COMPUTER FORENSICS LAB

London digital forensics & e-discovery · Established 2007

What Is a Forensic Hash and Why Does It Matter?

MD5, SHA and Integrity, in Language a Court Will Follow

A hash value is a mathematical fingerprint: the same data always produces the same hash, and changing a single bit produces a completely different one. It is how a copy proves it is a true copy, and how an exhibit proves it has not changed since acquisition. It is also one of the most misunderstood ideas in litigation: matched hashes are cited as proof of authenticity (they are not), and "MD5 is broken" is deployed as if it dissolved the evidence (it does not). This guide explains the MD5 and SHA families, what hashes genuinely prove and genuinely cannot, and gives lawyers the exact sentences to use, and to challenge, in court.

⏱ Estimated reading time: 27 min read

§ ABOUT THE AUTHOR

PREPARED BY COMPUTER FORENSICS LAB E-DISCOVERY TEAM

Computer Forensics Lab is a London-based digital forensics and electronic disclosure practice established in 2007. Hashing runs through everything its examiners do: computed at every acquisition, re-verified at every transfer, exhibited in every CPR Part 35 and CrimPR Part 19 report, and explained, regularly, to judges and juries in exactly the plain terms this guide teaches. Its eDiscovery arm, **e-discovery.uk**, applies the same mathematics at scale: hash-based deduplication, known-file filtering and production-integrity verification across disclosure exercises.

ESTABLISHED 2007 · LONDON | ISO 17025-ALIGNED PROCEDURES | ISO 27001-ALIGNED SECURITY | ACPO / NPCC DIGITAL EVIDENCE PRINCIPLES | CPR PART 35 EXPERT REPORTS | COURT-EXPERIENCED EXPERT WITNESSES | FULL CHAIN-OF-CUSTODY DOCUMENTATION

§ CONTENTS

In this guide

§ 0 1 · ORIENTATION

Executive summary

THE HEADLINE POINT: A HASH PROVES IDENTITY AND INTEGRITY, AND ONLY THAT

A cryptographic hash is an algorithm that reduces any quantity of data (a document, an email, a full disk image) to a short fixed-length value, its fingerprint. The mathematics gives that fingerprint three properties the law can use: the same data always produces the same hash, with any tool, run by anyone; a change of even one bit produces a completely different hash; and the hash cannot be reversed to reveal the data. From these follow the two legitimate courtroom claims: a copy whose hash matches the original's is **bit-for-bit identical** to it, and an exhibit whose hash today matches the hash recorded at acquisition is **unchanged since acquisition**. Everything else commonly claimed for hashes is a misconception: a matching hash does not make a document authentic (a forgery hashes as happily as an original), does not show who created a file or when, and proves nothing about the period before the hash was first computed, which is why late hashing is weak hashing. And "MD5 is broken" does not dissolve evidence: the known weaknesses concern engineered collisions between two specially crafted files, not the ability to fabricate a file matching an existing evidence hash, which remains computationally infeasible; modern practice uses SHA-256, often alongside MD5, and both claims survive.

- **The two model sentences** every litigator should own: "The hash value of the exhibit computed today matches the value recorded at acquisition; the exhibit is therefore mathematically demonstrated to be identical, bit for bit, to what was collected." And its attacking twin: "No hash was computed until [date]; nothing in the mathematics speaks to what happened to this material before then."
- **Hashes do the disclosure heavy lifting too:** deduplication, known-system-file filtering (deNISTing), locating copies of a file across devices, and verifying that productions arrived intact are all hash operations; the fingerprint that authenticates the exhibit also culls the review set.
- **Timing is the whole game.** A hash certifies the data from the moment of computation forward: computed at acquisition, it certifies the exhibit's whole evidential life; computed the week before trial, it certifies a week. Chain-of-custody and hashing are one discipline seen from two sides.
- **Similarity is a different tool.** Cryptographic hashes see identical or utterly different, nothing between: they cannot show that two documents are near-duplicates or that one derives from another. That is fuzzy or similarity hashing, useful, but a distinct technique with distinct weight.
- **Challenge is procedural, not mathematical.** The productive attacks on hash evidence are about when values were computed, what exactly was hashed, and whether the chain between computations is complete: the questions of §9, none of which require disputing the algebra.

