronic evidence (ESI) should be addressed relative to eDiscovery.

From the perspective of digital laws, the Computer Fraud and Abuse Act (CFAA) [6] was primarily created to address computer abuse by malicious actors. Before this act, the Counterfeit Access Device and Abuse Act of 1984 targeted fraud and computer crimes U.S. Congress adopted CFAA in 1986 as an amendment to the 1984 act. As part of the National Information Infrastructure Protection Act of 1996, the CFAA was amended to cover extortion that threatens harm to a protected computer. Additional amendments were made in 2001 and 2006 with the addition of the U. S. PATRIOT Act [7]. The Identity Enforcement and Restitution Act [8] of 2008 made an amendment to the CFAA allowing prosecution if the victim and perpetrator are in the same state. In U.S. litigation, in addition to other laws, the above rules and laws are relevant for court cases.

In U.S. courts, legal precedent requires that potentially relevant information must be preserved at the instant a party "reasonably anticipates" litigation or another type of formal dispute [9]. ESI originates from common data repositories of the legal parties or stakeholders. ESI data sources range from computers, email, documents, social media, instant messaging, smartphone applications, databases, web browser data to more obscure ones like automation devices. The rules, processes and technologies around eDiscovery are sometimes complex due to the sheer volume of electronic data produced, stored and securely disposed of (destroyed). Unlike hardcopy evidence, ESI contains metadata such as time-date stamps, author and recipient information, and file properties. Thus, preserving the original content and metadata of ESI is required to eliminate claims of spoliation or tampering with evidence later in the litigation. Proper disposal of ESI is also necessary post case-closure.

Electronic Discovery Reference Model (EDRM) [10] is a conceptual framework of the eDiscovery process that outlines standards for the recovery and discovery of digital data. EDRM consists of nine steps namely; Identification, Preservation, Collection, Processing, Review, Analysis, Production and Presentation for ESI management as described below.

- Identification – Locating potential sources of ESI and scope, breadth and depth identification
- Preservation – Establish safeguards for ESI protection against inappropriate access, alteration or destruction.
- Collection – Gathering ESI for use in the eDiscovery process (processing, review, etc.).

- Processing – Condense the volume of ESI and convert if necessary to forms more suitable for review and analysis.
- Review – Evaluating ESI for relevance to the case.
- Analysis – Assessment of ESI for content and context, including key patterns, topics, people and discussions.
- Production – Delivering ESI to stakeholders or opposing parties in appropriate forms and private delivery mechanisms.
- Presentation – Displaying ESI before audiences (at depositions, hearings, trials, audits etc.), especially in native and near-native forms.

Fig 1 shows the various stages of the ERDM process and the related ESI data locations possibly involved. One might repeat the same step numerous times, or cycle back to earlier steps. As the steps of the model progresses, the security risk of ESI increases over time. Forensic challenges also increase with time in case of an unexpected security incident, intrusion or attack. This shows the increasing risk over time when dealing with ESI data taken from company repositories into the law firm networks during the life of the case under litigation. During the lifecycle of the eDiscovery process incident forensics readiness keeps evolving and security risk increasing over time. The cleanup activity post case-completion has been added as the last stage to the EDRM process to deter malicious forensic acquisition and extraction of this data from the eDiscovery storage locations. This stage can also benefit from physically destroying any storage media or hard drives using a high-security and verifiable media destroyer. Residual data on storage devices cannot be retrieved even when using modern forensic techniques when adequate data destruction measures are employed. Proper data destruction is also key to regulatory compliance and hefty fines.

