SCIENCE COLUMN: RECONSTRUCTION: THE EXPERIMENTAL SIDE OF DIGITAL FORENSICS

Fred Cohen
California Sciences Institute
California, USA

ABSTRACT

Many in digital forensics seem to forget that the science part of digital forensics means experimentation and that implies a whole lot of things that most practitioners never learned.

MOST CASES

Most cases don't seem to involve anything in the way of experimentation or science. They involve bag and tag, transport, search, and present. Looking for contraband content? No problem. If you know what you are looking for and it is found, it is present. The reason it is seemingly so simple is that there is a whole lot of research and development surrounding the searching process that has been done for the last 50+ years. And if found, it's usually pretty easy to verify that what was found was what you were looking for. Find a file, compare the cryptographic checksum to a known item for a presumptive positive, compare bit-for-bit to the original for a definitive answer, selectively look at them with another tool just to make sure, and tell the court: "[This] is what I did, [this] is what I saw." But maybe that's just most of your cases... it's almost none of mine.

MY CASES

Maybe I am unusual. I don't really know. I don't tend to take on cases where there are 2 hours allowed for an answer. I tend to take on cases where there is a period of months over which different issues come up. I tend to be in cases where so-called experts say or write things that don't, to me, appear on their face to be strictly true. And I tend to be in cases where I have to make sure before I make a declaration. These cases almost always involve some experimentation. Or as I call it–reconstruction.

Suppose someone claims that some e-mail was sent from some place because an email header says so. My first reaction is to ask why they come to that conclusion. What is the basis? And in most of the cases I have encountered, they haven't told me the basis for their opinions in their reports. So I check it

out. Suppose they mistakenly say something like "all [X] are [Y]". That's almost never true, so I check it out. And so should you.

CHECK IT OUT

"When in doubt, check it out." If it isn't already an official saying for digital forensics, it should become one. Cite me on it. The lawyers prefer this, I think. They seem to me to really want to know the truth. If it goes against them, they want to know that so they can prepare for the contingency. And if it goes for them, they want to know because it helps their case. In either case, it really hurts if they think one thing and later find out something else.

To check it out, the first thing I usually do is try to repeat whatever the other side did. It's not that I don't believe them. I just think it's your job to check out the claims of the folks on the other side of the isle. If they say that Google searches don't come up with the same result much of the time, even though my experience agrees with them, I will still check it out. I just did. I did the following search:

> site:all.net supercalifragilisticexpialidocious

It came back with:

> Your search–site:all.net supercalifragilisticexpialidocious - did not match any documents.

Suppose you repeat it. See if your result differs. If not, we have just refuted by experiment something that we may well have thought was true. I know, if you do it in another language, it might come out differently, but if it means the same thing then the result is equivalent (the meaningful part–none found–if that is the thing we are trying to understand).

You think this is cheating? Because it's only true for searches returning no results? Try looking for something else at all.net and see what happens. Try disabling the advertising or ignoring it if it's not the relevant part of the issue. So maybe I need to update my saying... "Always doubt, check it out."

MOST EXPERIMENTS ARE PRETTY SIMPLE

I like to think that most of the most meaningful experiments in digital forensics are really simple enough for the jury or judge to do on their own. If not, they take a lot more explaining. I don't mind all that explaining, but I think it works better when there is less of it and the judge and jury say to themselves "I get it" as opposed to "Huh?"

CAUSALITY AND REFUTATION BASICS

I have said and will likely say this often and again. Effect does not imply cause. Rather cause (C) acting through mechanisms (m) produces effects (E), expressed as $C \rightarrow^{m} E$. To have a scientific hypothesis, it is not enough to state that C produces E, it is also necessary to identify the mechanism by which C produces E. Testing can then be repeatedly done to confirm or refute the hypothesis of $C \rightarrow^{m} E$ by trying to refute the hypothesis. If refutation fails, it is a confirmation, while if refutation succeeds, they hypothesis as stated cannot be correct.

If that sounds like exactly what I said last time, it is. It continued:

While many confirmations may be found, any number of confirmations of a universal statement do not prove it to be correct, while a single refutation demonstrates its falsehood[1]. Typically, science progresses when a refutation is identified, the errors in the $C \rightarrow^{m} E$ hypothesis are identified, and an updated $C' \rightarrow^{m'} E$ version of the hypothesis is created to mitigate the refutation cases, or the hypothesis abandoned.

