

Defending an Organization

SIO

deti universidade de aveiro
departamento de eletrónica,
telecomunicações e informática

João Paulo Barraca

Welcome to the lecture on defending an organization. Across the coming slides we will build a complete picture of what it means to protect a modern organization: the threat landscape it faces, the legal obligations that bind it, the frameworks used to structure security work, and the teams and technologies that carry out the day-to-day defence. Keep in mind that security here is not a single product you buy, but a continuous engineering and management discipline.

The central idea of this entire lecture is that defence is asymmetric. An attacker only needs to succeed once, and can choose the weakest point in the organization. A defender, in contrast, must protect everything, all the time, against every possible technique. This imbalance shapes every decision we will discuss, from which assets we protect first to how we decide what is worth investigating.

Throughout the presentation we will move from strategy to operations. We begin with the context in which organizations operate, both the physical world they have always inhabited and the cyberspace they have only recently entered. We then look at the laws and frameworks that force organizations to act, and finally we descend into the operational world of security teams, monitoring, detection engineering, and threat intelligence.

As you follow the lecture, try to keep two questions in your mind for every topic we cover. First, what is the organization actually trying to protect, and why does that particular asset matter to its mission? Second, given limited time, money and people, where should the organization spend its effort first? Most of the trade-offs we will discuss come down to answering those two questions well.

The current organizational landscape

- Organizations are complex and must reach everyone
- **Physical space:** where we live since >10000y
 - We know it, it's slow, it involves moving matter around
 - Laws are plentiful and cover most interactions
- **Cyberspace:** to which organizations just tapped into
 - We do not know it, it's fast, there are no barriers
 - Everything can be hidden, laws are limited

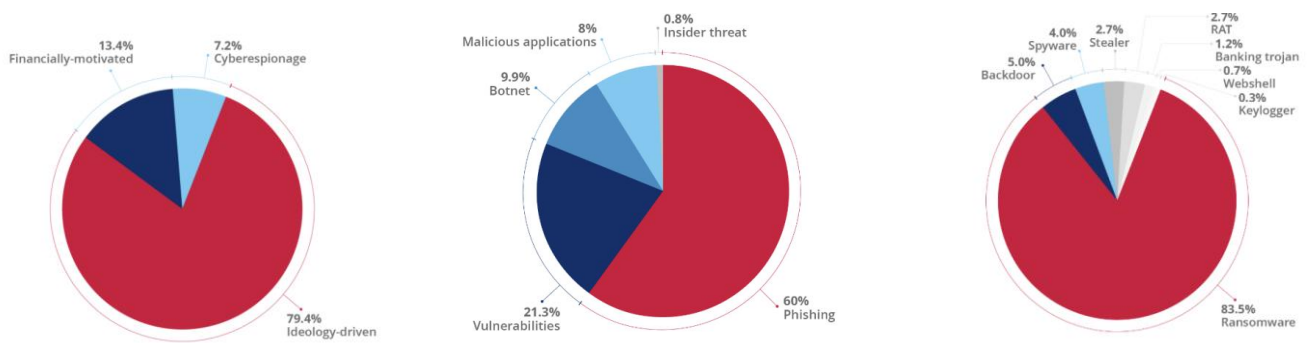
Organizations today live simultaneously in two very different worlds. The physical world has been home to human societies for many thousands of years, and we understand it well. It is slow, it requires physically moving materials and people around, and it is governed by laws that have had centuries to develop and cover most of the interactions we can imagine. In this world, boundaries are visible, property is tangible, and the rules are widely known and largely respected.

Cyberspace is different in almost every dimension. Organizations have only recently connected to it, and they are still learning how to behave there. It is extremely fast, information moves at the speed of light, and there are no natural barriers or borders. Everything can be hidden, copied, or altered without leaving a trace in the physical world. Crucially, the laws that govern this space are limited, fragmented across jurisdictions, and often lag far behind the technology they are meant to regulate.

This contrast explains many of the security problems organizations face. In the physical world, an attacker who breaks into a building leaves physical evidence, must travel to the scene, and can be caught by police. In cyberspace, an attacker in another country can compromise a server in seconds, steal data, and disappear, all while hiding behind anonymity tools and compromised machines. The defender must therefore assume that anything that can be reached over a network can be attacked, and design accordingly.

The practical lesson for students is that we cannot simply transplant physical-world assumptions into cyberspace. We cannot assume that something is safe because it is hard to reach, or that a boundary fence will work the way it does in the physical world. Instead, we must think in terms of identities, access, and data, and we must design defences that assume the network is hostile and that any single assumption may be wrong. This mindset, sometimes called zero trust, will appear again later in the lecture.

Malicious actors motivations and methods



To defend an organization effectively, we first need a clear picture of who attacks it and why. The motivations of malicious actors vary enormously, and understanding them helps us predict what they will do and how much effort they are willing to invest. Some attackers are financially motivated and treat crime as a business. Others are politically motivated, seeking to make a point or damage a target they disagree with. Still others are driven by ideology, curiosity, or simply the desire to prove their technical skill.

The European Union Agency for Cybersecurity, ENISA, publishes an annual threat landscape that aggregates the most pressing threats observed across Europe. This document is a useful starting point because it is based on real incident data reported by member states and organisations, and it ranks threats by their prevalence and impact. Reading it gives a security team a sense of what is actually happening, rather than what the media imagines is happening.

The methods actors use are equally diverse and evolve quickly. Ransomware remains one of the most damaging threats because it targets business continuity directly, encrypting data and demanding payment. Distributed denial of service attacks aim to take services offline. Supply chain attacks compromise trusted software so that many organizations are affected at once. Each method exploits a different weakness, and each requires a different defensive response.

As a student, you should learn to read a threat landscape critically. Look at which threats are rising and which are falling, which sectors are targeted most, and how the tactics change over time. This kind of intelligence, properly interpreted, is what allows a security team to prioritise its defences, focusing limited resources on the threats that are most likely to materialise against the organization it protects.

The current legal landscape (PT)

- Must comply with new regulatory frameworks
 - **2016:** NIS – Defines basic cybersecurity requirements
 - **2018:** GDPR – Defines requirements for private data
 - **2018:** RJSC – Legal Framework for the national Cyberspace
 - **2021:** DL65 – Defines processes for inventory, reporting, formalize strategy
 - **2025:** CRA – Defines mandatory security standard for hardware/software
 - **2025:** DORA - Digital Operational Resilience Act – Financial Institutions
 - **2026:** NIS 2 – Defines cyber teams and processes for critical/essential services
- Strategies are based on risk and maturity
 - Risk: identify assets and determine their risk
 - Maturity: determine organization maturity over multiple areas

Security is no longer a purely technical discipline; it is also a legal one. Over the last decade, a dense web of regulations has been created that forces organizations to take specific actions, whether or not they want to. Understanding this legal landscape is essential for any security professional, because failing to comply can result in fines, legal liability, and loss of the right to operate in certain markets.

The slide lists the key European and national instruments in chronological order. The Network and Information Security directive, usually abbreviated as NIS, was the first attempt to impose basic cybersecurity requirements on essential services. The General Data Protection Regulation, or GDPR, established strict rules for the processing of personal data, including duties to protect that data and to report breaches. Portugal then added its own national framework, the RJSC, and later issued a decree-law, DL65, which defines concrete processes for asset inventory, incident reporting, and the formalisation of a security strategy.

More recently the landscape has expanded further. The Cyber Resilience Act, or CRA, imposes mandatory security requirements on hardware and software products placed on the market, effectively moving security obligations into product design. DORA applies a similar logic to the financial sector, requiring digital operational resilience. And the revised NIS 2 directive, which supersedes the original NIS, strengthens obligations for a wider range of critical and essential entities, including requirements for dedicated cyber teams and processes.

Two concepts underpin how organizations actually comply with all of this regulation: risk and maturity. A risk-based strategy begins by identifying the organization's assets and assessing the risk each one faces, and then prioritises controls accordingly. A maturity-based strategy assesses how developed the organization is across a number of security areas, and works to raise that level over time. Most frameworks and regulators expect organizations to adopt one or both of these approaches, so you should be comfortable with the vocabulary of risk assessment and maturity models.

National Cybersecurity Framework (QNRCS)

Objectives



<https://www.cncs.gov.pt/pt/quadro-nacional/>

The National Cybersecurity Framework of Portugal, known by its acronym QNRCS, is the national translation of internationally recognised good practice into a practical structure that Portuguese organizations are expected to follow. It is maintained in coordination with the Portuguese cybersecurity authorities, and its details are published online for reference. Before diving into its content, it is worth understanding why national frameworks exist at all.

A national framework serves several purposes. It gives organizations a common vocabulary for talking about security, so that an auditor, a regulator, and a company can discuss the same concepts without confusion. It provides a checklist of areas that need attention, helping organizations avoid the common mistake of focusing on only one part of security while ignoring others. And it gives regulators a yardstick against which the security posture of an organization can be measured and compared.

The framework is designed to be risk-based and to consider all the stakeholders of an organization. It expects an organization to know its own assets, from its services and products to the supporting infrastructure, and to base its security decisions on the risks those assets face. It is not a set of prescriptive rules that tell you exactly what to buy, but rather a structured way of thinking about what to protect and why.

When studying the framework, remember that it is deliberately aligned with international standards, in particular the ones produced by ISO/IEC and the NIST Cybersecurity Framework. This means that the skills you learn here, and the terminology you pick up, will transfer to any other country or organization that follows the same international consensus. The specific national details may differ, but the underlying logic is shared across the world.

National Cybersecurity Framework (QNRCS)

Objectives

- **Identify:** Understanding the organization's context, the assets that support the critical business processes and relevant associated risks.
- **Protect:** Implementation of measures aimed at protecting the business processes and company assets, regardless of their technological nature.
- **Detect:** Definition and implementation of appropriate activities aimed at identifying incidents on time.
- **Respond:** Definition and implementation of appropriate measures in case of incident detection.
- **Recover:** Definition and implementation of activities aimed at managing the recovering plans and actions to restore impaired processes and services...

The QNRCS framework is organised around five high-level functions: Identify, Protect, Detect, Respond, and Recover. These functions are not sequential phases that you complete once and forget. Instead, they describe the ongoing capabilities an organization must maintain, and they should be seen as a continuous cycle that repeats throughout the life of the organization.

Identify is about understanding. The organization must know its context, its critical business processes, and the assets that support them, and it must be able to assess the risks those assets face. You cannot protect what you do not know you have, which is why asset inventory and risk assessment sit at the very foundation of any security programme. Many incidents become worse simply because the organization did not know an asset existed until it was attacked.

