INVESTIGATION · DIGITAL INVESTIGATION

Cybersecurity Incident Investigation

When an organisation's computer systems are breached, it must find out quickly what happened, how far the attacker reached and how to stop them. This article explains how cybersecurity incidents are investigated, from the first alert through evidence collection, analysis and attribution to reporting and lessons learned.

ModQu research article · 2,465 words · 23 references (APA 7th edition)

A cybersecurity incident is an event that threatens the confidentiality, integrity or availability of an organisation's information or systems: a ransomware attack that encrypts files, a stolen password used to read email, malware that quietly copies customer data, or an insider who takes confidential documents. When an incident occurs, the organisation needs answers to urgent questions. How did the attacker get in? What did they do? Are they still present? What data were affected? Who must be told? Incident investigation is the discipline that answers them.

Incident investigation differs from many other kinds of investigation because it usually happens while the threat is still active. The investigator must gather evidence and build understanding at the same time as colleagues try to contain the attacker and restore normal operations, and decisions about one affect the other. Shutting down a compromised server may stop the damage, but it can also destroy evidence held in memory and alert the attacker that they have been discovered.

This article explains the main frameworks for incident handling, the sources of evidence, the analytical methods used to reconstruct an attack, the difficult problem of attribution, and the reporting and learning that follow.

The incident response lifecycle

The most widely used framework is the incident handling guide published by the United States National Institute of Standards and Technology (NIST). It describes a lifecycle with four phases: preparation; detection and analysis; containment, eradication and recovery; and post-incident activity (Cichonski et al., 2012). The international standard on information security incident management describes a similar sequence of planning, detection and reporting, assessment and decision, response, and lessons learned (International Organization for Standardization [ISO], 2016).