§ 0 2 · FIRST PRINCIPLES

The problem in plain English: "how do we know it is the same file?"

Every electronic exhibit invites the same question. The email was collected in 2024, processed in 2025, produced in 2026 and is now on a screen in court: how does anyone know that what the judge is reading is what sat in the mailbox, unaltered by two years of copying, conversion and handling? Paper answered with physical continuity: the same sheet, in the same file, initialled. Digital evidence is copied dozens of times by design, and every copy is an opportunity for corruption, truncation or interference, real or alleged.

The hash is the answer, and the analogies that land with tribunals are worth having ready. It is a **fingerprint**: unique in practice to the data, taken at acquisition, comparable forever after. It is a **wax seal**: intact, it proves nothing was opened and changed; and unlike wax, it cannot be resealed, because nobody can produce different data with the same seal. It is the reason a forensic examiner can hand a copy to the other side's expert with equanimity: both sides can compute the fingerprint independently, with different tools, and the numbers either match or they do not. No trust is required, which is precisely the point: the hash replaces "believe us" with "check for yourself", and evidence that invites checking is evidence a court can rest weight on.

§ 0 3 · THE MECHANISM

How hashing works, for non-mathematicians

A hash algorithm reads data from first bit to last and mixes it, deterministically, into a fixed-length output. Four properties matter, and each is checkable in seconds by anyone with a laptop:

- **Deterministic:** the same input always yields the same output. Hash a file today in London with one tool and next year in Leeds with another, and the values match to the last character, or the data differs. This is why opposing experts can verify each other without cooperation.
- **Fixed length:** whether the input is a six-word note or a two-terabyte disk image, the output is the same short length (32 hexadecimal characters for MD5, 64 for SHA-256): a fingerprint, not a summary; nothing of the content is readable in it.
- **The avalanche effect:** the smallest change transforms the whole output. The sentence "The quick brown fox" and the same sentence with a full stop added produce values with no visible relationship: not one similar character, no gradient of "nearly the same". This is why a matching hash means identical, and why hashes cannot measure similarity.
- **One-way:** the data produces the hash; the hash cannot be run backwards to produce the data. A schedule of hash values discloses nothing of content, which is why hash schedules can be exchanged freely, even for privileged or sensitive material, as integrity anchors.

That is the whole machine. Everything a court is asked to accept about hashes reduces to these four properties, and an expert who cannot explain them this simply should be pressed until they can.

§ 0 4 · THE NAMES

The algorithm families: MD5, SHA-1, SHA-256

ALGORITHM	OUTPUT LENGTH	STATUS AND ROLE IN PRACTICE
MD5	128 bits · 32 hex characters	The long-serving workhorse, embedded in decades of forensic tooling and case law. Engineered collisions are achievable (see §5), so it is deprecated for security uses; for evidence-integrity verification it remains widely used and defensible, typically alongside a SHA value
SHA-1	160 bits · 40 hex characters	Similar story a decade later: demonstrated engineered collisions, retired from security applications, still encountered in tooling and legacy evidence sets; treated like MD5 in forensic use
SHA-256 (SHA-2 family)	256 bits · 64 hex characters	The current standard: no practical collision or preimage attacks; the default for new acquisitions, laboratory procedure and anything security-sensitive. SHA-2 siblings (SHA-384, SHA-512) and the newer SHA-3 family exist and behave equivalently for these purposes

- **Dual hashing is standard laboratory practice:** computing MD5 and SHA-256 together at acquisition. The pairing is belt and braces (a fabricated collision would have to defeat both simultaneously, which is not a live prospect) and keeps new evidence comparable with older tool ecosystems and historical hash sets.
- **The choice of algorithm is rarely the issue.** Any of the three, computed at acquisition and re-verified through the chain, does the evidential work; a SHA-256 value computed for the first time last week does less than an MD5 computed at seizure three years ago. Discipline outranks algebra.