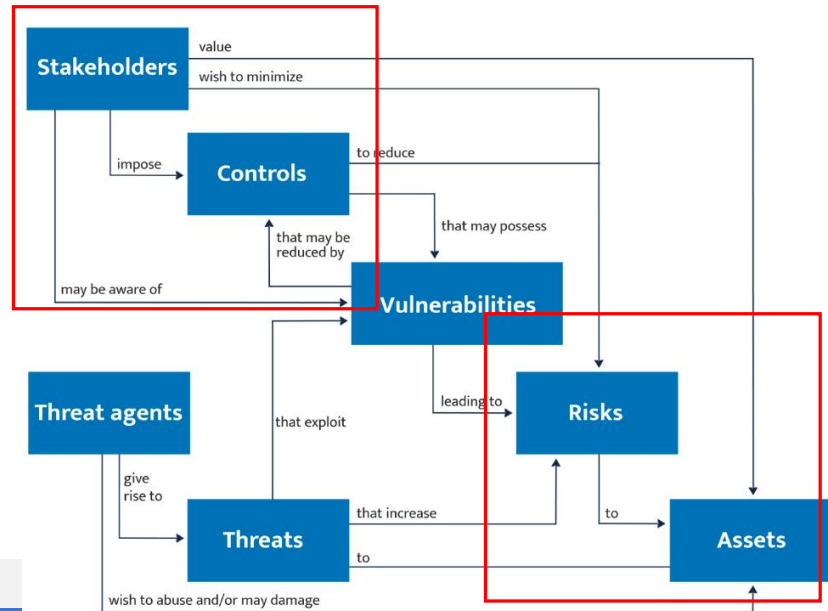
Protect is about putting safeguards in place. Once the organization knows what matters, it must implement measures that protect its business processes and assets, whatever their technological nature. These measures include technical controls such as access control and encryption, but also organisational measures such as policies, training, and incident procedures. The phrase whatever their technological nature is important, because it reminds us that protection extends to paper records, buildings, and people, not only to servers and software.

Detect, Respond, and Recover complete the picture. Detect requires the organization to identify incidents in a timely manner, which is the domain of monitoring and detection engineering. Respond covers the measures taken once an incident is detected, from containment to eradication. Recover is about restoring impaired processes and services, and returning the organization to normal operation. Together these five functions provide a complete model of what a defensive organization must be able to do, and later in the lecture we will see the specific teams and tools that implement several of them.

National Cybersecurity Framework (QNRCS)

ISO/IEC 27032, Basic concepts and high level relationships

- Risk Based
 - Aims to minimize risk
- Consider Stakeholders
 - Decision Level
- Consider Assets Inventory
 - Services
 - Products



João Paulo Barraca, André Zúquete

The QNRCS framework is grounded in international standards, and this slide points to one of them, ISO/IEC 27032, which describes basic concepts and high-level relationships in cybersecurity. Standards like this exist to give organizations a common reference point, and they are useful because they have been developed through consensus among experts and are recognised across borders. When a framework references a standard, it is essentially saying that the vocabulary and the relationships it uses come from an agreed international source.

The key idea on this slide is that the framework is risk based. It does not promise to eliminate all risk, which would be impossible, but instead aims to minimise risk to a level the organization considers acceptable. This is a pragmatic view of security: you cannot protect against everything, so you must decide where your effort is most valuable and accept the residual risk that remains.

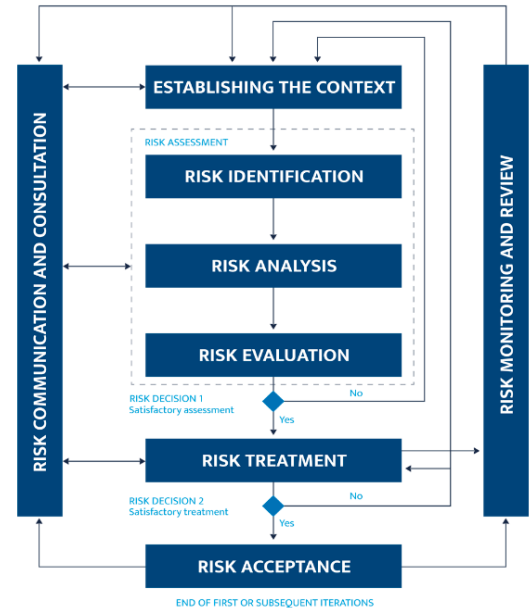
The framework also explicitly considers stakeholders and a decision level. Security decisions are not made only by technicians; they involve the people who own the business risk and who must accept the consequences of a decision. This is why the framework refers to a decision level, the point in the organization where risk is formally accepted or where investment is authorised. A security team can recommend, but someone accountable must decide.

Finally, the framework expects an organization to maintain an asset inventory that covers not just its services and products but also the underlying components. The reason is simple: risk assessment and prioritisation are impossible without knowing what exists. An organization that does not know what it has cannot decide what to protect first, cannot estimate the impact of losing something, and cannot justify its security budget. Inventory is therefore not an administrative chore; it is the foundation of a risk-based security programme.

National Cybersecurity Framework (QNRCS)

ISO/IEC 27005:2022, Basic concepts and high level relationships

- Strategy focused on **Risk Management**
- **Risk** used to decide what to address
 - Vulnerabilities to handle
 - Controls to deploy
 - Policies
 - Mechanisms to apply
 - Investment in cybersecurity



João Paulo Barraca, André Zúquete

This slide introduces the standard that governs the risk management methodology within the framework, ISO/IEC 27005. Where the previous standard gave us the vocabulary and the relationships, this one tells us how to actually reason about risk in a structured way. The current edition of the standard, published in 2022, is aligned with the latest versions of the related ISO/IEC 27001 and 27002 standards, which means the whole family of information security standards now shares a common structure.

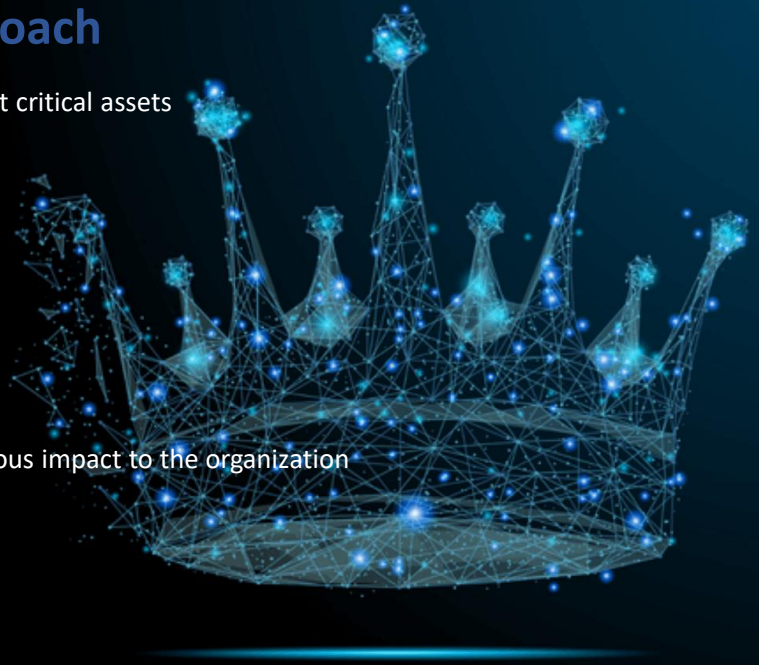
The central message of the slide is that the strategy is focused on risk management, and that risk is the criterion used to decide what to address. In practical terms this means the organization will not try to fix every possible weakness. Instead, it will identify the vulnerabilities that matter most, because they are exposed, because they are easy to exploit, or because the asset they affect is critical, and it will prioritise those.

Risk also drives two other categories of decision. It determines which controls to deploy, that is, which specific safeguards to implement. And it determines how much investment to make in cybersecurity. A useful way to think about this is that risk management is a resource allocation problem: you have a limited budget of money, time, and people, and you must allocate it so that the overall risk to the organization is minimised.

For students, the important skill to develop is being able to perform a risk assessment, even a simplified one. Identify an asset, identify the threats and vulnerabilities that could harm it, estimate the likelihood and the impact of an incident, and then propose controls that reduce either the likelihood or the impact. This is exactly the reasoning the framework expects an organization to be able to document and defend, and it is a skill that will serve you well in any security role.

Assets: Crown Jewels Approach

- Focused on identifying and protecting the most critical assets
 - To the organization mission!
- What is a **crown jewel**?
 - Essential Sensitive Data
 - Essential Servers
 - Essential Software Systems
 - Any other asset (HVAC, Generators...)
- Disruption to the crown jewels will pose a serious impact to the organization
- Objective: Protect the crown jewels
 - and grow from there to the rest of the organization
 - based on a risk assessment



Not all assets are equally important, and this slide introduces the Crown Jewels approach, a strategy that focuses security effort on the assets that matter most to the organization's mission. The idea is deliberately simple: identify the small set of assets whose loss would cause serious harm, protect them exceptionally well, and then build outward from there. It is a form of prioritisation that avoids spreading limited resources too thinly.

The slide lists what typically counts as a crown jewel. It includes essential sensitive data, the data the organization cannot afford to lose or leak. It includes essential servers and software systems, the systems that if they stopped working, the organization would stop working. And it also includes assets that are easy to forget but essential to operations, such as heating, ventilation and air conditioning systems, generators, and other physical infrastructure. An attacker who disables the cooling of a data centre can achieve the same effect as one who breaks into the servers, which is why these physical assets deserve the same protection.

The defining property of a crown jewel is that disruption to it poses a serious impact to the organization. This is what distinguishes it from an ordinary asset. A lost laptop is a problem, but a leaked database of customer records may be existential. The approach therefore asks the organization to be honest about which few assets, if lost, would truly threaten its survival or mission.

The recommended sequence is to protect the crown jewels first and then to grow outwards to the rest of the organization, but always on the basis of a risk assessment rather than guesswork. This is a pragmatic and widely used strategy because it acknowledges that you cannot protect everything equally. When resources are limited, the question is not whether to prioritise, but how to choose the right priorities, and the crown jewels approach is one disciplined answer to that question.

Security Plan

- Live document describing the security posture
 - Allows organizations to know where they are and where they want to go
 - Considers authentication, backups, risk, access control, policies, etc.
- Accepted by the organization, signed by Security Principal
 - Periodically reviewed and improved
- Written and accepted policies implies higher maturity
 - Organizations frequently only have word of mouth or informal frequent practices

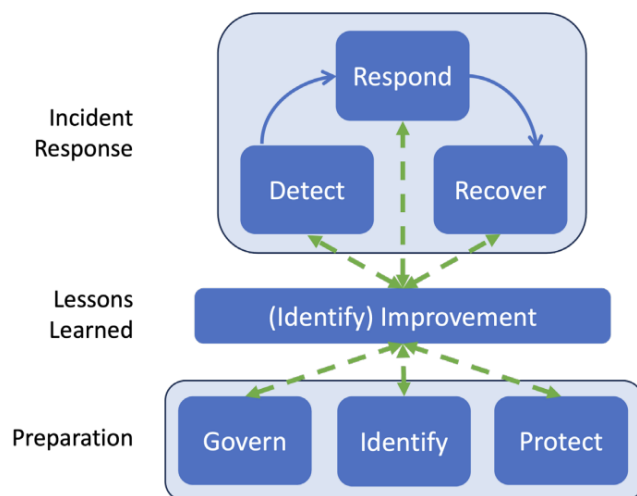
A security plan is the document that captures the organization's security posture in a structured form. It is not a set of slides or a one-off report, but a live document that is maintained over time. Its purpose is to give the organization a shared understanding of where it currently stands on security and where it intends to go, so that everyone, from technicians to executives, is working towards the same picture.

The plan is broad by design. It must consider authentication, backups, risk management, access control, and policies, among many other areas. Each of these is a thread that, woven together, defines how the organization protects itself. For example, the plan might state who is allowed to access which systems, how data is backed up and restored, and what the organization considers an acceptable level of risk. Because it covers so much, it is typically the document that an auditor or a regulator will ask to see first.