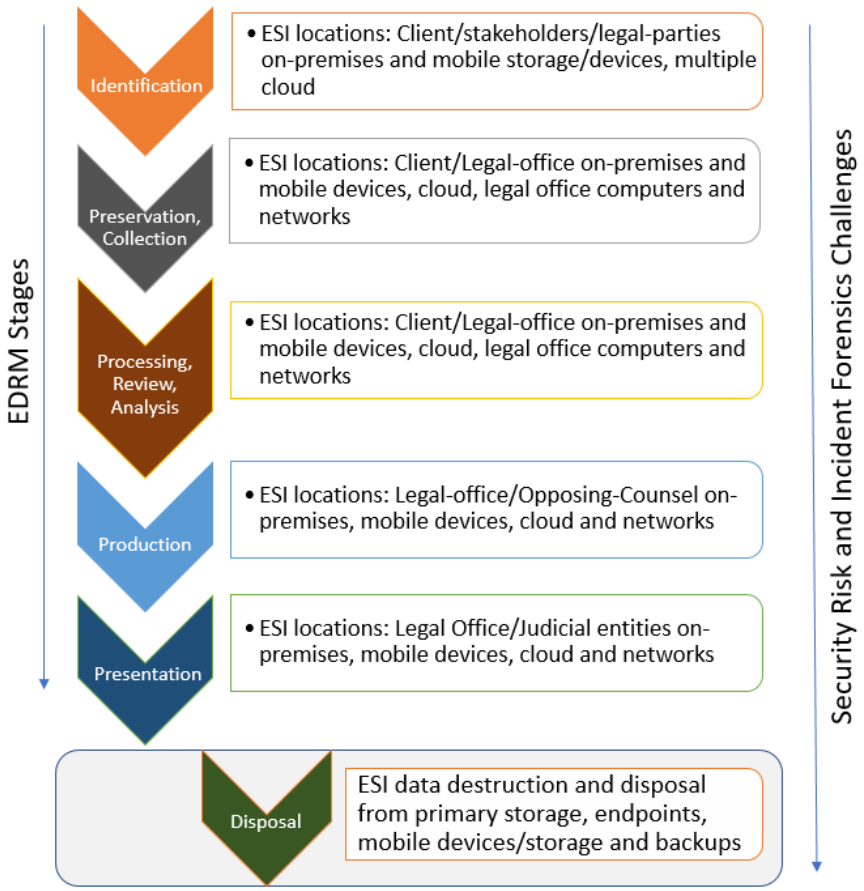

FIGURE 1: EDRM - Security Risk and Forensic Challenges.

The relationship between eDiscovery and digital forensics can sometimes be confusing. eDiscovery is the procedure by which parties involved in a litigation case collect, preserve, process, review and exchange information in an electronic format to use it as an evidence in the case. True digital forensics on the other is used to perform a specific deeper recovery of say a computer disk looking for hidden data or unallocated disk space for identifying who, what, where, why from a computer. The overall processes involved in both eDiscovery and digital forensics are very similar. Both involve the identification, preservation, collecting, analyzing and reporting of data with minimal tampering. However, eDiscovery involves dealing with 'active' and visible data using superficial forensic identification techniques of searches, queries etc. applied through file repositories and application programs installed on the machine whereas digital forensics can more technically complex as it involves digging a little deeper and looking at hidden areas of the system, logs and deleted files for example.

Incident forensics is part of the incident response and management process triggered by a security event (incident). The primary goal of the incident response is to validate, isolate and contain the incident. While incident forensics largely follow the digital forensics procedures and tools, the difference is largely in the setting and goals. Incident forensics is to identify the attack vectors, actors, etc. that triggered the incident. Both incident forensics and digital forensics require strong log analysis, disk acquisitions, file carvings, network packet analysis, and malware analysis capabilities. Cloud forensics is a cross-discipline of cloud computing and digital forensics. It can also be considered as a subset of network forensics. Pichan et al. [11] describe the issues in cloud computing using the phases of traditional digital forensics thereby helping the forensic incident investigators to better understand the problems in a cloud environment. These issues manifest in the eDiscovery industry due to the dependency of eDiscovery products on the cloud environment. In this paper, we survey and explore eDiscovery product features focusing on overall security, digital forensics, cloud forensics, and evidence security. We strive to highlight the pitfalls and challenges in the eDiscovery field when accessing and handling vast volumes of ESI in a forensically secure way. In addition, we also hope that the presented results will stimulate further research in securing the eDiscovery process and ESI.

2. RELATED WORK

E-discovery is a billion-dollar industry and is growing continuously but is facing new challenges each year such as expensive outsourcing, evolving technology, large amounts of ESI data per case and the continuous learning curve for legal staff on eDiscovery technology to name a few. There are numerous challenges in this industry [12] highlighting the issues of disappearing and reappearing redactions by eDiscovery product users (legal staff). Performing this work manually is laborious and it is considered a nightmare. Document files can be thousands of pages long and manually performing the redaction process is not feasible. Further, documents without Optical Character Recognition (OCR) is again an issue as many in the legal world refuse to OCR their documents that have not been already OCR-ed.

Inconsistent coding is another challenge as interpretation is difficult to understand. The need for searchable database with OCR makes the task efficient for all the attorneys and also helps in preventing sanctions by courts. In a survey [13] that was carried out against thirty judges few key trends and insights surfaced such as; attorneys being slow to catch up with the eDiscovery competence, attorneys having an attacking mindset to tackle the problems of eDiscovery and that eDiscovery education was required and training should be made available. Attending adequate training(s) and educational webinars periodically can help in solving the aforementioned issues. The survey also noted that a small percentage of the legal staff were confident enough to counsel their clients on eDiscovery matters. While eDiscovery is blooming on its own and has its own challenges and merits, industry trends indicate that companies moved from the left-side of EDRM (Identification, preservation, and collection) phases to in-house first [14]. Industry trends also indicate that companies are preserving more from sources than ever before [3]. Cheap cloud and on-premises storage options can be attributed to the large volume of data being preserved. Growth in technology around storage, preservation, project management and analytics are

empowering in-house teams to tackle even more challenges [14]. The report [14] also confirms that Law firms are striving hard to incorporate technology to better collaborate with clients and service providers and in the past two years, 71% of the companies have successfully done so. Companies are using data repositories and document preservation so that they can process the litigation data with their eDiscovery products. There is one more issue and that is the issue of spoliation of ESI. Many companies rely on data custodians for preserving data [14]. eDiscovery products companies are also providing Artificial Intelligence (AI) technologies in their products so that legal teams have better efficiency when analyzing a case. EDiscovery professionals are also facing managerial challenges such as controlling costs, completing tasks efficiency, ensuring that the process is defensible [14].