§ 0 5 · THE HARD QUESTION

The collision question, answered honestly

Sooner or later, an opponent or a juror will have read that MD5 is "broken". The honest, and reassuring, answer requires one distinction:

- **A collision attack** means constructing *two new files, together,* engineered so that they share a hash. This is demonstrably achievable for MD5 and SHA-1: researchers have published pairs of crafted documents with identical values. It is why these algorithms are retired from security roles like certificate signing, where an attacker controls both artefacts.
- **A preimage attack** means fabricating a file to match *an existing, given* hash: taking the value recorded in the acquisition log and constructing different data that produces it. For MD5, SHA-1 and SHA-256 alike, no practical preimage attack exists; the computation remains infeasible by margins that are not close.
- **Why the distinction decides the forensic question:** the evidential claim is always of the second kind. The hash was computed from real evidence at acquisition; the suggestion that the exhibit was later swapped requires an adversary to have fabricated substitute data matching that pre-existing value, which is a preimage, which nobody can do. The collision literature, spectacular as it is, describes a different trick with no purchase on the scenario.
- **The complete courtroom answer, in three sentences:** "The published weaknesses in MD5 concern creating two specially crafted files that share a value. They do not enable anyone to construct a file matching the value of this evidence, recorded at acquisition. For that task MD5 remains computationally secure, and in any event the exhibit was dual-hashed, and the SHA-256 value also matches."

§ 0 6 · THE WORKHORSE

What hashes are used for in practice

- **Acquisition verification:** the image or export is hashed as it is taken; the value is the exhibit's birth certificate, recorded in the acquisition log and chain-of-custody record.
- **Transfer and handling verification:** re-computation at every movement (into storage, to the review platform, to the other side) proves each copy true; an unbroken chain of matching values is the exhibit's whole biography in numbers.
- **Deduplication:** identical documents share a hash, so review platforms suppress duplicates by fingerprint: exact and reliable for true bit-level duplicates, with the important limits (a re-saved or differently exported "same" document hashes differently) covered in this series' deduplication guide.
- **Known-file filtering:** reference libraries of hashes for standard operating-system and application files (the NIST NSRL) let processing discard millions of irrelevant system files unexamined: deNISTing, the free first cull of every disclosure exercise.
- **Known-item identification:** the same mechanism in reverse: matching against hash sets of known illegal or targeted material identifies items without human viewing; law-enforcement image classification runs on it, and civil investigations use the technique to find copies of specific documents (the pricing book, the leaked report) across an estate by fingerprint.
- **Levels of hashing:** a full-image hash certifies the entire acquisition; per-file hashes certify each document individually; block-level hashing supports resumable and verifiable imaging of failing media. A competent report says which level each quoted value operates at, and a competent reader checks.

§ 0 7 · THE BOUNDARY

What a hash proves, and what it does not

A MATCHING HASH PROVES	A MATCHING HASH DOES NOT PROVE
The two hashed objects are bit-for-bit identical	That the content is true, authentic or genuine: a forgery hashes perfectly
The exhibit is unchanged since the hash was first computed	Anything about the period before that first computation
Every copy in a matching chain is a true copy	Who created the file, when, on what device, or why
Both parties can verify independently, without trust	That the file was lawfully or reliably collected (that is the acquisition record's job)
Two files on different devices are the same file	That two similar files are related: one bit's difference destroys the comparison

The right-hand column is where cases go wrong. Integrity (unchanged since acquisition) and authenticity (genuinely what it purports to be) are different questions: the hash answers the first completely and the second not at all. Authenticity is established by the surrounding evidence: metadata, artefacts, context, the other guides of this series; the hash's contribution is to guarantee that the material being analysed for authenticity is the material that was collected. Both matter; neither substitutes for the other; and the examiner or advocate who elides them hands the other side its best point.

§ 0 8 · THE GALLERY

The misconception gallery

- **"The hashes match, so the document is genuine."** No: the document is *unchanged since acquisition*. If it was forged before collection, it is a perfectly preserved forgery. Integrity is not authenticity.
- **"The hashes differ, so someone tampered with it."** No: any difference (a re-save, a format conversion, an added byte of metadata, a different export route) produces a different value. Differing hashes prove the objects are not identical; the innocent explanations are legion and must be excluded before "tampering" is said aloud.
- **"MD5 is broken, so the integrity evidence fails."** No: the break is engineered collisions between two crafted files, not fabrication against an existing evidence value (§5); and dual hashing forecloses the point entirely.
- **"The hash proves when the file was created."** No: a hash has no date inside it. The dates live in metadata and logs; the hash merely guarantees those have not changed since acquisition.
- **"These two documents are 95 per cent similar; the hashes nearly match."** Impossible: cryptographic hashes have no "nearly". Near-duplicate and derivation analysis uses similarity (fuzzy) hashing and other techniques, which are real, useful and differently weighted evidence.
- **"We hashed everything last month, so the collection from two years ago is verified."** No: verified *as of last month*. The hash certifies forward from computation, never backward, and the gap is exactly the period a challenge will occupy.
- **"Hash values are sensitive and cannot be shared."** The reverse: they are one-way and content-free, and exchanging hash schedules is how parties give each other verifiability without giving content.