An important requirement is that the plan is accepted and signed by the security principal, the person who is ultimately accountable for security in the organization. This signature matters because it converts the plan from a recommendation into a commitment. It also forces accountability: if the plan says something will be done and it is not, there is a clear record of who accepted it. The plan is also periodically reviewed and improved, because threats, technology, and the organization itself all change over time.

The slide closes with an important observation about maturity. Organizations that have written and accepted policies demonstrate a higher level of maturity than those that rely on word of mouth or informal practices. The reason is that written policies are explicit, reviewable, and testable, whereas informal practices vary from person to person and disappear when people leave. If you are ever assessing an organization, one of the quickest signals of its security maturity is whether it has written, signed, and current policies, rather than an oral tradition of how things are done.

Incident Response



NIST SP 800-61r3 – Incident Response Life cycle
<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r3.pdf>

No organization can prevent every incident, so the ability to respond well when one occurs is a core capability. This slide introduces the incident response lifecycle, and the reference framework used here is the NIST Special Publication 800-61, a widely recognised guidance document on computer security incident handling. Understanding this lifecycle gives you a mental model of everything that has to happen, before, during, and after an attack.

The lifecycle is typically described in a series of phases. Preparation is the work done before any incident, such as writing playbooks, building a response team, and acquiring the tools you will need. Detection and analysis is where the incident is noticed and understood, which depends on the monitoring we will discuss later. Containment, eradication, and recovery are the phases where the organization stops the spread of the attack, removes the attacker's presence, and brings systems back to normal operation. Finally, post-incident activity is the review that captures lessons learned and improves future response.

Each of these phases deserves attention, but two are commonly neglected. Preparation is often undervalued because its benefits only appear during a crisis, yet a team that has rehearsed its response will perform immeasurably better than one improvising under pressure. And post-incident review is often skipped when the organization is relieved that the crisis is over, yet it is the phase that prevents the same mistake from recurring.

The legal and regulatory dimension is worth noting too. Modern regulations impose reporting deadlines, which means the detection and analysis phases must be fast enough to support them. An organization that takes weeks to notice an intrusion cannot meet a legal obligation to report within a matter of days. This is one more reason why the lifecycle is not a theoretical model, but a set of constraints that shape how the detection and monitoring functions must be designed.

- **FIRST: Forum of Incident Response and Security Teams**

- Global forum of incident response and security teams.
- Aim to improve cooperation between security teams on handling major cybersecurity incidents.
- FIRST is an association of incident response teams with **global coverage**.

- **ENISA: European Union Agency for Cybersecurity**

- Contributes to EU cyber policy, improving trust and resilience



Major incidents rarely respect the boundaries of a single organization, and this slide introduces the coordination structures that exist to handle them at national and international level. The first of these is FIRST, the Forum of Incident Response and Security Teams. FIRST is a global association of incident response teams whose aim is to improve cooperation when handling major cybersecurity incidents, and its membership spans the world, giving it genuine global coverage.

The value of an organisation like FIRST is that trust and working relationships are built before a crisis. When an incident strikes, there is no time to exchange business cards and establish credibility. Because the relevant teams already know each other and share common procedures and vocabulary, they can coordinate quickly and effectively. This principle, build the relationship before you need it, underlies most incident response coordination.

The second structure is ENISA, the European Union Agency for Cybersecurity. ENISA's role is to contribute to EU cyber policy and to improve trust and resilience across the Union. It does this by producing analyses and threat landscapes, supporting member states, and facilitating cooperation. It is important to understand, however, that ENISA is primarily a coordination and support agency rather than an enforcement body; enforcement of the various directives remains with the national authorities of each member state.

For a student, the lesson is that incident response is a team activity that extends well beyond the walls of any one organization. Knowing who to call, how to reach them, and what information to share with them is part of being prepared. This is why organisations document these contacts and relationships in their incident response plans, so that when a serious incident occurs, the right people can be brought together quickly rather than discovered under pressure.

- **CERT: Computer Emergency Response Team**
 - One per country, coordinating all significant events
 - Helps companies identifying, preparing and recovering from attacks
- **CSIRT: Computer Security Incident Response Team**
 - One per organization, coordinating the response in coordination with the CERT
 - <https://www.cncs.gov.pt/pt/certpt/>
- **CSIRT Networks:** Groups of CSIRTs to facilitate joint actions
 - E.g. training, taxonomy, Threat information exchange
 - <https://www.redecsirt.pt/>



Within any country there is a need for a single point of coordination for significant cybersecurity events, and in Portugal this role is fulfilled by the national CERT, the Computer Emergency Response Team. The slide uses CERT.PT as an example, and its function is to coordinate all significant events at a national level and to help companies identify, prepare for, and recover from attacks. When an incident is too large or too widespread for one organization to handle alone, the national CERT becomes the hub that connects everyone involved.

Alongside the national CERT, each organization typically maintains its own CSIRT, the Computer Security Incident Response Team. The CSIRT is the team within the organization that coordinates the response to incidents, working in coordination with the national CERT when the incident has broader implications. The relationship is a practical one: the organization handles its own incidents day to day, but escalates to the national level when the incident involves other organizations, requires legal involvement, or reveals a pattern that affects the wider community.

The slide also introduces the idea of CSIRT networks, which are groups of CSIRTs that come together to facilitate joint action. These networks organise training, share taxonomies and terminology, and exchange threat information. The Portuguese example given is the national CSIRT network, which provides a forum for teams across the country to cooperate. Such networks are powerful because they multiply the knowledge and capabilities of any single team.

There is a clear division of responsibility worth remembering. The CERT is national and coordinates significant events, while the CSIRT is organisational and coordinates the response within a single organization. Between the two there is a spectrum of coordination, from the individual organization to the national level, and an effective response depends on each level knowing its role and how to escalate. This is the practical machinery that turns the incident response lifecycle from a model into a functioning capability.

- **Support Activities**

- Networks, projects
- E.g. <https://www.ccc-centro.pt> (Competence Center)
- Increase the security posture and resilience of organizations
 - Training and awareness
 - Exchange strategies, information, and tools
 - Incident Response
 - Funding

- **Police Authorities**

- Polícia Judiciária
- Unidade Nacional de Combate ao Cibercrime e à Criminalidade Tecnológica (UNC3T): unc3t@pj.pt



Incident response does not happen in isolation; it is supported by a wider ecosystem of networks, projects, and institutions. The slide lists some of these support activities. One example is the Competence Center, which groups academic and research expertise to increase the security posture and resilience of organizations. Such centres contribute training and awareness, and they provide a channel for exchanging strategies, information, and tools between researchers and practitioners.

The existence of these support structures reflects a broader principle: no single organization has all the knowledge needed to defend itself. Threats evolve quickly, and the defensive techniques that work today may be obsolete tomorrow. By participating in networks and competence centres, an organization can learn from the experiences of others, access specialised expertise it could not afford to hire itself, and contribute its own lessons back to the community.

The slide also mentions the funding dimension and the involvement of police authorities. In Portugal, the relevant unit is the Polícia Judiciária's national unit for fighting cybercrime and technological criminality, abbreviated as UNC3T. This unit is the law enforcement contact for cybercrime matters, and its involvement is essential when an incident constitutes a crime, such as fraud, extortion, or unauthorised access. Knowing when and how to involve the police is itself part of incident response preparation.

There is an important practical lesson here about building relationships in advance. An organization that has never contacted the police until a crisis occurs will find the process slower and more uncertain than one that has established a working relationship and knows the correct channels. The same applies to the competence centres and networks. The preparation phase of the incident response lifecycle is therefore not only about internal playbooks, but also about knowing the external partners you will need before you need them.

Security Teaming

- Security operations are frequently organized in teams
 - Blue Team: Defends an organization from malicious actors
 - Red Team: Attacks an organization to help finding weak spots
 - Purple Team: Mixed attack defense role
- Each team uses specific tools and methods



Security operations in a large organization are frequently organised into distinct teams with complementary roles, and this slide introduces the three classic ones: the Blue Team, the Red Team, and the Purple Team. Understanding these roles is essential because they define how security work is divided, and each uses specific tools and methods tailored to its mission.

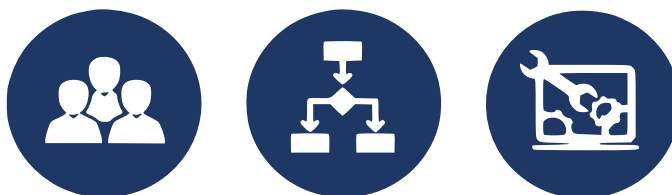
The Blue Team defends the organization against malicious actors. This is the team that operates the defences we will discuss throughout the rest of this lecture, monitoring systems, detecting and responding to incidents, and keeping the organization safe on a day-to-day basis. The Blue Team is essentially the operational face of the organization's security, and it is the primary focus of this lecture.

The Red Team plays the opposite role. It attacks the organization, but with the explicit purpose of helping it find weak spots. By simulating the behaviour of real attackers, the Red Team reveals vulnerabilities that the Blue Team and the organization did not know about, before a genuine attacker can exploit them. This adversarial testing is a powerful complement to defensive work because it sees the organization from the attacker's perspective.

The Purple Team sits between the two, combining attack and defence roles. In practice, a purple team is not so much a separate permanent team as a collaboration between blue and red, where the red team's findings are used to improve the blue team's defences. The value of this combination is that it closes the loop: attacks are not just reported and forgotten, but turned into concrete defensive improvements. Later in the lecture you will see specific techniques and tools that each of these teams would use.

Blue Teams

- Defend organizations from malicious actors
 - Abusing and Careless actors, and general failures also
- Typical fundamental tasks to address:
 - People: training, awareness, culture
 - Processes: analysis, investigation, data, reporting
 - Technology: monitoring, detection, scripting, automation



The Blue Team is the team responsible for defending the organization from malicious actors. Importantly, the slide clarifies that this defence is not limited to deliberate attackers. The Blue Team must also deal with careless insiders who make mistakes, and with general failures, such as misconfiguration, bugs, or human error, that open the organization to harm. In security, the most damaging incidents are often not sophisticated attacks but simple accidents amplified by a lack of preparation.

The slide organises the fundamental tasks of the Blue Team into three categories: people, processes, and technology. The people dimension is about training, awareness, and building a security culture, because many incidents begin with a human action, such as clicking a malicious link or sharing a password. No amount of technology will help if the people using the systems repeatedly make the same mistakes.

The processes dimension covers analysis, investigation, data handling, and reporting. When something suspicious happens, someone must investigate it, gather and preserve the evidence, and produce a clear account of what occurred. These processes must be defined in advance, because during an incident there is no time to invent them. The technology dimension is about the tools the team uses: monitoring, detection, scripting, and automation. Technology multiplies the limited human attention of the team, allowing it to watch far more than any person could.

These three dimensions are deeply connected, and a good Blue Team is strong in all of them. Technology without trained people is useless, and skilled people without defined processes are unpredictable. When you think about a career in defence, you should therefore consider not only the technical skills, but also the ability to write clear reports, design repeatable processes, and communicate with non-technical colleagues, because these are the skills that turn individual talent into an effective team.