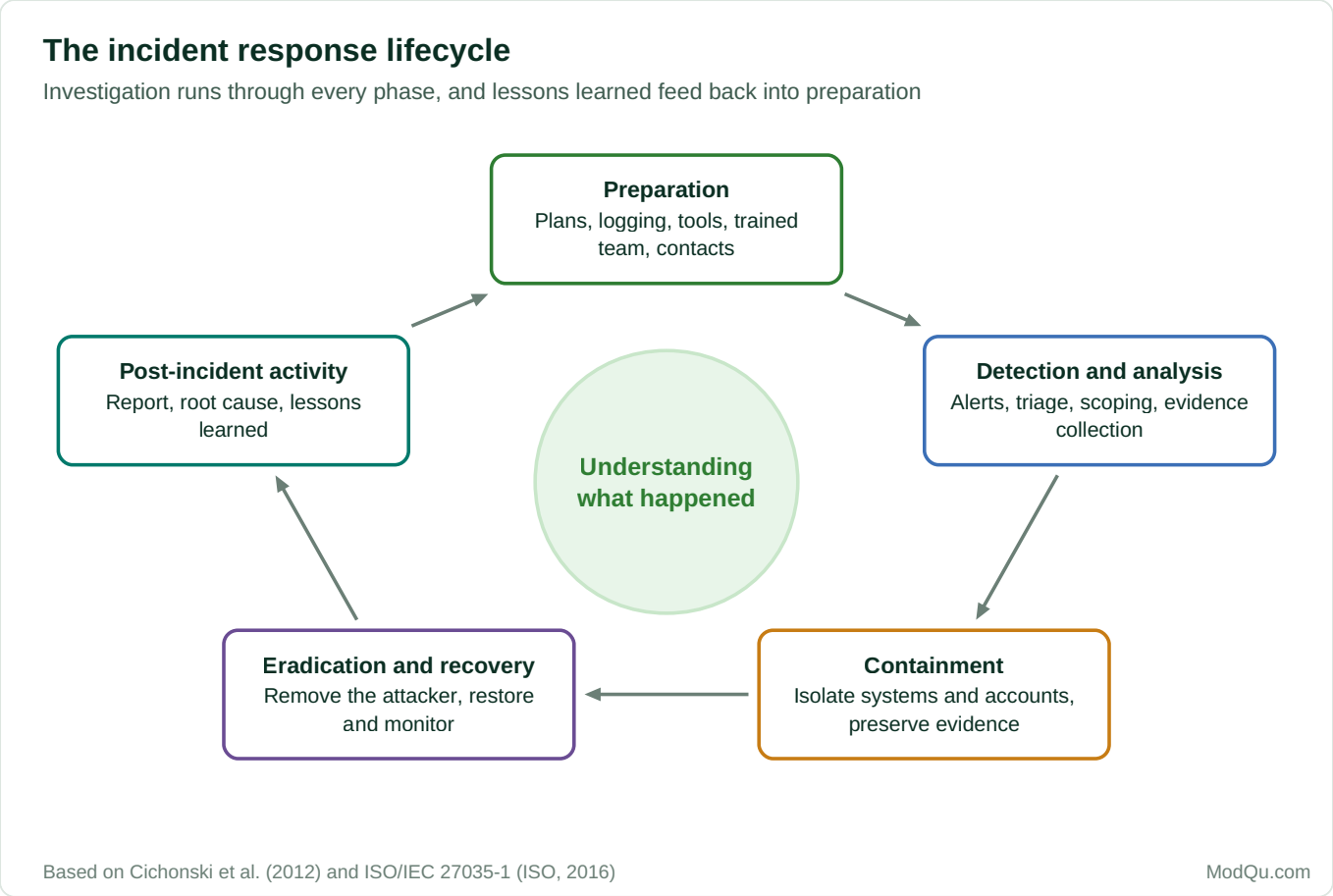


Figure 1. The incident response lifecycle. Investigation is not a separate stage: it begins with detection and continues until the root cause is understood and the lessons have been applied. Diagram by ModQu.

Preparation determines how much an investigation can achieve. Organisations that keep detailed logs, synchronise their system clocks, maintain an inventory of their systems and have a trained team and agreed procedures can reconstruct an attack far more completely than those that do not. Ahmad et al. (2012) found that incident response teams often focused on restoring technical systems quickly and gave less attention to investigation and learning, which limited what their organisations learned from incidents.

Detection and analysis begins with an alert or report: unusual network traffic, an antivirus warning, a user who notices strange behaviour, a ransom note, or a notification from a third party such as a law enforcement agency or a security researcher. The team triages the alert, decides whether it is a real incident, assesses its severity and begins to establish its scope.

Containment, eradication and recovery stop the damage, remove the attacker's access and restore normal operations. Containment choices are informed by the investigation: the team needs to know which systems and accounts are compromised to isolate them all at once, because removing an attacker from one system while they retain access to another simply teaches them to hide better.

Post-incident activity includes the final investigation report, analysis of the root cause, and changes to defences and procedures.

Sources of evidence

Cyber incidents leave traces in many places. Kent et al. (2006) group the sources of digital evidence relevant to incident response into four categories: data files, operating systems, network traffic and applications. In practice, investigators draw on the following.

Endpoint evidence comes from computers, servers and mobile devices. It includes files and file system metadata, which show when files were created and changed (Carrier, 2005); operating system records such as event logs, lists of programs that run at start-up, and records of program execution; and the contents of memory. Memory is especially important because some malicious software runs only in memory and leaves little trace on disk, and because memory holds information about running processes, network connections and encryption keys. Ligh et al. (2014) describe methods for capturing and analysing memory to detect such threats.

Network evidence includes logs from firewalls, proxies and domain name servers, records of network connections, and sometimes full captures of traffic. It can show which internal systems communicated with external addresses controlled by the attacker, and how much data left the network.

Log evidence from applications and cloud services records authentication attempts, administrative actions, email access and file downloads. In cloud environments, provider logs may be the only record of what an attacker did, and their availability depends on the organisation's subscription and configuration.