§ 0 9 · THE LANGUAGE

Hashes in court: the exact language

Presenting your own evidence

- **The core sentence:** "Exhibit [X] was hashed at acquisition on [date]; the values are recorded at [reference]. The values were re-computed on [dates of each transfer] and today, and match on every occasion. The exhibit is therefore demonstrated, mathematically, to be identical bit for bit to the data collected on [date]."
- **The invitation that disarms:** "The master copy and its recorded values have been available to [opponent]'s expert throughout; verification requires no cooperation from us, and none has been suggested to fail."
- **The boundary, volunteered:** "The hash establishes integrity since acquisition. Whether the document is authentic (what it purports to be) is addressed separately at [section], on the metadata and artefact evidence." Conceding the limit before it is put is what credible expert evidence sounds like.

Challenging the other side's

- **When:** "On what date was each value first computed, and by whom? What record of that computation exists?" (A hash first computed at report-writing certifies the report-writing.)
- **What:** "Is that value the hash of the image, the file, or the produced copy? Do the per-file values derive from the verified image, and can that derivation be shown?"
- **Between:** "Values were computed at acquisition and again at production, fourteen months later. What verification exists for the period between, and where was the material held?"
- **The absent hash:** "No hash values exist for this material at all. On what basis, other than your say-so, is the court invited to find that the exhibit is what was collected?" This is the question casual collection cannot answer, and the reason guide 31's disciplines exist.

§ 10 · IN THE WILD

Worked examples

EXAMPLE 1 · THE PLANTING ALLEGATION, DISSOLVED IN FOUR QUESTIONS

A defendant asserted that an incriminating file had been added to his laptop's image after seizure. Cross-examination of the examiner ran: When was the image hashed? ("During acquisition, on 14 March; MD5 and SHA-256, recorded in the worksheet exhibited at C4.") Verified since? ("On each of the five occasions the image was copied or accessed, listed at C6; all values match.") Could the file have been inserted without changing the values? ("No. Any alteration of the image, of any size, changes both values.") Has my expert been able to check? ("Yes; your expert re-computed both values in June and confirmed the match in the joint statement.") Re-examination was unnecessary. The allegation was not renewed in closing.

EXAMPLE 2 · "MD5 IS BROKEN" MEETS THE DISTINCTION

Counsel put to an examiner that MD5 was "cryptographically broken" and the integrity evidence therefore worthless, armed with a printout about collision research. The answer followed §5: the research constructs two new files engineered together to share a value; it does not enable anyone to fabricate data matching the pre-existing value recorded at this acquisition, a task that remains computationally infeasible; and the exhibit was in any event dual-hashed, with the SHA-256 value, to which no published attack applies, also matching throughout. The judge's ruling adopted the collision-versus-preimage distinction in a paragraph, and the point has not troubled that courtroom since.

EXAMPLE 3 · THE LATE HASH THAT CERTIFIED A WEEK

An employer's disclosure included documents copied from a leaver's laptop by IT, unhashed, ten months before proceedings. Shortly before exchange, the employer's provider hashed the copied set and exhibited a tidy schedule. The tribunal accepted the claimant's submission on timing: the values proved the set unchanged since the week they were computed, and nothing whatever about the ten months before, during which the laptop had been reissued and the copy had lived on an open network share. The documents stayed in evidence, at a weight that decided nothing; the employer's counsel was reduced to relying on the IT manager's recollection, which is the precise condition hashing exists to prevent.