Blue Teams

- **Mandatory** for all organizations!
 - Good amount of job opportunities
 - extreme shortage of professionals
- Very demanding due to **high asymmetry**
 - Attackers must succeed once, using their preferred TTPs
 - Defenders must defend continuously, from all attacks
 - To the entire organization attack surface, using any TTP
- Challenging and interesting
 - Many topics to address: prog, forensics, AI/ML, training...
 - Continuously evolving with new techniques and tools

This slide makes a strong claim: a Blue Team, or at least the defensive capability it represents, is mandatory for all organizations. It is not an optional extra for large companies, but a fundamental requirement for any organization that holds data, provides services, or depends on its digital systems. The reasoning is simple, if you are connected to the internet, you are being attacked, and someone must be responsible for defending you.

The slide also highlights a significant labour market reality. There is a good number of job opportunities in this field, driven by an extreme shortage of professionals. This shortage is a recurring theme in cybersecurity, and it means that the demand for qualified defenders consistently outstrips supply. For students considering a career, this is a field where skills are scarce and therefore valued. The same shortage, however, also means organizations struggle to find the people they need, which is one reason automation and tooling are so important.

The work is demanding because of a fundamental asymmetry that the slide spells out clearly. The attacker must succeed only once, and can use their preferred tactics, techniques and procedures, attacking the specific weakness they have chosen. The defender, in contrast, must defend continuously, against all attacks, across the entire attack surface of the organization, using any technique the attacker may choose. This is an inherently uneven contest, and understanding it is key to appreciating why defence is so hard.

Despite the difficulty, the field is challenging and interesting precisely because it is never static. It draws on many disciplines: programming, forensics, artificial intelligence and machine learning, and training and awareness, among others. And it evolves continuously as new techniques and tools appear. If you enjoy solving problems, learning continuously, and working under pressure, this is a field that will reward you for the rest of your career, and it will never become routine.

Blue Team Defence Techniques

- Everything Everywhere All at Once?
 - No! Prioritize according to the organization mission
- Current approaches focus on:
 - the CIA triad
 - the crown jewels
 - Risk assessment
 - with the least pain
 - security plan



A common mistake in security is to try to defend everything everywhere all at once, and this slide explicitly warns against it. The phrase everything everywhere all at once describes the instinct to cover every system, every user, and every technique equally. The slide rejects this instinct and instead argues that the Blue Team should prioritise according to the organization's mission. You cannot protect everything, so you must protect what matters most.

The slide lists the considerations that current approaches use to set priorities. The first is the CIA triad, which stands for confidentiality, integrity, and availability, the three properties that information security aims to preserve. The second is the crown jewels approach, which we discussed earlier, focusing on the assets whose loss would be most damaging. The third is risk assessment, which provides a systematic way to decide where threats are most likely and impact is greatest.

Two further considerations complete the picture. The first is choosing controls with the least pain, meaning the least disruption to legitimate users and business operations. A control that brings the business to a halt to stop a minor threat is usually a bad trade-off. The second is the security plan, which we discussed earlier, because it is the document that records the organization's chosen priorities and provides the reference point that keeps everyone aligned.

The practical message for the Blue Team is that prioritisation is not a one-off decision but a constant discipline. Every new control, every new monitoring rule, and every new tool should be evaluated against the organization's mission and its risk assessment. This is what turns a collection of security measures into a coherent defensive posture, and it is the difference between a team that is busy and a team that is effective.

SOC – Security Operations Center

- Responsible for continuously monitoring
 - Organization's digital infrastructure
- Monitor, detect and respond
 - To cybersecurity threats
- Empowered with skilled analysts and technology
 - Security assessments
 - Information security
 - Incident response

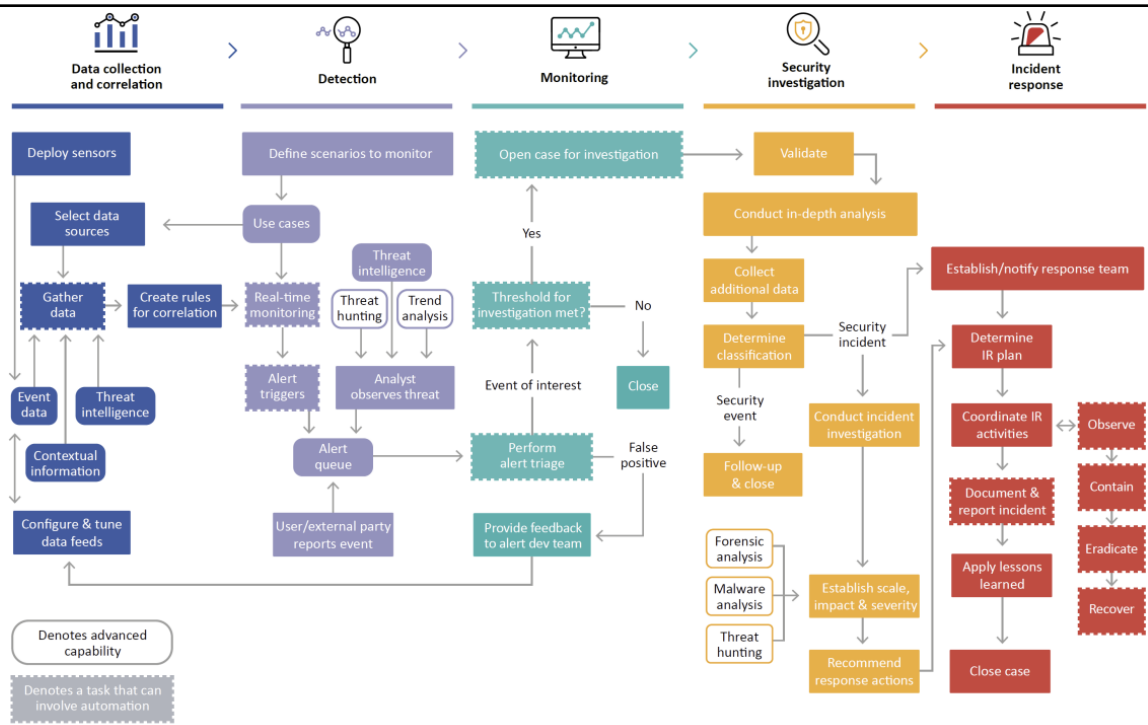


The Security Operations Center, universally abbreviated as SOC, is the operational heart of the organization's defence. This slide defines its role: the SOC is responsible for continuously monitoring the organization's digital infrastructure, and for monitoring, detecting, and responding to cybersecurity threats. The word continuously is crucial, because threats do not respect office hours, and an attack at three in the morning is just as dangerous as one at three in the afternoon.

A SOC is not just a room full of screens; it is an organisation of people, processes, and technology working together. The slide emphasises that the SOC is empowered with skilled analysts and technology. The analysts are the human judgement that interprets signals, decides what is worth investigating, and coordinates a response. The technology is the monitoring tools, detection rules, and automation that scale the analysts' reach across the entire infrastructure.

The slide lists some of the capabilities that a SOC typically encompasses. These include security assessments, information security, and incident response. In practice this means the SOC is involved in evaluating the security of systems before they are deployed, in managing the information that flows from monitoring, and in running the response when something goes wrong. The exact scope varies from organization to organization, but the core mission, continuous monitoring, detection, and response, remains constant.

There is an important point about the relationship between a SOC and the wider organization. The SOC is the operational layer, but it depends on the strategic layer we discussed earlier: the security plan, the risk assessment, and the priorities set by management. The SOC detects and responds within the framework those decisions create. When you study a SOC, you should therefore think of it as the execution arm of the organization's overall security strategy, rather than as a standalone tool.



This slide presents a visual representation of a Security Operations Center, and while the image carries most of the detail, it is worth reflecting on what the picture conveys about the nature of this work. A typical SOC is a room designed for sustained, high-attention monitoring, filled with analysts watching dashboards, screens, and alert queues. The environment is deliberately arranged to support long shifts of focused work, because the people in it are the organization's first line of detection.

The image also illustrates the scale of what a SOC must observe. Multiple screens are common, because a single analyst may need to watch network traffic, endpoint telemetry, authentication logs, and threat intelligence feeds at the same time. Each screen represents a different source of data that could reveal an attack, and the challenge is to notice the signal among the noise. This is exactly the monitoring and detection function we will examine in more detail shortly.

Behind the visible room lies the reality that a SOC is only as good as the data it receives and the processes it follows. The analysts need high-quality, well-structured logs from the organization's infrastructure, and they need clear procedures for what to do when they find something suspicious. Without good data, the SOC is blind; without good processes, even a talented analyst cannot respond consistently.

For students, the SOC is a useful lens through which to understand the rest of this lecture. The direction phase, which tells the SOC what threats to look for, the collection phase, which supplies the data, and the analysis and response phases, which turn data into action, all come together in the day-to-day work of a center like the one in this image. As we move through the technical material, keep the image of this room and its analysts in mind as the place where all the theory becomes operational.

Main concepts

- **Defensive Security (Engineering)**
 - Firewalls, backups, logs
 - Secure Software Development Lifecycle
 - Security related requirements (e.g., OWASP ASVS)
 - Training and Awareness
- **Incident Response**
 - Have processes and procedures to handle incidents
 - Involve stakeholders (Decision maker, Clients, Lawyers) and communicate (Public Relations)
- **Detection Engineering**
 - designing, developing, testing, and maintaining threat detection logic

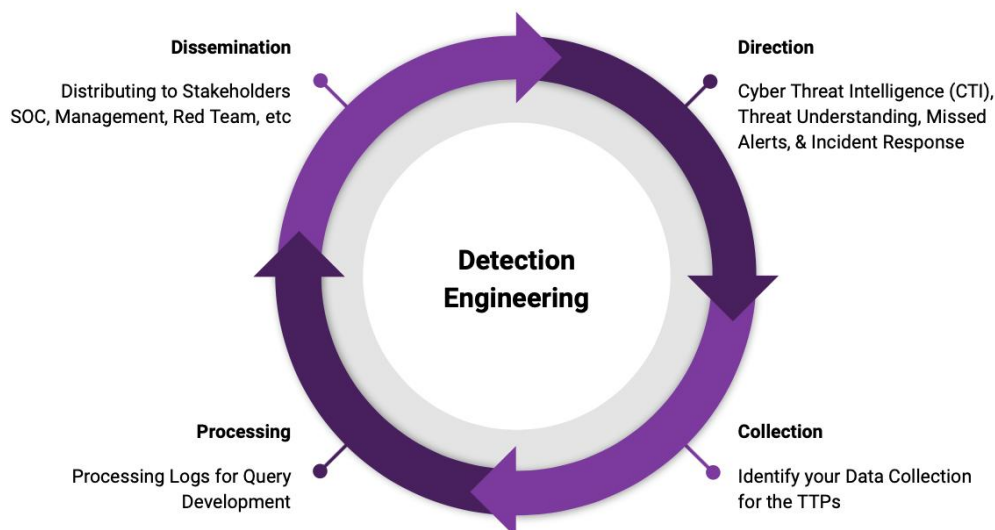
This slide gathers the main concepts that together form the practice of defensive security, and it is worth seeing them as a set rather than as isolated topics. The first group is defensive security engineering, which is the discipline of building security into the organization's systems from the start. It includes the classic technical controls such as firewalls, backups, and logs, but it extends well beyond them into how software is developed and how people are trained.