When faced with a statement that is not couched in the proper specifics, always try to refute it by experiment. Predict a cause and effect per the proposed claim. Create a cause that meets the specifics of the claim, observe the effect, and see if the prediction fails to match the expectation. Try edge cases, like my "no result" search. Try center cases. Try the exact thing the other side tried, but with unspecified conditions set in other ways. Try things! And report them all. Don't just say I found this refutation. Say that you tried some cases that worked out as predicted (if you find them).

To serve your client in a system based on an adversarial approach, you might end up not testing exhaustively. The client may not support it because it has a real financial cost. And you are not obligated to try to prove the case for the other side. But on the other hand, sometimes you end up exhausting the space and never coming up with a confirmation. But hopefully, when you are making such statements, that won't happen... because...

TEST YOURSELF BEFORE YOU EXPRESS OPINIONS

Even when you know you are right, you should still perform tests to try to refute your own claims. You could take the position that it has to be done because I might end up on the other side and you don't want to look bad when I come up with a counterexample. But a better reason is because it's the right thing to do in a scientific field. But I admit that's not why I do it. I do it because

[1] K. Popper. The Logic of Scientific Discovery (1959), Hutchins and Company, London. ISBN10: 0415278449.

it's not enough to believe I am right, I need to prove it to myself experimentally before I am willing to tell it to the judge and jury.

One of the areas I most often encounter in this regard is metadata. Before going into more detail, I should point out that metadata as used in the computing field differs from metadata as used in the archives and records management space. Notably, this difference is not always made clear to the legal community which is used to metadata associated with archives and public record-keeping systems as the information that makes definitive the chain of custody and provenance of such records, rending them accurate and reliable. This is not the case with regard to metadata as it is used in the computing fields, where, for example, the date and time stamp on a file can be readily changed by the user.

Regardless of its ready changeability in computer systems, the sort of metadata associated with file and directory date and time stamps and ownership is produced by automated mechanisms that are specific to particular operating environments. Because they differ depending on a wide range of different factors, claiming meaningful information about metadata without adequate experimental basis is likely to be problematic if challenged. And it often yields results that are not accurate, are less precise then claimed, and end up asserting events in an order different from the reality of what took place.

In order to be more certain about such results, or in order to test the claims of others, reconstruction is really the only available method. While the literature on how date and time stamps are produced may indicate one thing or another, there have been various experiments performed and results published that indicate different results for the apparently same tests. These differences may be due to patch variations or other environmental conditions, and thus testing in situ is the optimal approach for getting the right answer.

PRECISION, ACCURACY, AND RELIABILITY

Precision has to do with the number of digits provided. For example, a time stamp may indicate results to the nearest second, microsecond, millisecond, or some other time span. But just because the indication is to this many digits of precision doesn't mean the measurement itself is. A timestamp, in some cases, is only recorded based on the date, ignoring the time. But the record may be in a format including fields ranging to the millisecond. Or the routine that reads the record may return results with fields that don't exist in the original data, set to pre-defined default values. In examination and presentation, limiting the precision of reporting is important to providing the trier of fact with the real information presented without a false sense of precision. If the measurement is to the nearest second, adding the millisecond fields is misleading. In comparing two measurements recorded to different levels of precision, if they are too close to each other, you cannot reliably indicate an ordering and should

indicate that you cannot do so, presenting the ordering information so that the multiple possible orderings are readily apparent.

Accuracy has a lot to do with whether the things being measured are properly calibrated. For example, even though a measurement may be precise to the nearest millisecond, it could be offset from another measurement by hours or days. Even with network time protocol (NTP) working properly, variances of milliseconds to seconds are not unusual. And it is not always working properly. Sometimes, even though it is running, it is not connecting with remote servers. And recent demonstrations show that NTP is attackable and forgeable, as is global positioning system (GPS) data. Cellular phones use data from the telephone network which can and often is offset by seconds from other time sources. While the problem of accuracy has long been known and many folks have tried to subvert forensics by changing system clocks, even without malice, you have to be careful to get it right and report it properly.

Reliability is something entirely different. In the term of art of archives and records management, it is the correspondence of the records to reality. This is something that is very difficult in terms of the current state of the science. Tying a computer record to reality involves a lot of indirection (hopefully not misdirection).

RECONSTRUCTION TO THE RESCUE

Because of the nature of digital systems and the fact that effect doesn't imply cause, looking at an audit trail indicating a login of a user identity at a time from a terminal identifier is not adequate to tell us that someone logged in with that user identity from that terminal at that time. But we can drive causality forward by reconstructing the conditions on the system and doing tests of various theories of the case, demonstrating that the tests came out consistent with some set of theories of the case and inconsistent with another set of theories of the case.