Security is a major challenge impacting the growth of cloud services. Trying to adequately identify secure, acquire, examine and produce case-related data, let alone determine legal relevance, is becoming an increasingly daunting task [15]. Growing security concerns with cloud data storage have prompted increased attention leading to organizations developing independent cloud-focused security policies. Similar concerns also linger around third-party data from corporate sources in a legal network during eDiscovery. Forensic work in eDiscovery includes incident investigations during or after a security breach attempting to establish the key six questions of an incident; who, when, what, why, where and how [16]. While Digital Forensic Readiness can be considered as a proactive measure, there is a limited implementation as a policy across the legal industry. Park et al. [17] suggest an adoption of Digital Forensic Readiness as a mandatory requirement to protect personal information and efficiently implement information risk management for private and government entities. Complications with data privacy, data protection, data disposal continue to plague the cloud market. Subashini et al. [18] highlight different security issues that have evolved due to the nature of the service delivery models offered by cloud service providers. Similar risks are also prevalent in an organization should ESI data be stored on-premises of the legal offices. To overcome security challenges, risk assessments should be periodically undertaken. Security should be implemented in layers focusing on the three tenets of the security triad; Confidentiality, Integrity and Availability (CIA). Krishnan et al. [19] discuss the legal user-privacy and data-privacy challenges when working with Cloud that can be easily related to eDiscovery security scenarios. Chang et al. [20] discuss a multi-layered security approach to protect data in real-time in the cloud. However, there is minimal research addressing eDiscovery product features related to security and forensic readiness. Securing storage and access to ESI and being forensically prepared for a possible cyber incident are key factors to reduce risk and stay compliant with regulations and privacy laws. This paper tries to fill this gap by conducting vendor surveys and discusses the gaps around security and forensic readiness of eDiscovery products.

3. EDISCOVERY PRODUCT INDUSTRY, SECURITY AND FORENSICS

3.1 The eDiscovery Product Industry

A product of eDiscovery process is the ESI. The history of ESI can be documented as history: years 2000 to 2004, The Wild, Wild West; years 2004 to 2008, Standardization and Stabilization; year 2008, Depression; years 2009 to 2011, New Tools, Rules and Schools; years 2011 to 2013, Massive Maturation; and finally years 2013 to 2016, Consolidation [21]. With the development of ESI over the years, the eDiscovery industry has also matured and grown with major players like Exterro, Logikcull, Sherpa Software, CaseFleet, CloudNine, iCONECT, Relativity, etc.

3.2 Security During eDiscovery

Cybersecurity around eDisovery is still in its nascent phases requiring pioneers, better design, standardization, and more form to give it greater function [21]. With Cloud adoption as a strong argument for cost savings, corporate (client) data in litigation cases are often stored in the cloud. As cloud security is yet to be fully understood and properly implemented by industry, eDiscovery users should deploy adequate security layers to protect sensitive client ESI data until litigation is completed. Adequate ESI data destruction process and documentation by cloud providers should be established in service level agreements. Most eDiscovery product vendors are yet to

incorporate strict security controls around their product architecture and against ESI. Incorporating security controls in the products should not be dictated customers but by the industry security best practices.

3.3 Forensics In eDiscovery

During eDiscovery, ESI collected in the initial phases of the eDiscovery process arrive from different sources of devices and storage. During the course of the legal case, ESI can sometimes move across storage locations like from one cloud provider to another. Some of the ESI files are obtained via technical forensic techniques like ESI from computer forensics of disk drives or form chip-level forensic extraction methods of a smartphone. Some ESI files are collected from lesser technically intensive forensic processes like email searches, repository searches, etc. Gathering of relevant email(s) for the legal case ESI is by conducting searches of the mail exchange system or archives that in turn can also be argued to be similar to conducting forensics of the mail system.

3.3.1 Forensics During Unforeseen Cyber Incidents on ESI

Incident Forensics tends to be the human effort required in the case of a cyber/computer breach of the ESI during the eDiscovery process. Malicious actors are often attracted to the ESI repository as it could contain Personally Identifiable Information (PII), Protected Health Information (PHI), Confidential Business Information (CBI), etc. of clients and their customers. Incident investigations will require forensics of the eDiscovery product, legal office networks, users, ESI storage and associated systems to identify and close any gaps in security.

3.3.2 ESI and Metadata

Increasingly digital forensics contributes to the bulk of ESI for a legal case. ESI is far more than just communications sent between parties; a party's investigators can also create custom ESI [22]. In a lawsuit brought by PETA against at a zoo for the treatment of animals under the Endangered Species Act, the Defendants sought photos and videos surreptitiously recorded, and related investigatory reports [23]. In an ideal situation, ESI metadata should not change during the course of the legal case. However, uncontrolled or accidental access to ESI can also sometimes alter the metadata. Movement of ESI storage locations when a case is taken over by new law firms can also alter metadata.

4. PRODUCT SURVEY METHODOLOGY

In this study, the survey method used was a questionnaire prescribed to select eDiscovery product vendors. We administered this survey from Dec-2018 to Jan-2019. A total of 8 eDiscovery product vendors were contacted for the survey with 5 vendors completing the survey. eDiscovery vendor participation was voluntary and the product vendors were shortlisted based on the product vendors in the leaders and high performers quadrants [24]. Few product vendors contacted us for survey question clarifications before submitting their replies. Submitted questionnaires were consolidated for analysis. Scoring was based on an aggregate total of similar replies per question. Unanswered questions were skipped during analysis.