Secure software development is a central part of this picture. The idea is that security cannot be an afterthought bolted onto finished software, but must be considered in every phase of the development lifecycle, from requirements to design to testing and deployment. The slide mentions security-related requirements, citing OWASP ASVS as an example. ASVS, the Application Security Verification Standard, provides a detailed checklist of security requirements that a software product can be verified against, and it is a practical tool for making security concrete and testable.

The second concept is incident response, which we have already seen. The slide adds two important dimensions to it. The first is that incident response must involve stakeholders: the decision makers who can authorise expensive actions, clients who may be affected, and lawyers who handle legal and regulatory consequences. The second is communication, because how an organization communicates about an incident, with the public, with regulators, and with customers, can be as important as the technical response.

The third concept is detection engineering, which is the design, development, testing, and maintenance of threat detection logic. This is the discipline of deciding what to look for and how to express that as rules and models that can find malicious activity. Detection engineering is a relatively modern and specialised field, and it underpins much of what a SOC does. As the lecture progresses, we will look at each of these pillars in more depth, and you will see how they fit together into a coherent defensive capability.

Detection Engineering



Source: SANS

This slide introduces detection engineering, which is one of the most important modern disciplines in defensive security. The slide credits SANS as a source, reflecting that the terminology and framing come from well-established practitioner training. Detection engineering is the practice of designing, developing, testing, and maintaining the logic that identifies malicious activity, and it is the engine that turns raw data into the alerts a SOC acts upon.

At the centre of detection engineering is the concept of TTPs, which stands for Tactics, Techniques, and Procedures. This is a structured way of describing what an adversary does. The tactic is the goal, such as gaining initial access or achieving persistence. The technique is the general method used to achieve that goal, such as exploiting a vulnerability or abusing a legitimate tool. And the procedure is the specific, detailed way the technique is executed by a particular actor or tool. Understanding TTPs is fundamental because it moves detection away from chasing individual indicators and towards understanding behaviour.

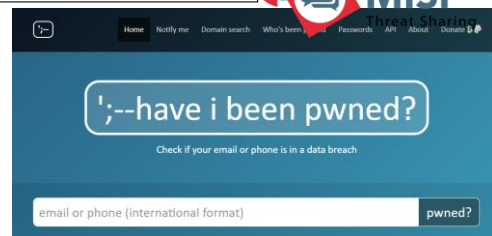
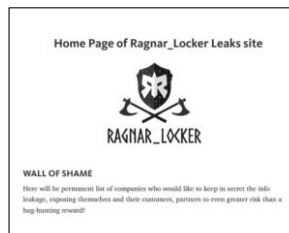
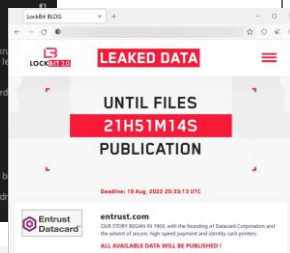
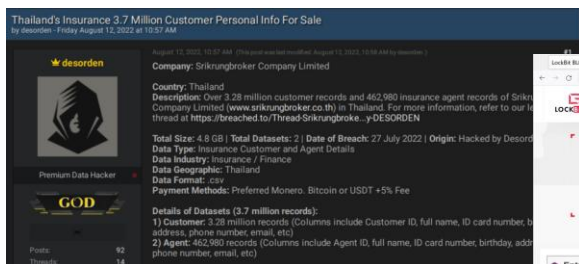
The reason TTPs matter is that attackers change their tools and indicators frequently, but their behaviour tends to repeat. A malware sample may be repacked and its hashes changed, but the way it moves across the network and the way it interacts with the operating system will often stay the same. Detection that is based on behaviour, on TTPs, therefore remains effective even when the specific artifacts change, whereas detection based only on static indicators becomes useless the moment the attacker changes them.

This is a crucial idea that will appear again when we discuss the Pyramid of Pain and the MITRE ATT&CK framework. Detection engineering is not about writing a single rule; it is about building a layered set of detections, at different levels of abstraction, that together make it hard for an attacker to operate without being noticed. The more the defender can detect based on behaviour and intent, the more durable the detection is against a changing adversary.

Direction: CTI

Assess the current threats from Cyber Threat Intelligence

- Cyber Threat Intelligence helps understanding the dynamics
 - The “Dark web”: Blogs & forums (hidden with Tor), TOX
 - Official reports: Security Researchers (Reversing, analysis)
 - How actors position themselves (hacktivists, crime)
 - How attacks to similar organizations are conducted



João Paulo Barraca, André Zúquete

SIO

23

The first function in a detection engineering pipeline is direction, and this slide introduces Cyber Threat Intelligence, abbreviated as CTI, as the source of that direction. Before the SOC can decide what to monitor, it must understand what threats it currently faces. Threat intelligence provides exactly that understanding by helping the organization assess the current threats relevant to it. This is the input that tells the detection team what to look for.

The slide describes some of the places where this intelligence is gathered, and it is worth being precise about what each represents. The dark web, the part of the internet reachable only through specialised clients such as the Tor network, hosts forums and markets where stolen data and tools are traded. Separately, there is the clearnet, the openly accessible internet, where criminals also operate through messaging applications, forums, and social media. Collecting from all of these sources, known collectively as open source intelligence, gives a picture of how actors position themselves and of their plans and capabilities.

The slide also notes that threat intelligence draws on official reports produced by security researchers. When researchers reverse engineer malware or analyse an intrusion, they publish reports describing how attacks are conducted. These reports are a rich source of information about the methods and infrastructure used against similar organizations. An intelligence analyst synthesises all of this raw material into something actionable for the organization.

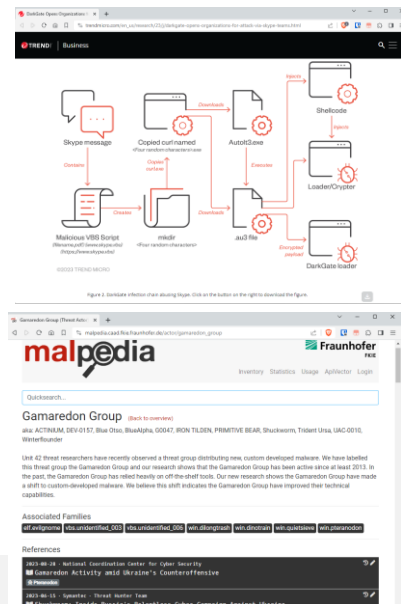
For a student, the key lesson is that direction is about answering the question: what should we be looking for? Without this input, a detection team would be guessing, chasing every possible threat with no sense of priority. CTI provides the focus that allows the organization to spend its limited detection resources where they matter most, on the threats that are actually relevant to it and to its sector.

Direction: CTI

- Threat Intelligence provides analysis and forecasts
 - Official entities, private orgs
 - Police Authorities
 - Government Ministries



João Paulo Ba



This slide continues the discussion of threat intelligence by describing what it produces. The key word is analysis and forecasts. Raw intelligence, such as a log of malicious domains or a report of a new malware family, is not itself useful until it has been analysed and turned into something that can guide decisions. The output of a good intelligence function is a clear picture of what is happening, what is likely to happen next, and what the organization should do about it.

The slide notes that threat intelligence is produced by a range of entities: official bodies, private organizations, police authorities, and government ministries. Each of these has a different vantage point and a different mandate. Police authorities, for example, see the criminal side of the threat landscape and can share information about active crime operations. Government ministries can provide a national perspective and coordinate responses to threats that affect critical infrastructure. Private intelligence firms can provide specialised analysis and cross-industry visibility.

Bringing these sources together is the real value of an intelligence function. No single source sees the whole picture. The police see the criminal ecosystem, the government sees the national picture, and the private sector sees what is happening inside specific industries. A good intelligence team fuses these different perspectives into a coherent assessment that is more than the sum of its parts.

It is also important to understand the limitations of threat intelligence. Intelligence is probabilistic, not certain. It provides forecasts based on the best available evidence, but forecasts can be wrong, and the future never arrives exactly as predicted. For the detection team, the practical value is that intelligence narrows the space of possibilities, telling it where to look first and what is most likely, even if it cannot tell it everything. This is what makes the direction phase effective: it converts the vast space of all possible threats into a manageable, prioritised set.

Direction: Alerts and Incidents

- Current alerts will tailor future rules
 - Identify popular threat actions
 - Reduce false positives
 - Keep the capability to detect new threats
 - Includes conducting controlled attacks to validate rules/controls
- Incident resolution impact resolution playbooks
 - Once a threat is found, what can the organization do?
 - Deficiencies in incident response define future improvements
 - Includes simulated incidents to test processes

This slide discusses the second source of direction for detection engineering: the organization's own alerts and incidents. While threat intelligence tells the team what threats exist in the wider world, the organization's own history tells it what is actually happening inside its network. These two sources together provide a much more grounded picture of what the detection team should focus on.

Current alerts tailor future rules in a very practical way. When the organization sees certain threat actions repeatedly, it learns that these are the actions that actually matter in its environment, and it can build and refine rules to catch them more reliably. The slide also makes the point that this process helps reduce false positives, because the team learns which detections produce noise that is not worth investigating, and it can tune or retire them.

The slide mentions conducting controlled attacks to validate rules. This is the practice of deliberately simulating an attack in a controlled way, sometimes called adversary emulation, to confirm that the detection rules actually fire when they should. This is important because a rule that has never been tested may fail at the worst possible moment. By running controlled tests, the team gains confidence that its detections work, and it discovers gaps before an attacker does.

The second half of the slide turns to incident resolution and its influence on detection. When a threat is found, the organization must decide what it can do, and the way incidents are resolved feeds back into the detection function. Deficiencies exposed during an incident, such as a rule that did not fire or a data source that was missing, become the basis for future improvements. Simulated incidents serve the same purpose: they test the organization's processes and reveal weaknesses in a safe environment, so that the response can be improved before a real crisis.

Collection: Data Harvesting

Engineer Data Collection

- Focus on relevant data sources to address threats
 - Cannot get all data
 - Visibility will be limited
- Potential targets
 - Servers: AD, email, HTTP, Databases
 - Wireless Controllers
 - VPN access
 - Firewalls
 - Endpoints: Laptops, VMs, IoT devices

The second major function in the detection pipeline is collection, and this slide addresses how data is harvested for detection purposes. The central engineering problem is that the organization cannot collect all data. Storage is finite, processing is expensive, and the volume of possible telemetry is effectively unlimited. The team must therefore engineer its data collection to focus on the data sources that are most relevant to the threats it cares about.

The slide lists some of the typical sources that matter. On the server side these include Active Directory, email systems, HTTP services, and databases, because these are the systems that attackers target and where evidence of compromise appears. Network infrastructure such as wireless controllers, VPN access points, and firewalls provides visibility into how users and attackers connect. Endpoints, including laptops, virtual machines, and IoT devices, are where malware executes and where much malicious behaviour ultimately occurs.