§ 11 · WHERE IT GOES WRONG

Common mistakes and technical limitations

Common mistakes

- **Hashing late,** or never: the baseline computed after months of casual handling certifies the casual handling.
- **Claiming authenticity from integrity:** the elision an alert opponent turns into a credibility point against the whole report.
- **Alleging tampering from a mismatch** before excluding conversions, re-saves and export-route differences.
- **Quoting values without their level:** image hash, file hash and production-copy hash asserted interchangeably, unravelling under the "what exactly was hashed?" question.
- **Broken chains of computation:** acquisition and trial values with nothing between, leaving fourteen silent months for cross-examination to furnish.
- **Refusing to exchange hash schedules,** forfeiting the cheapest credibility available in disclosure.
- **Treating "MD5 is broken" as either fatal (when defending) or sufficient (when attacking):** both sides of the same misunderstanding.

Technical limitations

- **Hashes are binary comparators:** identical or different, nothing between; similarity, derivation and near-duplication need fuzzy hashing and analytical techniques, which give probabilistic answers and are presented with correspondingly measured language.
- **The hash inherits the acquisition:** a live image is a verified snapshot of a moving system: the values prove the image unchanged, and the acquisition log carries what the image is a snapshot of. Hashes certify the copy, never the circumstances.
- **"Same document" is bigger than "same bits":** one letter saved as .docx, exported as PDF and printed to image is one document and three hash values; deduplication and identity arguments must respect the difference.
- **Verification requires the record:** a hash proves nothing unless the original computation is documented; the mathematics is only as strong as the worksheet it was written on, which returns everything, as always, to guide 31.

§ 12 · INTERROGATORIES & DRAFTING AIDS

Questions to ask · Suggested wording

Ask your client

- Were any copies of relevant data made before forensic involvement, and does anyone know whether they were hashed at the time? (The answer sets the integrity ceiling for that material.)
- Where do the acquisition records and hash schedules for our own collections live, and are they filed with the custody records as one set?

Ask your opponent

- Please provide, for each collection relied upon, the hash values computed at acquisition, the algorithm(s) used, and the dates and results of each subsequent verification.
- Please confirm what each provided value is a hash of (image, file, or production copy) and how per-file values derive from the verified acquisition.
- Please confirm that the master images or exports remain available for independent hash verification by our expert.

Ask your eDiscovery / forensic provider

- Which algorithms do you compute at acquisition, and is dual hashing your standard?
- At which handling events are values re-computed, and how are the results recorded?
- Can you produce a single hash schedule per exhibit for exchange, and verify an opponent's productions against their schedule on receipt?

SUGGESTED WORDING · REPORT OR WITNESS STATEMENT PARAGRAPH (INTEGRITY)

> "At acquisition on [date] I computed MD5 and SHA-256 hash values for [exhibit], recorded at [reference]. I re-computed both values on [dates], being each occasion on which the exhibit was copied, transferred or accessed, and again on [date] for the purposes of this report. On every occasion the values matched those recorded at acquisition. The exhibit before the court is accordingly demonstrated to be identical, bit for bit, to the data acquired on [date]. I make no wider claim: the hash values establish that the material has not changed since acquisition; the authenticity of the material's content is addressed separately at section [n]."

SUGGESTED WORDING · REQUEST FOR VERIFICATION MATERIAL

> "So that the integrity of the disclosed material can be verified independently, please provide within [14] days: (1) the hash values (stating the algorithm) computed at acquisition for each source collected, with the contemporaneous record of their computation; (2) the dates, results and records of each subsequent verification; (3) confirmation of what each value hashes (acquisition image, individual file, or production copy); and (4) confirmation that master copies remain preserved and available for re-computation by our expert. If no acquisition-time values exist for any source, please say so and identify the earliest date on which values were computed."

§ 1 3 · QUICK CONTROL

Checklist and red flags · When to involve a digital forensic expert

The hash-evidence checklist

- ☐ Values computed at acquisition, dual-hashed, recorded contemporaneously in the worksheet
- ☐ Re-computation at every transfer and handling event, results logged in the custody record
- ☐ Level of every quoted value stated: image, file or production copy, with derivations shown
- ☐ Hash schedules prepared per exhibit, exchange-ready, and offered to the other side
- ☐ Opponent's productions verified against their schedule on receipt; discrepancies raised early and neutrally
- ☐ Integrity and authenticity kept rigorously separate in every report, statement and submission
- ☐ The collision-versus-preimage answer prepared before anyone says "MD5" in a hearing
- ☐ Late-hashed or never-hashed material identified in your own case, and its weight assessed honestly first
- ☐ Master copies preserved and available for independent re-computation throughout
- ☐ Fuzzy-hashing and similarity claims, where made, labelled as such and never dressed as cryptographic identity

Red flags

- Disclosure with no hash values anywhere.
- Values first computed at report-writing, production, or the eve of trial.
- A schedule that cannot say what was hashed, or when, or by whom.
- "The hashes match, so it is genuine" in anyone's skeleton, including yours.
- Tampering alleged on a bare mismatch.
- Refusal to allow independent re-computation of preserved masters.
- An expert who cannot explain the avalanche effect in one sentence.