The survey questionnaire was developed based on the security and forensics features of eDiscovery products. We focused on simple yet basic features that should be part of a product given the sensitive information it would store, index and handle. The representation of the questionnaire was mostly around security and incident forensics. Most survey questions were designed as close-ended for easy measurement. The complexity of questions was limited and categorization was added for context.

The survey was circulated to nine vendor participants. Completed surveys were received from five vendor participants. Skipped questions on the survey were ignored from the final report with adequate notes. While individual answers to the survey questions are confidential, the aggregate results are published with commentary to highlight the existing security and incident forensics posture among these eDiscovery products.

5. RESULTS

In this section, we analyze and opine on the results from the eDiscovery vendor survey. The questionnaire was broken into sections; Storage, Search functionality, Redaction and logging, Access control (Authorization, Authentication and Accounting), File management, monitoring and metadata, Case Processing and workflows. Below is the analysis of each survey section with the questions and responses as on the survey questionnaire.

5.1 ESI Storage

The eDiscovery product vendors offer both on-premises and Cloud type storage. Storage can be for two different purposes; ESI storage and application functionality storage. ESI storage is for electronic data related to the case obtained from corporate (client, stakeholder) storage locations. Application usage generates data like custom user notes, logs etc. that also needs adequate storage and should be independent of ESI unless required. The storage location(s), types, access lists, backup volumes and frequency are few key factors for security design and during incident forensics. Survey questions related to storage are in Table 1.

Survey questions	Response
What are the storage size options offered for evidence (ESI)?	Elastic depending on customer needs
What are the location options for evidence storage offered (on-premises/cloud?)	3 out of 5 respondents offer cloud. Remaining 2 respondents solely offer on-Premises
Cloud storage; can customer pick the cloud provider or is it predetermined?	Mostly Pre-determined
If cloud provider is recommended by product vendor, what is the recommended cloud provider?	Mostly AWS
For on-premises storage, are there any storage perquisites like (RAID levels)?	None
Is the core product application intranet? (client/server or intranet web?)	Mostly Web/Internet based
Is there an application specific database? If so, can it be on-premises or in the cloud?	Yes applications have databases. 3 out of 5 respondents with cloud offering have databases in cloud, remaining 2 respondents offering on-Premises host databases on-Premises
Is there an application specific database? If so, what is the database product used (Oracle/SQLite/SQL Server etc.)	Yes, ranging from PostgressSQL, Access to SQL Server
Is storage for evidence data ESI encrypted by default?	Yes for cloud. Customer responsible for on-Premises

TABLE 1: Evidence Storage by eDiscovery Products.

From Table 1, it can be observed that very few vendors offered on-Premises storage options for the end customer. This highlights the dependency of cloud as a cheaper alternative. The vendors also seem to prescribe their preferred cloud providers for the customer making this is a potential limitation for the customer. If not properly planned, designed and managed, cloud integrations could bring about complications with security and cyber incidents. The details on the cloud service contracts can also limit attack investigations and incident forensics. The databases offered are wide-ranging which can be a drawback for the customer during future migrations to different eDiscovery products. The use of encryption-at-rest feature for storage has to be strong and keys adequately managed.

Security around storage is crucial especially when storing sensitive ESI for a long period of time. Storage size also determines the forensic processing time. Data Leak Prevention (DLP) techniques like strong encryption at rest, encryption during transmission and when in use is recommended on ESI storage. Adequate vulnerability and risk assessments should be carried out focusing on ESI storage. Product vendors may not undertake the forensic challenge of wiping ESI data from cloud servers or mobile devices upon a case closure. Such considerations should be taken into account by eDiscovery product customers when business agreements are signed with cloud providers. Establishment of a corporate Information Technology policy for clients and stakeholders geared towards cloud storage and handling of ESI at a third party location is also recommended. Routine backups of ESI and Disaster Recovery (DR) have to be planned keeping security risk in mind. Regular DR exercises should be planned irrespective of storage being in the cloud or on-premises.

5.2 Search Functionality

During the eDiscovery lifecycle [25] following hold or preservation, identification, and collection, it is necessary to normalize and process data before it can be moved to the next steps of the EDRM Lifecycle. The term "processing" encompasses many steps like De-NIST, De-duplication, Embedded Objects, Exceptions, Password Protected Files, Time Zone, etc. [26]. Indexing is another step in the "processing" stage and can take time to complete, depending on the volume of data that needs to be indexed. Indexing of ESI greatly helps with the efficiency and performance of the search functionality. Each time there is a change in the ESI repository like the addition or removal of a file, re-indexing is required. Processing almost always includes the forensically sound search and extraction of the necessary file, metadata and text for the subsequent review stage. Survey questions related to search functionality are in Table 2.

Survey Questions	Response
What are the Search types supported (keyword, wildcards, RegEx, strings, logical, contextual)?	Keyword, date, fuzzy, stemming, conceptual, near duplicate, wildcards, Full Lucene, RegEx etc.
Are language translations offered during searches? (can a keyword search in English also pick a French equivalent?)	Not currently
Are language translations offered alongside contextual searches? [can a keyword search for "Butter" in English also pick "Lard" or "Fat" in French?]	Not currently
What are the search target formats supported? [.pst, pdf, .doc etc.]	Various. Files without text are OCRed.
Can the search be targeted in scope and made granular with filters?	Yes
Can a search be scheduled?	Not currently
Are there schedule management features available for long running searches?	Not currently

TABLE 2: Search Functionality.