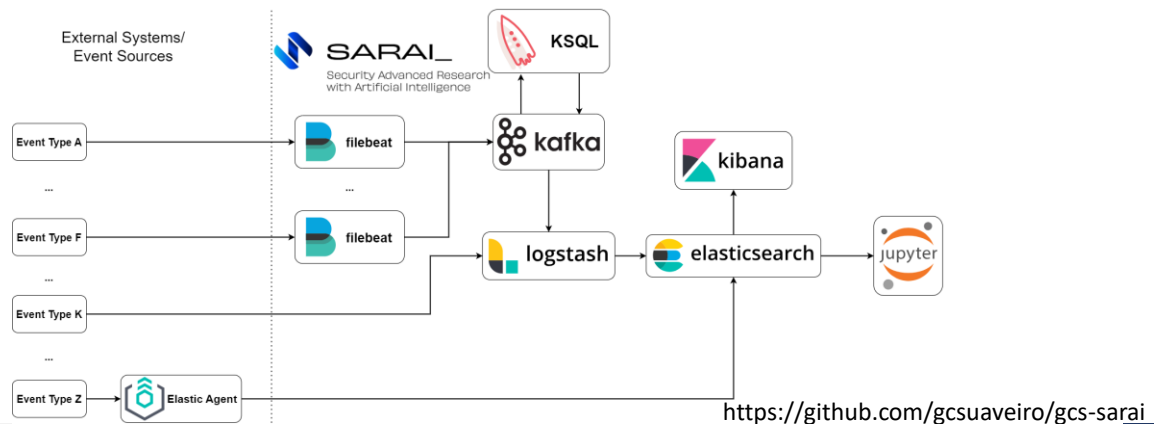
A key message is that visibility will be limited, and the organization must accept and plan around that limitation. You cannot see everything, and attackers will find ways to operate in the gaps you cannot see. This is not a failure of engineering but a fundamental constraint. The skill is in choosing where to spend your limited collection budget to maximise the value of what you can see, prioritising the sources that cover the highest-risk systems.

It is worth noting that the list of sources has grown over time. Modern environments also generate critical telemetry from cloud services, identity providers and authentication systems, DNS, web proxies, and endpoint detection agents. The slide's list is a starting point, and the engineering principle, choose the sources that matter most to your threats and your highest-risk assets, is what you should take away. Collection is a deliberate, prioritised design decision, not an attempt to record everything.

Collection: Data Harvesting

Engineer Data Collection

- Current approaches focus on a large data lake
 - Algorithms match rules, ML models, signatures, behavior



João Paulo Barraca, André Zúquete

SIO

27

This slide continues the discussion of data collection by describing how the collected data is organised, and it references an approach that has become common in many organizations: the data lake. The idea is to bring large volumes of security-relevant data together in a central, searchable repository, where it can be analysed by the detection logic. This is a significant engineering undertaking because of the sheer scale of the data involved.

The slide notes that the analysis is performed by algorithms that match rules, run machine learning models, compare signatures, and analyse behaviour. Each of these approaches has a different strength. Rules are precise and explainable: if a specific condition is met, an alert is raised. Machine learning models can find patterns that are too subtle for a human to define in advance, but they can be harder to explain and can produce false positives. Signatures are effective for known artifacts but useless for new ones. A mature detection capability uses a combination of all of these.

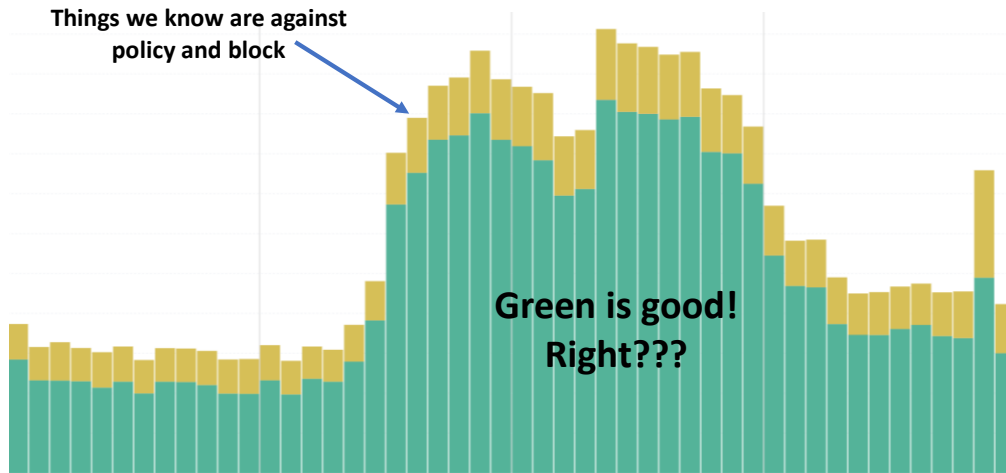
It is important to understand that a data lake is not an end in itself. Raw data sitting in a repository has no value until it is queried, and the value of the lake depends entirely on the quality of the data that is put into it and the quality of the analysis performed on it. This is why the engineering decisions discussed earlier, about which sources to collect and how to structure them, are so important.

The slide also provides a practical example with a link to an open source project. Studying such projects is an excellent way to learn how real data collection and analysis pipelines are built. Rather than expecting to design everything from scratch, you will often be working with and extending existing platforms, so understanding how they are organised and how their components fit together is a valuable skill for anyone entering this field.

Collection: Data Harvesting

Processing: Pain?

Millions of
events/hour



João Paulo Barraca, André Zúquete

SIO

28

This slide shifts attention to the processing stage and introduces the volume problem that every SOC faces. The slide makes a pointed observation: the organization may be dealing with millions of events per hour. This is not an exaggeration; large organizations routinely generate this volume from their endpoints, servers, and network devices. The first challenge of processing is simply coping with the scale.

The slide uses the phrase green is good, which refers to the way monitoring dashboards often colour compliant events. The instinctive reaction is to assume that green events are harmless and can be ignored. The slide challenges this assumption with an explicit question mark, asking whether that confidence is really justified. This is a warning that the apparent safety of the dashboard can hide real problems.

The slide also notes that the organization knows certain things are against policy and should be blocked. These are the events that the detection team is confident about, the ones with clear rules and clear intent. Even so, the volume of events that must be examined, the ones that are not obviously bad, remains enormous. The processing problem is to separate the small number of truly interesting events from the vast majority of noise.

This sets up a theme that runs through the next few slides: the sheer scale of the data means that no human can manually review it all. The processing stage therefore depends on automated filtering, correlation, and prioritisation to reduce the millions of events to the handful that deserve human attention. Understanding this volume problem is essential, because it shapes every decision about how detection is designed and how the SOC operates.

Collection: Data Harvesting

Processing: Pain?

Millions of
events/hour

Things we know are against
policy and block

Green is good!
Right???

Thousands of malicious agents (detect or block)

João Paulo Barraca, André Zúquete

SIO

29

This slide adds another layer to the processing problem introduced in the previous slide. In addition to the millions of events per hour, the organization must contend with thousands of malicious agents that it must either detect or block. These are the attackers, bots, and compromised systems that are actively probing or attacking the organization, and each of them is a potential source of a serious incident.

The slide is deliberately building a picture of how hostile and noisy the environment really is. The organization is not dealing with a single, well-defined attacker. It is dealing with a constant background of thousands of malicious agents, all of which must be handled somehow, whether by blocking them at a firewall, detecting them in the network, or identifying them on an endpoint. This is the everyday reality of operating an internet-connected organization.

The combination of millions of events and thousands of malicious agents illustrates why manual review is impossible. No team of analysts, however large, could manually examine millions of events per hour while also tracking thousands of attackers. The processing stage must therefore be highly automated, using rules, signatures, and models to filter, correlate, and prioritise before anything reaches a human.

This also explains why the quality of the detection logic matters so much. If the processing stage is poorly designed, it will either bury real attacks under noise or allow genuine threats to slip through. The goal is to find the right balance: automated processing that is sensitive enough to catch real threats, but precise enough not to overwhelm the analysts. This balance is one of the central challenges of the whole detection function.

Collection: Data Harvesting

Processing: Pain?

Millions of
events/hour

Things we know are against
policy and block

Green events are:

- Compliant events
- Malicious events that were not detected
- Malicious events that cannot be detected

Thousands of malicious agents (detect or block)

SIO

30

This slide completes the picture of the processing problem by examining what the green events, the ones the dashboard shows as compliant, actually are. The slide lists three categories, and the first is compliant events, the legitimate activity that truly is harmless and expected. These are the vast majority of events, and they are genuinely not a concern.

The second category is malicious events that were not detected. These are attacks that slipped through the detection logic and were therefore never flagged. They appear green only because the organization did not notice them. This is the most dangerous category, because it gives a false sense of security while an attacker may be operating undetected inside the network.

The third category is malicious events that cannot be detected. These are attacks that are designed to evade detection entirely, such as living off the land, where an attacker uses legitimate tools and services that blend in with normal activity. Some of these are genuinely undetectable with current methods, which is a sobering reality that any honest security team must accept.

The slide's final line, so many flows, captures the overwhelming scale of the problem, and the following slide will expand on the idea that activity is coming from everywhere, all around the organization. The lesson is that the dashboard's green colour is not proof of safety. It simply reflects what the organization knows and can see. The challenge of the processing stage is to be honest about how much of the green activity may actually be malicious and undetected, and to keep looking for it.



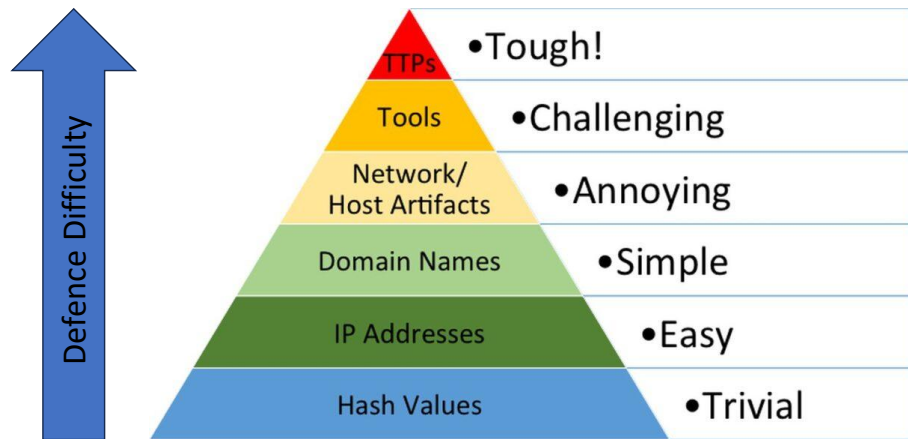
This slide makes a point that follows directly from the discussion of the volume of events and flows. The slide states, in large type, that there are so many flows coming from all around us. This is the visual representation of the scale of the problem: the organization is receiving network flows and telemetry from everywhere, from its own users, from partners, from the internet at large, and from the constant background of malicious agents we have been discussing.

The slide then makes an important conceptual point: the notion of us, meaning internal, versus others, meaning external, is not a robust distinction. In the traditional view, an organization drew a perimeter around itself and trusted everything inside it. But this view has broken down. Users connect from outside the office, applications run in the cloud, partners need access to internal systems, and attackers, once inside, behave just like legitimate users.

This is the reasoning behind the modern security architecture known as zero trust. Rather than assuming that anything inside the perimeter is trustworthy, zero trust assumes that no user or device is trusted by default, regardless of where it is located. Every request is verified, every session is validated, and access is granted on the basis of identity and policy rather than location. This is a fundamental shift from perimeter-based thinking.

For students, the practical implication is that you cannot build your mental model of a network around the idea of a safe inside and a hostile outside. The boundary is blurred, and threats can originate from any flow, internal or external. Defences must therefore be designed around identity, access, and continuous verification rather than around a single network border. This is a theme that runs through modern defensive security, and it is worth carrying this idea into every topic that follows.