When to involve a digital forensic expert

For anything beyond the schedule itself: designing the verification protocol when productions will be exchanged at volume; re-computing and verifying an opponent's values against their masters; diagnosing a mismatch (conversion, corruption or something worse) before anyone pleads it; deploying similarity hashing where derivation or near-duplication is the real question; and giving the CPR Part 35 / CrimPR Part 19 evidence when integrity is contested, including the §5 answer delivered with the authority of someone who computes these values daily. And earliest of all: at collection, because every strength this guide describes is created in the first hour of guide 31 and merely reported ever after.

§ 14 · COMMON QUESTIONS

Frequently asked questions

Can two different files ever share the same hash?

In theory yes (the outputs are finite and the inputs infinite), but for a court's purposes the honest framing is this: no two meaningfully different files have ever been found to share a SHA-256 value; the engineered MD5 and SHA-1 collision pairs are laboratory constructions built together on purpose; and the accidental collision probabilities are so small that every hash comparison ever run in litigation, worldwide, has not produced one. Dual hashing takes even the theoretical residue off the table.

Can a hash be reversed to reveal the document's contents?

No: the algorithms are one-way, and the 64 characters of a SHA-256 value contain no recoverable content: a two-terabyte image and a birthday card produce values of identical length and equal opacity. This is why hash schedules can be exchanged freely as verification anchors, even where the underlying material is privileged, without disclosing anything of what it says.

Should we insist on SHA-256 everywhere?

For new work, expect it as the default (usually alongside MD5); for existing evidence, no: an MD5 computed at acquisition and verified through the chain is sound integrity evidence, per §5, and re-hashing the preserved masters with SHA-256 now is a cheap supplement any laboratory will perform. The instruction worth giving is not "use this algorithm" but "hash at acquisition, dual where practicable, verify at every step, and record everything".

What can be done if hashes were never taken?

Compute them now, and be honest about what they mean: integrity is established from today forward, and the earlier period rests on whatever else exists: custody records, storage access logs, consistency with other sources, witness evidence of handling. Sometimes that composite is persuasive; sometimes Example 3 happens. Either way, the material is rarely excluded for the gap alone in civil proceedings; it is weighed with it, which is precisely the risk assessment your case plan should make before the other side makes it for you.

What is dual hashing, and is it necessary?

Computing two algorithms' values (conventionally MD5 and SHA-256) over the same acquisition, so verification can proceed against either and any theoretical attack must defeat both at once. Necessary, strictly, no: a verified SHA-256 alone is robust. But it is effectively free at acquisition, forecloses the entire "MD5 is broken" theatre in advance, and keeps evidence comparable across old and new tooling, which is why it is standard laboratory practice and a reasonable expectation of any provider.

Do hashes help us find our document on the other side's systems?

Yes, precisely: give the examiner the file and its hash, and any bit-identical copy anywhere in an estate is findable by fingerprint, regardless of filename or location: the standard technique in leak and exfiltration cases. Its limit is the avalanche effect: a copy that was edited, re-saved or format-shifted no longer matches, and the hunt shifts to similarity hashing and content analysis, which find the family rather than the twin.

§ 15 · REFERENCE

Glossary

Avalanche effect
The property that any change to input data, however small, completely transforms the hash.

Collision attack
Constructing two new files together that share a hash; achievable for MD5/SHA-1, irrelevant to matching an existing evidence value.

deNISTing
Discarding known operating-system and application files by matching against reference hash libraries (NSRL).

Dual hashing
Computing two algorithms (typically MD5 and SHA-256) at acquisition; standard laboratory practice.

Fuzzy (similarity) hashing
Techniques scoring how alike two files are; probabilistic, distinct from cryptographic identity.