External evidence includes threat intelligence about known attacker tools and infrastructure, information shared by industry groups and government agencies, and reports from other victims of the same campaign.

Preserving evidence

The principles of digital forensics apply: identify relevant evidence, collect it in a way that preserves its integrity, document every action, and work on copies rather than originals (ISO, 2012; Casey, 2011). Investigators collect the most volatile evidence first, such as memory and active network connections, before it is lost, and record hash values to show that the copies have not changed.

In incident response, full forensic copies of every affected system are often impractical, especially across hundreds of machines. Teams instead collect targeted evidence, such as particular logs, registry entries and memory images, from many systems using remote tools, and make full copies only of the most important ones. The trade-off between speed and completeness should be recorded, because it affects what the investigation can later show. If legal action or regulatory scrutiny is likely, the chain of custody of key evidence must be maintained from the start. Many organisations also involve legal counsel early, so that investigation findings are handled appropriately and communications about the incident are accurate and consistent.

Analysis: reconstructing the attack

The core analytical task is to build a timeline of the attacker's actions, from initial access to the present. Investigators correlate time stamps from different sources, such as an email log showing a malicious attachment arriving, an endpoint record showing it being opened, and a network log showing a connection to an external server a few seconds later. Accurate and synchronised clocks, and an understanding of time zones and time stamp formats, are essential.

Several frameworks help investigators organise what they find.

The intrusion kill chain describes the stages an attacker must complete: reconnaissance, weaponisation, delivery, exploitation, installation, command and control, and actions on objectives (Hutchins et al., 2011). The model helps investigators ask which stages they have evidence for and where defences could have stopped the attack.

The diamond model describes every intrusion event in terms of four linked features: the adversary, their capability, the infrastructure they used, and the victim (Caltagirone et al., 2013). Finding one feature helps the investigator search for the others; for example, an attacker's server address may lead to other victims or to other tools.

MITRE ATT&CK is a public knowledge base of attacker tactics and techniques observed in real intrusions, organised by goals such as initial access, persistence, privilege escalation, lateral movement and exfiltration (Strom et al., 2018). Investigators map their findings to its techniques, which helps them look for related activity and communicate clearly with others.

Investigators also extract indicators of compromise: file hashes, internet addresses, domain names and other artefacts that identify the attacker's activity. These are used to search other systems for signs of the same attacker. Bianco (2013) described a "pyramid of pain" that ranks indicators by how hard it is for an attacker to change them once defenders detect them.