From Table 2, it is evident that files that are not-text based are OCRed. eDiscovery products seem to have excellent features for indexing, search types and search filtering. Most products incorporate search against the metadata of the ESI files. Some of the search engines used in eDiscovery products are Lucene, SQL Server and Elasticsearch, etc. An existing drawback is the lack of search management features to manage and schedule a long running search. Another drawback is the unavailability of decrypting a file before adding to the ESI repository thereby inhibiting a search within these files. It is, however, unclear if these products can handle multiple simultaneous search queries against the same ESI dataset triggered by multiple data custodians.

Keywords used for searches is recommended to be tracked by user. Malicious insiders and external actors who breach security may execute searches that could help incident investigators if adequately logged. Alternatively cloud service providers may be leveraged to assist with reporting on searches made against storage. Access controls should be factored when searching PHI, CBI, PII data. If leveraging cloud offerings such as search-as-a-service for product's search functionality development, compliance to industry standards like ISO 27001, SOC 2, HIPAA, PCI etc. can be sought from cloud service providers. Else, following compliance to such standards is recommended in-house. Audits of all searches performed by user should be factored into regular security risk audits. Limitations on search functionality against non-searchable ESI and the indexes used for supporting searches should also be documented to assist with incident investigations. Searches against encrypted data may need some level of dynamic decryption using crypto keys extracted from vaults requiring such design undertaken with security risk in mind.

5.3 Redaction and Logging

With the growth of ESI from the corporate community over the years, there has been a corresponding increase in focus on how the data that has been collected, processed and reviewed is ultimately produced in civil litigation and regulatory investigations. The FRCP Privacy Rule 5.2 [27] outlines the general rules of redaction. Redaction is sometimes known as the necessary evil during eDiscovery. Redactions obscure confidential or privileged information on files or their metadata before production. Redactions are typically added by the review team prior to production. If stamps or redactions are required, the native and near-native files are converted to image (near paper) formats so stamps and redactions can be applied [28]. Redaction if not properly managed can lead to security challenges, regulatory fines, and privacy risks. While there is no single best method to limit privacy risk exposure, charting out a privacy map of ESI during the collection phase of EDRM may be recommended as best practice. Redacted data should be encrypted, vaulted and securely stored. Survey questions related to redaction and logging are in Table 3.

Survey questions	Response
Are Redaction features offered?	Mostly Yes
Is there a Redaction management feature to track and manage redaction activity by various users over time?	Mostly No
Does the product use any AI or machine learning?	Mostly No
Does AI or machine learning features/capabilities come at an additional cost to the customer?	3 respondents replied with N/A. When available, AI is leveraged by integration of 3rd party tools or depends on the vendor contract.
Does the product have alerts/notifications like visual flags, emails etc.?	Mostly Yes
Does the product allow users to edit an ESI evidence file?	Mostly No
Does the product track history of changes, features to rollback/ override if available?	User activity is tracked if and when allowed.
Is there a feature within the product to rollback/override changes with elevated permissions?	Mostly No

TABLE 3: Redaction and Logging.

In Table 3, redaction readiness of the eDiscovery products is surveyed. Many products from the survey do not seem to have built-in redaction features but allow for integration with third-party redaction specific products. Logging user access activity seems sparse and has to be improved

for Identity and Access audits and continuous security monitoring. Remote access of ESI data between legal partners during eDiscovery should also be securely managed. When producing documents with redactions on the load file, legal teams should maintain a checklist that ensures that image redactions are burned-in, that redacted native files (if produced natively) are properly redacted, and that associated text files and metadata have been checked to ensure that redacted data has been removed from those as well. Else, the production of redacted materials can still contain sensitive information that should not be shared with the opposing legal team. Tracking and logging redactions is recommended for security risk audits and swift incident forensics.

5.4 Access Control (Authorization, Authentication and Accounting)

The legal team, company staff, and third-party eDiscovery service providers may collect and transfer large amounts of Personally Identifiable Information (PII), Protected Health Information (PHI), and Confidential Business Information (CBI), etc. for production in legal matters. The need for access must balance user demands against difficult security requirements. With social engineering growing into an increasingly common and effective attack vector, remote access to case files and case ESI is a security challenge. Survey questions related to access control are in Table 4.

Survey questions	Response
Does user authentication support 2FA (2-factor authentication)?	2 respondents support 2FA
How is user authentication managed on the product?	User ID/Passwords, SAML based SSO, cookie/session based and in few cases MFA/2FA.
Can user access and authentication be managed and audited from within the product?	Mostly Yes
Can product logs be sent to a SIEM for monitoring? (SIEM Integration)	Mostly No
How is the first-time user login details shared?	Self-enrollment, via eMail or following company policies.
Is evidence storage vaulted with password rotation and/or 2FA?	Mostly Yes
Is evidence storage accessible from workstations via Single Sign-On (SSO)?	Mostly No
Is evidence storage accessible from non-workstations like smartphones, tablets etc.	2 respondents replied with “Yes”, 2 respondents replied with “No” and the respondent uses a client-server architecture.

TABLE 4: Access Control (Authorization, Authentication and Accounting).