Hash schedule
The exchange-ready record of values per exhibit: algorithm, value, computation dates, verifications.

Hash value
The fixed-length fingerprint an algorithm derives from data; identical data, identical value.

MD5
128-bit legacy algorithm; deprecated for security, defensible for acquisition-anchored integrity, used in dual hashing.

NSRL
NIST's National Software Reference Library of known-file hashes, the deNISTing reference set.

One-way function
The property that a hash cannot be reversed to reveal the data that produced it.

Preimage attack
Fabricating data to match a given existing hash; computationally infeasible for all algorithms in forensic use.

SHA-256
The current-standard 256-bit algorithm of the SHA-2 family; no practical attacks.

Sources and authoritative references

The verification disciplines in this guide draw on materials published by our laboratory and e-discovery practice, referenced here as sources:

- cflab.uk, digital forensics services (hashing at acquisition, verified transfer, preserved masters and CPR Part 35 / CrimPR Part 19 reporting of integrity evidence): cflab.uk/digital-forensics-services
- cflab.uk, the digital-evidence guide for defence lawyers (hash values as the digital fingerprints proving copies identical) and the lawyer guides library: cflab.uk/guides
- e-discovery.uk, EDRM Phase 04 · Processing (hash-based deduplication and known-file filtering) and Phase 06 · Production (manifest and hash exchange at production): e-discovery.uk/edrm/processing · e-discovery.uk/edrm/production
- e-discovery.uk, practice method and Knowledge Centre: e-discovery.uk · e-discovery.uk/knowledge
- NIST, National Software Reference Library (known-file hash sets) and Secure Hash Standard (FIPS 180-4): nist.gov, NSRL · csrc.nist.gov, FIPS 180-4
- NPCC / College of Policing digital evidence guidance (audit trails and integrity): college.police.uk, digital devices guidance
- Forensic Science Regulator, statutory Code of Practice: gov.uk/forensic-science-regulator
- ISO/IEC 27037:2012 and ISO/IEC 17025: iso.org, 27037 · iso.org, 17025
- CPR Part 35 (experts) and CrimPR Part 19: justice.gov.uk, Part 35 · justice.gov.uk, CrimPR Part 19

DISCLAIMER

This publication provides general information about cryptographic hash values in digital evidence in the United Kingdom as at August 2026. The worked examples are fictional illustrations. Cryptographic research develops; the status of algorithms described here should be checked against current authority for any contested matter. This guide is not legal advice, does not create a professional relationship, and should not be relied upon in place of advice on the facts of a specific matter from a qualified lawyer or a suitably qualified forensic practitioner. Links were live at the date of publication.

§ HOW A SPECIALIST LABORATORY CAN ASSIST

Working with Computer Forensics Lab

Hashing is where our laboratory's daily practice and the courtroom meet most cleanly. We compute dual values at every acquisition, verify at every transfer, and deliver exhibit-ready hash schedules alongside custody records as a single set; we verify opponents' productions against their schedules, re-compute values on preserved masters for independent confirmation, diagnose mismatches before anyone pleads them, and deploy similarity hashing where the question is derivation rather than identity. Our examiners give the §5 collision-versus-preimage evidence under CPR Part 35 and CrimPR Part 19 in terms judges adopt, and through **e-discovery.uk** the same mathematics runs the disclosure engine: deduplication, deNISTing, known-item searches across estates, and manifest-and-hash verified productions. If a single question from this guide is live in your matter (a missing baseline, a mismatch, a "broken MD5" skeleton), an examiner will answer it the same working day. Conflicts checks, confidential scoping discussions and fixed-fee estimates are routine.

INSTRUCT THE LAB

Speak to a forensic examiner, not a salesperson.

Describe your matter in confidence and an examiner will respond the same working day. Conflicts checks and scoping calls with US counsel are routine.

NEW ENQUIRIES
+44 (0)20 7164 6971

EMAIL
info@cflab.uk

E-DISCOVERY
e-discovery.uk

Computer Forensics Lab Ltd · Euro House, 133 Ballards Lane, Finchley, London N3 1LJ, United Kingdom
cflab.uk · Customer service +44 (0)20 7164 6915 · ICO ZB395023 · NPCC · ISO 17025 aligned practice · UK Gov Digital Marketplace