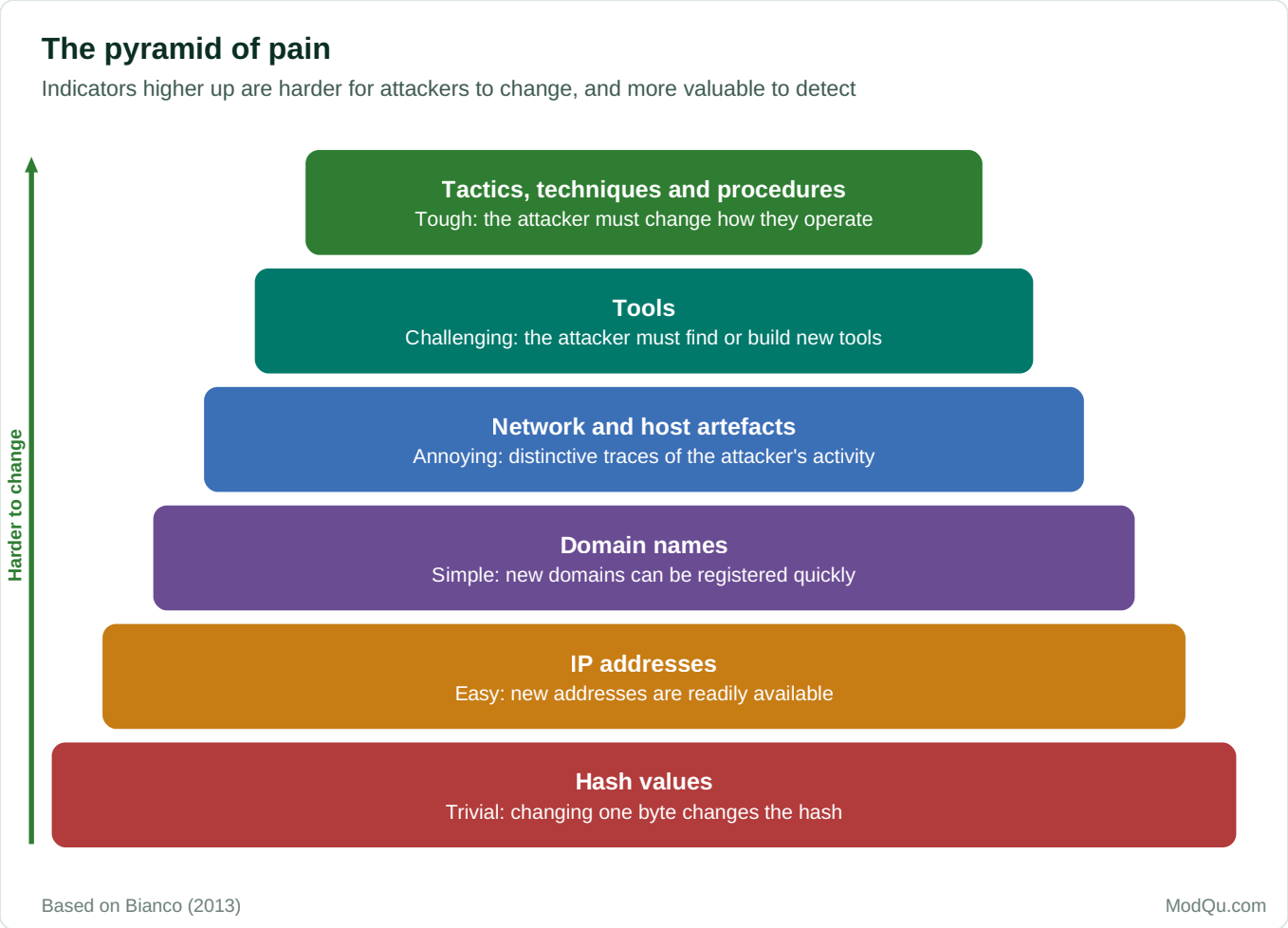


Figure 2. The pyramid of pain. Detecting an attacker's file hashes or addresses helps only briefly, while detecting their methods forces them to change how they work. Diagram by ModQu.

Scoping is a crucial part of analysis. The question is not only what happened on the first system found, but which other systems and accounts the attacker reached. Attackers who gain access to one machine typically try to obtain administrator credentials and move to others. An investigation that misses part of the attacker's footprint leaves the door open for them to return.

Automated tools play an increasing role in detecting suspicious activity, but they do not replace analysis. Sommer and Paxson (2010) explain why machine learning methods that work well in other fields often struggle to detect network intrusions in practice: attacks are rare, normal traffic is highly variable, and errors are costly. Alerts must still be interpreted by people who understand the systems involved.

Determining impact

Organisations must determine what the attacker accessed or took, both to manage harm and to meet legal obligations. This can be difficult when logs are incomplete or when the attacker deleted records. Investigators look for evidence of data being gathered, compressed and transferred out, and estimate the volume from network records. Where certainty is not possible, they state what is known, what is likely and what cannot be excluded.

Legal obligations shape the timetable. Under the European Union's General Data Protection Regulation, organisations must notify the supervisory authority of a personal data breach without undue delay and, where feasible, within 72 hours of becoming aware of it (Regulation (EU) 2016/679, 2016, art. 33). The NIS2 Directive requires essential and important entities to give an early warning of significant incidents within 24 hours (Directive (EU) 2022/2555, 2022). In the United States, public companies must disclose material cybersecurity incidents within four business days of determining that they are material (U.S. Securities and Exchange Commission [SEC], 2023). These deadlines mean that early findings are often reported before the investigation is complete, and later findings may revise them.