The Pyramid of Pain



- Increase defence capabilities **from the bottom to the top**
- Detecting URLs/files/emails is trivial (also for attackers to change them)
- Understanding how actors behave is very difficult (also for attackers to change)

This slide introduces the Pyramid of Pain, a concept proposed by David Bianco in 2013 to help security teams think about which indicators are most valuable to detect and how difficult they are for an attacker to change. The pyramid is organised in layers, and the central insight is that the higher up the pyramid you can detect, the more pain you cause the attacker and the more durable your detection becomes.

The bottom of the pyramid contains the easiest indicators to obtain and detect: things like hashes of files, and indicators such as IP addresses, domain names, and email addresses. The slide correctly notes that detecting URLs, files, and emails is trivial, but it also makes the crucial point that these are equally trivial for the attacker to change. An attacker can re-hash a file, register a new domain, or rotate an IP address in minutes, which is why detecting at this level is easy but fragile.

Moving up the pyramid, the indicators become harder for the attacker to change. The higher layers involve the attacker's tools, their infrastructure, and ultimately their behaviour and intent, the tactics, techniques, and procedures they use. The slide notes that understanding how actors behave is very difficult, both for the defender to detect and for the attacker to change, because behaviour is deeply embedded in how an adversary operates. This is why behaviour-based detection is so valuable.

The practical lesson of the pyramid is to guide where the detection team spends its effort. While it is necessary to collect the easy indicators at the bottom, the organization should invest in detection at the higher levels, based on behaviour and tradecraft, because these are the indicators that the attacker cannot simply change overnight. The pyramid provides a useful vocabulary for explaining why some detections are worth more than others, and it connects directly to the earlier discussion of TTPs and detection engineering.

Triage

Or how to select relevant events?

- Could be one of several definitions
 - Attack near completion
 - Targeting / affecting high-value items
 - Critical hosts, business processes, users, data
 - Advanced targeted attackers or simple attacks
 - Unique, never fired before or lowest count
- Will depend on the organization



Once the detection pipeline has produced alerts, the next problem is triage: how to select which events are worth investigating. This slide introduces triage as the process of deciding which of the many alerts deserve the limited attention of the analysts. Because the volume of events is enormous and analysts are scarce, triage is one of the most important functions in the whole SOC, and getting it right determines the quality of everything that follows.

The slide lists several criteria that can be used to prioritise events. The first is whether the attack appears to be near completion, because an attacker who is close to achieving their goal must be stopped immediately. The second is whether the event targets or affects high-value items, such as critical hosts, business processes, users, or data. An alert affecting a crown jewel is inherently more important than one affecting a routine system.

The third criterion is whether the attack appears to come from an advanced, targeted attacker, because such attacks are rarer and often more dangerous than opportunistic ones. The fourth is whether the event is unique, never fired before, or has the lowest count. Unusual events are more likely to represent something genuinely new and interesting, whereas frequent events are more likely to be routine or expected.

The slide makes an important closing point: the exact definition of what is important will depend on the organization. There is no universal answer to what deserves attention. An alert that is critical for a bank may be irrelevant for a university. This is why triage policy is defined in the context of the organization's mission, its crown jewels, and its risk assessment, the same priorities we have discussed throughout, and why it must be periodically reviewed as the organization changes.

Definition of Dangerous

- Could be one of several definitions
 - Attack near completion
 - Targeting / affecting high-value items
 - Critical hosts, business processes, users, data
 - Advanced targeted attackers
 - Unique, never fired before or lowest count
- Will depend on the organization
- Anything that will cause relevant damage
 - It has a high cost to recover from
 - Or it is difficult to remedy



This slide addresses the question of what should be considered dangerous, and it is closely related to the discussion of triage. The slide lists criteria that are very similar to the ones used for triage, including attacks near completion, attacks targeting high-value items, advanced targeted attackers, and unique or never-before-seen events. This overlap is intentional: the same signals that make an event worth investigating are the signals that make it dangerous.

The slide then adds a more fundamental definition of danger. Anything that will cause relevant damage to the organization is dangerous. This is a broad and useful definition because it ties danger directly to impact. An event that causes no real harm, however technically interesting, is not dangerous, while an event that causes serious harm, however mundane, is.

The slide also introduces two additional criteria. The first is that an event is dangerous if it has a high cost to recover from. An attacker who encrypts the organization's backups, or steals data that must be disclosed, has caused damage that is expensive to undo. The second is that an event is dangerous if it is difficult to remedy, meaning the damage cannot easily be reversed. These two criteria capture the reality that the true measure of danger is the cost and difficulty of recovery.

Taken together, these definitions give the analyst a practical way to think about danger. It is not about how exotic an attack is, but about the harm it causes and the difficulty of recovery. This perspective helps the SOC prioritise consistently, because it provides a clear yardstick: does this event, if real, cause damage that is costly or difficult to remedy? If so, it is dangerous and deserves attention.

(Fantastic) Threats and Where to Find Them?

- Behavior matching: mostly ML
 - Known patterns
 - Anomaly detection
- File signature matching: **YARA**
 - Signatures for malware are created and disseminated
- Reputation evaluation: **IP addresses /domains**
 - Low reputation addresses may generate alert or block
- Known threats are identified by vendor software
 - Challenge: **Unknown**/Tailored threats

This slide moves from the question of what to look for to the question of how to find threats, and it introduces several detection techniques. The first is behaviour matching, which is largely based on machine learning. This approach builds a model of normal behaviour and then flags deviations, or looks for known patterns of malicious behaviour. It is powerful because it can find threats that have never been seen before, but it can also produce false positives when legitimate users behave unusually.

The second technique is file signature matching, and the slide cites YARA as an example. It is important to understand what YARA actually is. YARA is a language for defining patterns that can be used to identify and classify malware samples during tasks such as triage and hunting. You write rules that describe characteristics of a malicious file, and YARA matches those patterns against files or memory to flag samples of interest. This is distinct from network-level signature systems that block traffic in real time, and the distinction matters for choosing the right tool.

The third technique is reputation evaluation. IP addresses and domains can be scored by reputation, and addresses with a low reputation, meaning they have been associated with malicious activity, may generate an alert or be blocked. This is a useful signal, but it has weakened over time because attackers rotate addresses and domains quickly, so reputation should be treated as one input among several rather than a reliable pillar on its own.

The slide closes by noting that known threats are identified by vendor software, meaning the commercial products the organization deploys, such as antivirus and endpoint detection, carry signatures and telemetry feeds for known threats. The overarching challenge, stated at the bottom, is that unknown or tailored threats are much harder to find, because they do not match any known signature. This sets up the next slide, which explores what happens when we do not know whether something is malicious.

(Fantastic) Threats and Where to Find Them?

- What if we do not know if something is malicious?
 - What is a malicious website or file?
 - Most dangerous threats are not really classified as Malware
 - They have no well known signatures
- New malware potentially has high impact
 - It is not detected by Anti-virus
 - Explores unpatched vulnerabilities or flaws (0 day)
- A new malicious asset is just a **new program/website**
 - May be a variation of an existing malware
 - Different language/obfuscated/encrypted/packed
 - May simply bypass existing signatures
 - There is a robust market selling malware

This slide confronts the hardest question in detection: what do we do when we do not know if something is malicious? The slide notes that the most dangerous threats are often not classified as malware at all, because they have no well-known signatures. This is a profound challenge for any detection team that relies on matching against known indicators.

The slide explains why new malware is so dangerous. It is not detected by signature-based antivirus, because no signature exists for it yet. It may explore unpatched vulnerabilities or previously unknown flaws, sometimes called zero-days, because it can take advantage of weaknesses that no one has patched. And it can have a high impact precisely because it is new and unexpected, catching the organization unprepared.

The slide also clarifies what a new malicious asset actually looks like. From the defender's perspective, a new piece of malware is just a new program or a new website. It may be a variation of an existing malware, rewritten in a different language, obfuscated, encrypted, or packed to hide its true nature. It may simply bypass existing signatures because the signature was built for the original version. In other words, the artifact is new, but it may be built on familiar foundations.

The slide closes by noting that there is a robust market selling malware. In modern practice, this is less about buying a binary outright and more about malware as a service, where attackers subscribe to tools and infrastructure, or operate as affiliates who share in the profits of successful campaigns. This commercialisation of attack tooling means new malware is produced constantly and at scale, which is why the detection team cannot rely on signatures alone and must invest in behaviour-based detection, threat research, and the hunting techniques discussed in the following slides.

Threat Research

- Threat Research allows detection of **new offenses**
 - Takes a Indicators and determines its behavior
- Includes several knowledge areas
 - Open Source Intelligence
 - Social Networks, DNS/TLS Records, Dark Web
 - Reverse Engineering
 - Networking concepts
 - Network traffic analysis
 - Cryptography
 - Machine Learning

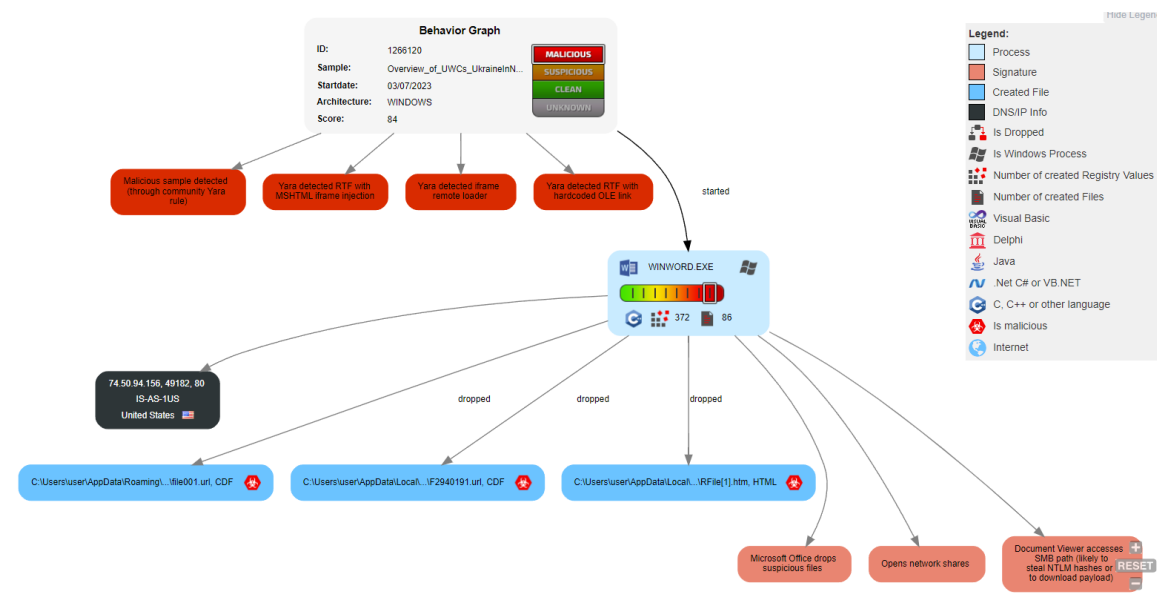
This slide introduces threat research, which is the discipline of understanding new and unknown threats well enough to detect them. The slide frames it as allowing the detection of new offenses, meaning new threats and attacks that have not been seen before. Where the previous slide described the problem of unknown malware, threat research is the answer: the work of figuring out what a new sample or a new behaviour actually does.