A password required to access ESI is inadequate given the ease at which they can be cracked. Access from mobile platforms like smartphones, laptops, and tablets should be layered with additional access protection as they can use insecure or public Wi-Fi when accessing ESI. From Table 4, two-factor or multi-factor authentication (2FA or MFA) support for access to the eDiscovery system and ESI is recommended. While two-factor authentication is offered when accessing ESI storage, it should be strictly enforced with a short token expiry time and not left optional (bypass) for any user. Single Sign-On (SSO) with third-part (client) networks is also recommended after reviewing the integration design for security risk. Security Information and Event Management (SIEM) integration seems to be weak and needs attention. As SIEM integration requires log ingestion from the eDiscovery products, various types of logs are required to be produced especially around access control, change management, metadata changes, and redactions. While product’s mobile integration is desired, adequate security around the device like

data encryption and jailbreak/root check functionality, remote wiping should be incorporated. DLP techniques should be implemented irrespective of the endpoint.

5.5 File Management, Monitoring and Metadata

As mentioned earlier, electronic data can be in a wide variety of formats like; Social media posts and communications, Email communications, Website content, Browser data, Instant messages, and text messages, Voicemail messages, Team-wide workflow applications, Company databases, IoT Data, Microsoft Office files, etc. Safeguarding ESI is a challenge these days as they can sometimes in terabytes. Metadata extraction enables the legal teams to see, search, and sort specific data about a particular file. File hashing helps to locate duplicated files. This helps the legal team to avoid searching for duplicate files. Survey questions related to file management, monitoring and metadata are in Table 5.

Survey questions	Response
Does the product automatically extract metadata against evidence files when a new file is added to the evidence storage?	3 respondents replied Yes
Does the product automatically detect and remove metadata against evidence files when a file is removed from evidence storage?	2 respondents replied with “Yes”, 2 respondents replied with “No”
Does the product have any inbuilt malware/virus detection scans or can it leverage a 3rd party service for automated scans?	No
Does the product allow 3rd party external scans against evidence storage for malware/virus detections?	Mostly Yes
Does the product integrate with 3rd party native viewing tools or offer any inbuilt features for native file format views?	Mostly Yes
What are the reporting file formats allowed by the product?	PDF, CSV, Excel, XML and others
Does the product assist with decryption if necessary keys are supplied?	Mostly No
If Yes, how is the key managed, stored and secured?	Mostly No
Does the product allow custom work notes/comments by users?	Yes
Are case work notes/comments stored alongside evidence?	Yes
Do work notes/comments get added to the evidence storage automatically based on user roles? (expert’s notes is evidence)	3 of 5 respondents replied “No”

TABLE 5: File Management, Monitoring and Metadata.

From Table 5, it is evident that metadata is a focus area in the eDiscovery products. By integrating with a virus scanner, ESI storage can be periodically scanned thus avoiding access to malicious files. Sometimes, scanning files can alter metadata and therefore, care should be taken to configure the virus scans to be non-intrusive. During decryption of client’s ESI data is performed using the eDiscovery product, adequate crypto key management techniques should be factored to vault client’s keys (to decrypt evidence) and eDiscovery user’s organization keys. Application data like custom notes, memos, tasks, contacts, etc. should be tracked and stored separately to avoid ESI contamination unless part of the evidence. Most vendors maintain such application data within a relational database. Encryption of this data on the database is also recommended. Tracking of ESI storage changes is also essential for audit purposes.

5.6 Case Processing and Workflows

Timeline visualizations can be useful to know when ESI data in a case assignment or file were sent, received, or last modified. Visualizations can also help with a chronology of case facts, ESI file dependency, and redaction impact. Redundant, obsolete, or trivial (ROT) management involves the culling of ESI files that are derailing the efficiency and cost of the eDiscovery process. Many companies still exist "in a state of ROT" and need good data hygiene [29]. Survey questions related to case processing and workflows are in Table 6.

Survey questions	Response
Does the product have timeline graphs features (graphical visualization) against case evidence?	3 of 5 respondents replied "Yes"
Do the timeline graphs allow for user customization?	3 of 4 respondents replied "No"
How is Redundant, obsolete, or trivial (ROT) management/culling managed and tracked?	Various methods allowed
Does the product offer any regulatory/compliance integrations (smart/intelligent lookups) against FERC, Sedona etc.?	Mostly No
Does the product integrate with live/online text messaging, social website extractions, online repositories for evidence collection?	Mostly No. Few allow Office 365 and Exchange integration
Does the product integrate directly with on-premises Exchange Server, RAID storage, external storage media etc.?	Mostly No. Few allow Office 365 and Exchange integration
Does the product integrate with judicial sites/applications for automated case updates?	No
What is the electronic transmission mechanisms supported (web, FTP etc.)?	One respondent replied HTTPS, another respondent replied as a client choice, the others do not support automatic electronic transmissions.
Is the electronic transmission secure with https or sFTP or SSL VPN?	One respondent replied SSL, another respondent replied as a client-choice, the others do not support automatic electronic transmissions.

TABLE 6: Case Processing and Workflows.