Attribution

Attribution means identifying who was responsible for an attack. For most organisations, the practical question is less "which person?" than "what kind of attacker, with what goals?", because that determines whether they are likely to return and what they were after. A criminal ransomware group, a state espionage unit and a disgruntled former employee call for different responses.

Rid and Buchanan (2015) argue that attribution is not a binary technical fact but a process of judgement at several levels: technical (which tools and infrastructure were used), operational (how the campaign was organised) and strategic (who ordered it and why). It draws on technical indicators, on patterns of behaviour, on language and working hours, and, for governments, on intelligence from other sources. Its quality depends on the skill of the analysts and the time available.

Public attribution reports have shaped understanding of particular threats. In 2013 the security company Mandiant published a detailed report linking a long-running series of intrusions to a specific unit of China's military, using technical evidence, infrastructure and the identities of individual operators (Mandiant, 2013). An analysis of the 2015 attack on Ukrainian electricity distribution companies reconstructed how the attackers had spent months inside the companies' networks before cutting power to customers (Lee et al., 2016). Langner (2011) described how analysis of the Stuxnet malware revealed that it had been designed to damage specific industrial equipment, which pointed to its target and the resources behind it.

Attackers can also plant false clues, such as words in another language or tools associated with other groups. Responsible attribution states its confidence and its basis, and avoids certainty that the evidence does not support.

Reporting and learning

The investigation report explains what happened, when, how the attacker got in, what they did, what was affected, how the incident was contained and what should change. Different audiences need different versions: technical teams need indicators and detailed timelines, managers need impact and decisions, and regulators need specific information in a defined form.

Post-incident review is where much of the value lies. The aim is to identify the root causes, such as an unpatched system, a lack of multi-factor authentication or a missing alert, and the reasons they existed, rather than to blame individuals. Public reviews of major incidents can spread these lessons widely. The United States Cyber Safety Review Board,

modelled partly on transport accident investigation, reviewed the response to the 2021 vulnerability in the widely used Log4j software library (Cyber Safety Review Board [CSRB], 2022) and later reviewed an intrusion into Microsoft's cloud email service, concluding that a cascade of avoidable security failures had allowed it (CSRB, 2024).

One of the earliest published accounts of an incident investigation remains instructive. Stoll (1989) described how a 75-cent accounting discrepancy at a university computer centre led him to trace an intruder who was searching military networks for secrets and selling information to the Soviet intelligence service. His account emphasised the value of detailed records, patient observation and cooperation with outside agencies, all of which remain central to modern practice.

A short example

The following hypothetical case illustrates the process. A company's security monitoring system flags an administrator account logging in to a file server at 03:00. The incident response team confirms that the administrator was asleep at the time and treats the event as an incident.

The team captures memory from the file server and finds a remote access tool running under a disguised name. Network logs show that the tool connects to an external server and that 40 gigabytes of data were sent to that server over two nights. Searching all company systems for the tool's file hash and the external address finds the same tool on four other machines.

Working back through the logs, the team builds a timeline. Three weeks earlier, an employee in the finance department received an email with a malicious document, opened it, and entered their password on a fake login page. The attacker used the password to access the employee's computer remotely, found stored administrator credentials, and used them to reach the file servers. Mapping the actions to ATT&CK techniques shows a pattern consistent with a known criminal group that steals data and then demands payment not to publish it.

The team isolates all five machines and resets the affected credentials at the same time, blocks the external server, and restores the systems from clean backups. Because the stolen data include employee personal information, the company notifies the data protection authority within 72 hours and later notifies the affected employees. The post-incident review identifies two root causes: administrator credentials stored on an ordinary workstation and the lack of multi-factor authentication for remote access. Both are fixed within a month.