The slide states that threat research takes an indicator and determines its behaviour. This is the core activity: given a suspicious file, domain, or sample, the researcher analyses it to understand what it would do if it ran, how it communicates, and what its purpose is. This behavioural understanding is what allows a detection to be built for a threat that has no known signature, because the detection is based on behaviour rather than on a static indicator.

The slide lists some of the knowledge areas that threat research draws upon. Open source intelligence involves gathering information from public sources, including social networks, DNS and TLS records, and the dark web. Reverse engineering is the analysis of software to understand how it works. Networking concepts and network traffic analysis are needed to understand how malware communicates. Cryptography is relevant because malware often uses encryption to hide its activity, and machine learning provides techniques for detecting patterns at scale.

For students, the important point is that threat research is a multidisciplinary field. It is not a single tool or a single skill, but a combination of many, and a good threat researcher is comfortable across all of them. This is also a field that is constantly changing, because the threats being researched are themselves constantly evolving. If you find this area interesting, it is worth investing in the fundamentals of all of these disciplines, because they are the foundation on which research is built.

Threat Research: Execution Graphs



This slide introduces execution graphs, a technique used in threat research to understand how a piece of malware behaves, and it cites Joe Sandbox as an example of the tools that produce them. An execution graph is a way of visualising the sequence of actions that a sample performs when it is executed in a controlled environment. By running the malware in a sandbox and observing what it does, the researcher can trace its behaviour from start to finish.

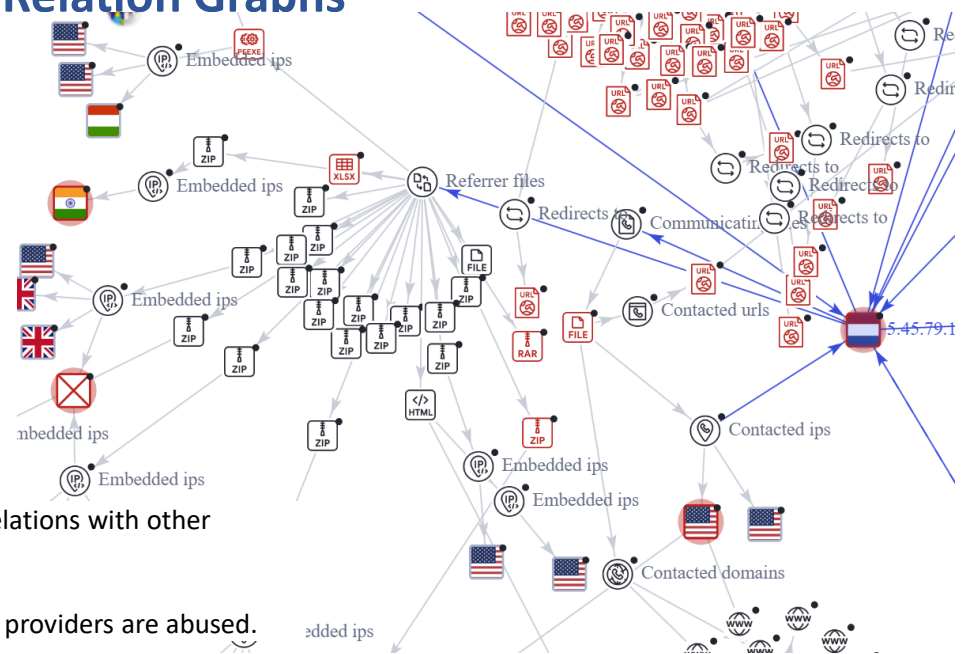
The value of an execution graph is that it converts an opaque binary into a readable story. Instead of staring at disassembled machine code, the analyst sees a flow of actions: which files are created, which processes are spawned, which network connections are made, and which registry keys or system settings are touched. This behavioural picture is far more informative than a static scan, because it reveals what the malware is actually trying to do.

The slide also raises an important operational security lesson, though it is implied by the choice of a cloud analysis service. When you upload a suspicious sample to a third-party analysis platform, you are disclosing that sample to the vendor and, potentially, to their partners. For samples from a public threat, this is usually acceptable, but for a sample from an internal incident, it can be risky, because you are telling a third party about your own compromise. Organizations must therefore be deliberate about when they use external analysis services.

For students, execution graphs are a useful window into the mindset of the malware author. By understanding the sequence of actions, you can infer the author's goals, such as establishing persistence, stealing credentials, or communicating with a command and control server. This understanding is exactly what feeds the detection engineering we discussed earlier: once you know what the malware does, you can build detections that look for that behaviour.

Threat Research: Relation Graphs

Virus Total



Some become suspicious due to relations with other malicious assets.

CDNs, Shared Hosting and Popular providers are abused.

João Paulo Barraca, André Zúquete

SIO

39

This slide introduces relation graphs, another technique used in threat research, and it cites VirusTotal as an example of the platforms that generate them. Where execution graphs show what a single sample does, relation graphs show how samples and indicators relate to each other. The idea is that some artifacts become suspicious because of their relations to other malicious assets, even if the artifact itself has no known signature.

The slide notes that certain assets become suspicious because they share infrastructure or characteristics with known-malicious ones. For example, a domain that resolves to the same infrastructure as a known malicious domain may be flagged, or a file that was submitted by the same actor as a known malware sample may be considered suspicious. This relational reasoning is powerful because it extends the reach of detection beyond the samples that are directly known.

The slide also includes an important warning. CDNs, shared hosting, and popular providers are frequently abused by attackers, but they are also used by perfectly legitimate services. This means that relational evidence, such as sharing an IP with a malicious asset, is a weak signal on its own. Many legitimate assets will be flagged simply because they happen to share infrastructure with something malicious, producing false positives. The analyst must therefore weigh relational evidence carefully rather than treating it as conclusive.

The practical lesson for students is that threat research is about connecting dots, but not every connection is meaningful. Understanding the strength and reliability of different kinds of evidence, and knowing when a relation is genuinely suspicious rather than just coincidental, is a key analytical skill. This is what separates a well-founded assessment from an alarm that turns out to be a false lead, and it is a skill that improves with practice and experience.

MITRE ATT&CK Matrix

- A globally-accessible knowledge base of adversary tactics and techniques
 - based on real-world observations.
 - A Matrix of techniques for each activity
- Allows organizations to map actions to an identified kill chain
 - Also facilitates tracking the Actor or how it evolves
 - Works because actors will reuse tools, tactics and techniques

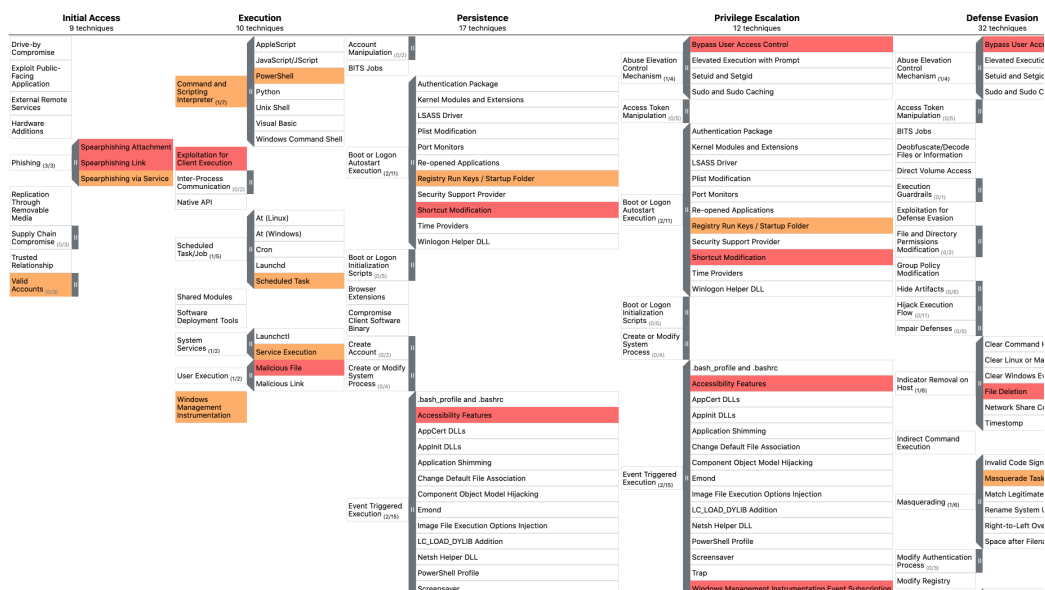
This slide introduces the MITRE ATT&CK framework, one of the most important reference tools in modern defensive security. ATT&CK is a globally accessible knowledge base of adversary tactics and techniques, built from real-world observations. The phrase from real-world observations is key: it is not a theoretical model, but a catalogue of what actual attackers actually do, derived from studying real intrusions.

The framework is organised as a matrix of techniques for each activity. It lists tactics, which are the goals an adversary pursues, such as initial access, execution, persistence, and exfiltration, and under each tactic it lists the techniques that can be used to achieve it. This structure gives the defender a comprehensive map of the adversary's playbook, and it is continuously updated as new techniques are observed.

The slide notes that ATT&CK allows organizations to map their observed actions onto a structured framework, and that it also facilitates tracking an actor and how it evolves over time. Because the framework is based on techniques, an organization can record which techniques a particular actor has used, and then watch how that actor's behaviour changes across campaigns. This is valuable for attribution and for anticipating what an actor might do next.

The slide also explains why the framework works: actors reuse their tools, tactics, and techniques. Even sophisticated attackers tend to repeat the methods they know well, which is why a structured catalogue of techniques is so useful. The final slide in this presentation shows the matrix itself. For the detection team, ATT&CK is a shared language that connects threat intelligence, detection engineering, and incident response, and it is worth learning to navigate it fluently.

MITRE Att&ck Matrix



João Paulo Barraca, André Zúquete

SIO

41

This slide displays the MITRE ATT&CK matrix itself, the actual visualisation of the framework introduced in the previous discussion. Seeing the matrix is the best way to appreciate its structure and its scale. It is organised as columns of tactics, the goals of an adversary, and within each column, rows of techniques, the methods used to achieve those goals. The result is a comprehensive map of adversarial behaviour.

The matrix is not a linear kill chain but a two-dimensional structure. This distinction matters. A kill chain describes a sequence of steps, whereas ATT&CK describes the full space of possibilities at each stage. An attack may not follow the matrix in order, and an adversary may jump between tactics, but the matrix still provides a way to describe and reason about any observed behaviour. This is why it is so widely used for detection engineering and for red team planning.

For the detection team, the matrix serves as a checklist and a language. It helps identify coverage gaps, because the team can ask which techniques it currently detects and which it does not. It helps communicate, because referring to a specific technique by its identifier, such as the one for a particular form of credential access, is unambiguous across teams and tools. This shared vocabulary is one of the framework's greatest strengths.

The matrix is also regularly updated as new techniques are observed in the real world, so you should treat it as a living resource rather than a fixed document. There are related tools and additional matrices for other platforms, such as industrial control systems and mobile devices, which extend the same approach to other domains. Learning to work with the ATT&CK matrix, and to think in terms of tactics and techniques, is a foundational skill for anyone entering the defensive security field.