From Table 6, we can safely conclude that all eDiscovery tools in the survey incorporated ROT features. Timeline visualizations greatly assist forensic investigators in large cases and can also highlight key ESI files changed or removed. Timeline graphs can also help validate data restoration points from backups during incident recovery. Intelligent lookups to external sources of reference is missing and can be of help during reviews. Integrations to dynamically extract data from websites, online forums, smartphones, social sites, third-party cloud databases, etc. can help with easy ingestion of data relevant for the legal case. Submission of final load files (post-production stage) in encrypted formats via secure transmission mechanisms is highly advocated.

6. CONCLUSION

With the increasing use of cloud services and mobile platforms, challenges around security, privacy, and forensics readiness continue to be instrumental in shaping the security risks of the organization. From this survey, we can safely conclude that eDiscovery products need to further embrace security features and incorporate forensic readiness capabilities to better assist in

incident investigations. Security posture and incident forensic readiness of eDiscovery products will need to stay abreast with the industry's security best practices. Employing better security design can greatly facilitate rapid incident containment, lower the attack surface, improve incident forensics and fulfill business continuity goals. As eDiscovery products often deal with the client's (legal party) data (evidence) coupled with increasing cloud usage, enhanced identity and accessibility security features should be supported by these products. eDiscovery product vendors also need to further incorporate security safeguards features to protect ESI during storage and transmission, secure redacted data, increase system and user logging, create crypto key vaults, provide audit features at the product level and deploy multi-factor authentication. While DLP controls at all levels are encouraged (rest, use and transport), further use of an Electronic Document Management System (EDMS) design approach is recommended to efficiently retrieve and organize large volumes of ESI data. Improved redundant, obsolete, or trivial (ROT) ESI management is critical to secure storage options and their policies. Digitally signed and encryption (or password-enabled) of reports from the product's case processing is also recommended. Storing copies of all case reports generated by these products is also suggested to facilitate audits. We also conclude that the implementation of artificial intelligence coupled with user behavior analytics (UBA) is promising towards decreasing processing time, orchestrating collection and security monitoring. Use of AI through machine learning, predictive coding, predictive analytics, etc. can also greatly assist with incident forensics and security audits of ESI. While we surveyed only a subset of the industry vendors offering eDiscovery products, and a broader study is proposed across the product industry focusing on their security integrations with the larger legal organization IT setup. Similarly, a study of their use of AI towards improving ESI security is also recommended.

7. ACKNOWLEDGMENT

We would like to thank the Cyber Forensics Intelligence Center at Sam Houston State University. The authors would also like to thank eDiscovery product vendors Zapproved, CaseFleet, Sherpa Software, iCONECT and CloudNine for their time and enthusiasm in completing the surveys and sharing their experience in the eDiscovery industry.

8. REFERENCES

[1] Committe on the Judiciary. (1994). "Federal Rules of Civil Procedure 1993," vol. 39, no. 6, Internet: https://www.uscourts.gov/sites/default/files/civil-rules-procedure-dec2017_0.pdf [Jan 08, 2019]

[2] The U.S. House of Representatives. (2014). "Federal Rules of Criminal Procedure". Internet: http://www.kyed.uscourts.gov/kyed_GOs/Privacy_Rules.pdf [Jan 20, 2019]

[3] U. S. Government, "Federal Rules of Evidence," *Legal Ref. Serv. Q.*, vol. 8, no. 3–4, pp. 219–232, 1988 Internet: https://www.uscourts.gov/sites/default/files/Rules of Evidence [Jan 18, 2019].

[4] Zapproved. (2018). "What is the FRCP? | Zapproved," *Zapproved*, Internet: https://www.zapproved.com/blog/what-is-the-frcp-facts-about-the-federal-rules-of-civil-procedure/ [Jan 11, 2019].

[5] The Sedona Conference. (2007). "The Sedona Principles, Second Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production (The Sedona Conference Working Group Series, 2007)" Internet: https://sosmt.gov/wp-content/uploads/attachments/A - Sedona Principles Second Edition.pdf?dt=1485383215701&dt=1519325634100 [Jan 04, 2019].

[6] H. Marshall Jarrett and E. W. Michael Bailie. (2015). "Prosecuting Computer Crimes," vol. 741, Available: https://www.justice.gov/sites/default/files/criminal-ccips/legacy/2015/01/14/ccmanual.pdf [Jan 20, 2019].

[7] U. S. Government. (2001). “The USA PATRIOT Act : Preserving Life and Liberty,” *USA Dep. Justice*. Internet: http://www.justice.gov/archive/ll/what_is_the_patriot_act.pdf [Jan 17, 2019]

[8] The Senate and House of Representatives of the United States of America. (2008). “One Hundred Sixth Congress of the United States of America TITLE II—IDENTITY THEFT ENFORCEMENT AND RESTITUTION ACT”. no. c, pp. 1–13. Internet: https://www.govinfo.gov/content/pkg/BILLS-110hr5938enr/pdf/BILLS-110hr5938enr.pdf [Jan 04, 2019] .

[9] Logikcull, “Chapter 2 - Data Preservation and Legal Holds| The Ultimate Guide to eDiscovery”. Internet: https://www.logikcull.com/guide/chapter-2-data-preservation-and-legal-holds. [Jan 11, 2019].

[10] D. Law, “EDRM Model,” *Duke Law*. Internet: https://www.edrm.net/frameworks-and-standards/edrm-model/. [Jan 11, 2019].