Conclusion

Cybersecurity incident investigation reconstructs what an attacker did so that the organisation can stop the damage, meet its obligations and prevent a repeat. It relies on preparation, especially good logging and synchronised clocks; on the careful collection of evidence from endpoints, networks and logs; on structured analysis using timelines and frameworks such as the kill chain, the diamond model and ATT&CK; and on honest statements of what is known and what is not. Attribution is possible but is a matter of judgement and stated confidence. The investigation's lasting value comes from the lessons it produces and the changes that follow.

References

Ahmad, A., Hadgkiss, J., & Ruighaver, A. B. (2012). Incident response teams – Challenges in supporting the organisational security function. *Computers & Security, 31*(5), 643–652. https://doi.org/10.1016/j.cose.2012.04.001

Bianco, D. J. (2013, March 1). *The pyramid of pain*. Enterprise Detection & Response.

Caltagirone, S., Pendergast, A., & Betz, C. (2013). *The diamond model of intrusion analysis*. Center for Cyber Intelligence Analysis and Threat Research.

Carrier, B. (2005). *File system forensic analysis*. Addison-Wesley.

Casey, E. (2011). *Digital evidence and computer crime: Forensic science, computers and the Internet* (3rd ed.). Academic Press.

Cichonski, P., Millar, T., Grance, T., & Scarfone, K. (2012). *Computer security incident handling guide* (NIST Special Publication 800-61 Rev. 2). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-61r2

Cyber Safety Review Board. (2022). *Review of the December 2021 Log4j event*. U.S. Department of Homeland Security.

Cyber Safety Review Board. (2024). *Review of the summer 2023 Microsoft Exchange Online intrusion*. U.S. Department of Homeland Security.

Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS 2 Directive). (2022). *Official Journal of the European Union, L 333*, 80–152.

Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. In *Proceedings of the 6th International Conference on Information Warfare and Security* (pp. 113–125). Academic Publishing International.

International Organization for Standardization. (2012). *Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence* (ISO/IEC Standard No. 27037:2012).

International Organization for Standardization. (2016). *Information technology — Security techniques — Information security incident management — Part 1: Principles of incident management* (ISO/IEC Standard No. 27035-1:2016).

Kent, K., Chevalier, S., Grance, T., & Dang, H. (2006). *Guide to integrating forensic techniques into incident response* (NIST Special Publication 800-86). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-86

Langner, R. (2011). Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy, 9*(3), 49–51. https://doi.org/10.1109/MSP.2011.67

Lee, R. M., Assante, M. J., & Conway, T. (2016). *Analysis of the cyber attack on the Ukrainian power grid: Defense use case*. SANS Industrial Control Systems & Electricity Information Sharing and Analysis Center.

Ligh, M. H., Case, A., Levy, J., & Walters, A. (2014). *The art of memory forensics: Detecting malware and threats in Windows, Linux, and Mac memory*. Wiley.

Mandiant. (2013). *APT1: Exposing one of China's cyber espionage units*. Mandiant.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). (2016). *Official Journal of the European Union, L 119*, 1–88.

Rid, T., & Buchanan, B. (2015). Attributing cyber attacks. *Journal of Strategic Studies*, *38*(1–2), 4–37. https://doi.org/10.1080/01402390.2014.977382

Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE Symposium on Security and Privacy* (pp. 305–316). IEEE. https://doi.org/10.1109/SP.2010.25

Stoll, C. (1989). *The cuckoo's egg: Tracking a spy through the maze of computer espionage*. Doubleday.

Strom, B. E., Applebaum, A., Miller, D. P., Nickels, K. C., Pennington, A. G., & Thomas, C. B. (2018). *MITRE ATT&CK: Design and philosophy* (Technical Report MTR180290). The MITRE Corporation.

U.S. Securities and Exchange Commission. (2023). *Cybersecurity risk management, strategy, governance, and incident disclosure* (Release No. 33-11216).