The question of what theories to test and to what level of certainty depends on the nature of the case and the standard of proof. For example, in a patent case for invalidity, an existence proof of prior art is all that is required. In such a case, a single reconstruction of a system existing prior to the relevant dates in the case that fulfills all of the elements of the patent may be perfectly adequate. But in another case, such as an attribution case where there are many possibilities of ways to bypass the normal authentication process, much more may be required.

Attribution of computer-related acts in a criminal case where the defendant claims they didn't undertake the relevant acts is an example of a far more complex situation. It's hard to prove a negative (I didn't do it), but it is also non-trivial to prove a positive beyond a reasonable doubt. For attribution,

regardless of the issues outside of the computer, which must be addressed to establish presence of the individual at the time and place asserted as used for access, the acts of the individual as opposed to alternatives become a critical issue, if challenged. Did someone else have access? If they did, could they have yielded the same traces identified? Did the system actually work as asserted? Was there a Trojan horse or other mechanism in place that altered the normal operation?

Given the assumption that recent revelations regarding government access programs for user and owner unauthorized access are true, this becomes more than a theoretical issue. Once we start to realize that the technology for covert unauthorized access has long been distributed and used on a large scale, the question of proof of activities becomes harder to show. How do we know that another actor did not commit the acts being denied by the suspect? If a Trojan is in place and well hidden, how can we find it? If we find one, is that the only one? If we are able to use it, can we produce the traces found? If we can produce those traces, how do we show that nobody else in fact undertook acts to produce the traces? The whole foundation of trust required to make assertions about traces reflecting reality and causality come into question.

What we can do is reconstruct scenarios in which the actor did undertake the act and seek to differentiate the alternatives from each other by examination of increasingly details sets of traces. As we move through this process, if we find that none of the traces we can reproduce meaningfully demonstrate the act, we cannot assert the attribution of the act to the actor and in fact must both refute those acts and confirm that something else we do not understand took place. Any hypothetical that is consistent with the available traces is possible, including system subversion by unknown 3rd parties.

If and to the extent we identify specific sets of event sequences consistent with theories of the case, with the traces found in a reconstruction, and with the traces found in the original writing, we can assert that the cause could have produced the effect to the extent the methods we used were able to differentiate, and through refutation, eliminate other possibilities while potentially leaving still other possibilities not yet confirmed or refuted. That is the limit of what we can do.

PRECISION, ACCURACY, RELIABILITY, AND RECONSTRUCTION

Reconstruction can rarely, if ever, produce exactly the same traces as the traces in the original writing. Because most digital systems are quite complex, even though many of their component parts are highly repeatable, timing, state information, external context, and other similar phenomena produce different traces for repetitions of the same experiment. It is impossible to perfectly reproduce an experiment and thus repetition of experiments are not always identical. The differences and similarities between experiments and their

results have to be understood in the context of the matter at hand in order to meaningfully reproduce events and understand results. This then comes down to understanding the base rates of differences, at a minimum, in terms of precision, accuracy, and reliability.

For example, to reproduce a Web access sequence to test the logs generated, files altered, and technical metadata produced by the process, we might reasonably perform a set of experiments in nearly identical conditions to see what the variances are in results of the measured phenomena as reflected in traces. Then we might vary parameters, such as system loads, time of day, cache conditions, and so forth. But which we have to do might also depend on the need to be accurate for a particular case, and the number of repetitions depends on the statistical model in use and the related assumptions. Sampling theory has to be applied to generate the answers to such questions, but only if and to the extent that multiple samples are relevant to the issues at hand.

If you don't know what these things are, chances are you aren't qualified to make decisions about whether and to what extent they are needed in order to get the right answers. While you may not usually have to worry because the experts and lawyers on the other side probably don't know what these things are either, you should still concern yourself with them, because otherwise, you could easily get the wrong answers. And isn't justice about getting at the truth and digital forensics a path toward that end?

EDITORIAL SUMMARY AND DISCUSSION

Building a science surrounding digital forensics requires a methodology for addressing experiments. The experimental branch of digital forensics is, in large part, what we call reconstruction.

Reconstruction involves identifying and repeating event sequences that test theories of the case in order to refute or confirm different theories based on available evidence. This is done by comparing the traces produced in the original writing to reconstructions of event sequences under different theories of the case.

In order to draw conclusions about the results of reconstruction, it is necessary to understand the issues of what constitutes a refutation and confirmation, precision, accuracy, and reliability issues, and the supporting fields such as relevant areas of statistics.

Disagree? Let us know! It's the only way we can grow.