[11] A. Pichan, M. Lazarescu, and S. T. Soh. (2015, Jun.). “Cloud forensics: Technical challenges, solutions and comparative analysis,” *Digit. Investig.*, vol. 13, pp. 38–57. Available: https://www.sciencedirect.com/science/article/pii/S1742287615000407 [Jan 07, 2019].

[12] Exterro. 2017. “E-Discovery Nightmares” Internet: https://www.exterro.com/blog/e-discovery-nightmares-5-experts-tales-of-terror/ [Jan 05, 2019].

[13] B. Y. Exterro and C. E. O. Bobby, “4 th Annual Federal Judges Survey Judicial Perspectives on the State of E-Discovery Law and Practice.” Internet: https://www.exterro.com/about/news-events/fourth-annual-federal-judges-survey-shows-biggest-problems-in-e-discovery-are-caused-by-mindset-of-practitioners/ [Jan 20, 2019]

[14] Exterro. (2018). “The State of eDiscovery 2018” Internet: https://www.exterro.com/about/news-events/state-of-e-discovery-report/ [Jan 06, 2019]

[15] S. Jorgensen. (2003). Convergence of Forensics, Ediscovery, Security, & Law. [On-line]. vol. 12, no. 2. Ave Maria School of Law. Internet: https://www.questia.com/library/journal/1G1-383980285/convergence-of-forensics-ediscovery-security [Jan 10, 2019]

[16] F. Freiling and B. Schwittay. (2007). “A Common Process Model for Incident Response and Computer Forensics,” in IMF 2007 : IT-Incident Management and IT-Forensics; proceedings of the 3rd International Conference on IT-Incident Management \&IT-Forensics; September 11 -13. Stuttgart, Germany. [On-line]. pp. 19--40. Available: http://ub-madoc.bib.uni-mannheim.de/23053/ [Jan 14, 2019]

[17] S. Park, N. Akatyev, Y. Jang, J. Hwang, D. Kim, W. Yu, H. Shin, C. Han, J. Kim (2018, Mar.). “A comparative study on data protection legislations and government standards to implement Digital Forensic Readiness as mandatory requirement,” Digit. Investig. [On-line]. vol. 24, pp. S93–S100. Available: https://www.sciencedirect.com/science/article/pii/S1742287618300446 [Jan 20, 2019]

[18] S. Subashini and V. Kavitha. (2011, Jan.). “A survey on security issues in service delivery models of cloud computing,” J. Netw. Comput. Appl. [On-line]. vol. 34, no. 1, pp. 1–11. Available: https://www.sciencedirect.com/science/article/pii/S1084804510001281 [Jan 14, 2019]

[19] S. Krishnan and L. Chen. (2014). “Legal Concerns and Challenges in Cloud Computing”. Internet: http://arxiv.org/abs/1905.10868 [Jun 27 2019].

[20] V. Chang and M. Ramachandran. (2016, Jan.). “Towards Achieving Data Security with the

Cloud Computing Adoption Framework," IEEE Trans. Serv. Comput. [On-line]. vol. 9, no. 1, pp. 138–151. Available: http://ieeexplore.ieee.org/document/7299312/ [Jan 22, 2019]

[21] Jared Coseglia. (2017). "State of the Industry: E-Discovery and Cybersecurity | Law Journal Newsletters". [On-line]. Available: http://www.lawjournalnewsletters.com/sites/lawjournalnewsletters/2017/06/01/state-of-the-industry-e-discovery-and-cybersecurity-3/?slreturn=20190005083511. [Jan 05, 2019].

[22] J. Gilliland. (2018). "Designating Privileged and Work Product Materials – CaseFleet". Internet: https://www.casefleet.com/blog/choosing-whats-privileged-or-work-product. [Jan 11, 2019].

[23] "People for the Ethical Treatment of Animals, Inc. v. Tri-State Zoologica...tern Maryland, Inc. et al Doc. 102". (1964). Internet: https://cases.justia.com/federal/district-courts/maryland/mddce/1:2017cv02148/396265/102/0.pdf?ts=1541239401 [Jan 12, 2019]

[24] G2Crowd. "Best eDiscovery Software in 2019 | G2 Crowd". Internet: https://www.g2crowd.com/categories/ediscovery. [Jan 06, 2019].

[25] D. EDRM. "Processing Guide." Internet: http://www.edrm.net/frameworks-and-standards/edrm-model/processing/ [Jan 4, 2019]

[26] Amy Bowser-Rollins. (2017). "Electronic Discovery – Indexing". Litigation Support Guru. Internet: https://litigationsupportguru.com/electronic-discovery-indexing. [Jan 08, 2019].

[27] "Federal Rules of Civil Procedure," Federal Rules of Civil Procedure, Administrative office of the Federal court System. Internet: http://www.kyed.uscourts.gov/kyed_GOs/Privacy_Rules.pdf. [Aug 8, 2019].

[28] D. Law. (2010). "Production Guide | EDRM". EDRM Duke Law. Internet: http://www.edrm.net/frameworks-and-standards/edrm-model/production/. [Jan 08, 2019].

[29] S. Williams and A. Gurney. (2016). "The 10 Steps to High Performance eDiscovery" Capax Discovery LLC. Internet: https://cdn2.hubspot.net/hubfs/1946272/Datasheets/10_Steps_for_High_Performance_eDiscovery.pdf [Jan 20, 2019].