

AnnoCTR: Cyber Threat Reports

Domain-Specific Annotation guidelines

Lukas Lange, Marc Mueller, Patrick Grau, Dragan Milchevski, Ghazaleh Torbati, Annemarie Friedrich
Bosch Center for Artificial Intelligence, Renningen, Germany

This document describes the domain-specific annotation guidelines used for the manual annotation of Cyber Threat Intelligence Reports. In addition, the documents have been marked with standard location (LOC), ORG (organization), and TIMEX (temporal expression) information.

Cyber-security specific concepts, tactics and techniques are marked and linked to the publicly available [MITRE ATT&CK](https://attack.mitre.org/) framework. In the following, we describe each of the tags and how they are linked.

GROUP

Annotate any APT (Advanced Persistent Threat = organized, sometimes state-sponsored hacker group) that is mentioned as GROUP. Set the `cysec_identifier` field to the Group that is mentioned and set the value on GROUP.

If the group cannot be found in the database, first search for synonyms/other common names of the group and if they are also not found, create a new instance for it in the database.

In the example below, APT40 and Leviathan are different names for the same group (see <https://attack.mitre.org/groups/G0065/>).

The screenshot displays the AnnoCTR interface. On the left, a text document is shown with line numbers 27 to 31. Annotations are visible as colored boxes and labels above the text. For example, line 28 has a label '[Technology | -org_or_sector_targeted | ORG]' above a URL. Line 30 has labels '[https://attack.mitre.org/groups/G0065 | -implicit | GROUP]' and '[https://attack.mitre.org/groups/G0065 | -implicit | GROUP]' above the text 'APT 40' and 'Leviathan' respectively. Line 31 has a label '[https://attack.mitre.org/groups/G0065 | -implicit | GROUP]' above the text 'APT 40'. On the right, a sidebar titled 'Cysec Named Entity' contains fields for 'Text' (APT 40), 'cysec_identifier' (empty), 'cysec_identifier_manual' (https://attack.mitre.org/grou), 'description' (empty), 'implicit' (Yes/No buttons, with 'No' selected), and 'value' (GROUP).

Software: MALWARE, TOOL, CONCEPT

Definition: *Software* is a generic term for custom or commercial code, operating system utilities, open-source software, or other tools. We distinguish two types of software in our annotation: MALWARE, which refers to software that has been written specifically for malicious purposes, and TOOLS, which refers to software not written for a malicious purpose but used with a malicious intent in a given context. Non-malicious software that is not used maliciously in a context is tagged as CONCEPT.

MALWARE

MALWARE are all types of malicious Software that are written to be malicious, for example any kind of ransomware or RATs (Remote Access Trojans) etc.

MITRE ATT&CK description: “Commercial, custom closed source, or open source software intended to be used for malicious purposes by adversaries. Examples include PlugX, CHOPSTICK, etc.”

66	![Another obfuscated script](/cdn-cgi/image/format=auto/sites/default/files/images/blogs/Trickbot%20Refresher/obsc2.png)
67	Image 6: Another obfuscated script
68	While in most cases, the final scripts responsible for downloading the https://attack.mitre.org/software/S0183 -implicit MALWARE TrickBot payload were identical and not obfuscated, in one case the script was slightly obfuscated:
69	![Decrypted downloader with slight obfuscation](/cdn-cgi/image/format=auto/sites/default/files/images/blogs/Trickbot%20Refresher/obsc_payload_dwonloader.png)
70	Image 7: Decrypted downloader with slight obfuscation

Cysec Named Entity

Text [+](#)

TrickBot

No links or relations connect to this annotation.

cysec_identifier

cysec_identifier_manual

<https://attack.mitre.org/>

description

implicit

Yes No

value

MALWARE

84	Cobalt Strike has been observed in a variety of attack chains alongside malware such as https://attack.mitre.org/software/S0266 -implicit MALWARE The Trick, https://attack.mitre.org/software/S0534 -implicit MALWARE BazaLoader, https://attack.mitre.org/software/S0386 -implicit TECHNIQUE Ursnif, https://attack.mitre.org/software/S0483 -implicit MALWARE IcedID, and many more popular loaders.
85	In these cases, the https://attack.mitre.org/tactics/TA0011 +implicit IMPL TECHNIQUE https://attack.mitre.org/software/S0154 -implicit MALWARE https://www.wikidata.org/wiki/Q29364585 preceding malware typically loads and executes Cobalt Strike
86	Likewise, there is a wide array of techniques leveraged in cases where https://attack.mitre.org/software/S0154 -implicit MALWARE https://www.wikidata.org/wiki/Q29364585 Cobalt Strike is delivered directly, such as https://attack.mitre.org/techniques/T1566/001 -implicit TECHNIQUE https://attack.mitre.org/techniques/T1059/005 +implicit IMPL TECHNIQUE via malicious macros in weaponized Office documents, compressed

Cysec Named Entity

Text [+](#)

BazaLoader

No links or relations connect to this annotation.

cysec_identifier

cysec_identifier_manual

<https://attack.mitre.org/>

description

implicit

Yes No

value

MALWARE

wiki_identifier

TOOL

TOOLS are all types of normal legit tools or pen test tools, which get abused by the attacker to do malicious actions. This category includes both software that generally is not found on an enterprise system as well as software generally available as part of an operating system that is already present in an environment. Note: We annotate this tag (TOOL) only if the software has a malicious use in a given context. If a software is mentioned and used in non-malicious way, we use the tag **CONCEPT**.

Examples: Netcat, wget, PsExec but also Mimikatz, Metasploit, PsExec...

160	execute the ransomware in bulk. This can be done in a variety of ways, ranging from manually starting the ransomware on the targeted machines , scheduling a task per machine , or using [PsExec](https://docs.microsoft.com/en-us/sysinternals/downloads/psexec) to launch the ransomware . ### Linking Groove to Babuk and BlackMatter As discussed above, there was a fallout within	<div>https://attack.mitre.org/techniques/T1486 -implicit TECHNIQUE https://www.wikidata.org/wiki/Q926331</div> <div>https://attack.mitre.org/techniques/T1486 +implicit IMPL_TECHNIQUE</div> <div>https://attack.mitre.org/techniques/T1486 -implicit TECHNIQUE https://www.wikidata.org/wiki/Q926331</div> <div>https://attack.mitre.org/techniques/T1053 -implicit TECHNIQUE</div> <div>https://attack.mitre.org/software/S0029 -implicit TOOL</div> <div>https://attack.mitre.org/techniques/T1486 -implicit IMPL_TECHNIQUE</div> <div>https://attack.mitre.org/techniques/T1486 -implicit TECHNIQUE https://www.wikidata.org/wiki/Q926331</div> <div>http://www.bosch.com/node1fsmh11c3x21 -implicit GROUP</div> <div>https://attack.mitre.org/software/S0638 -implicit MALWARE</div> <div>https://attack.mitre.org/software/S0638 -implicit MALWARE</div>	<div>Cysec Named Entity</div> <div>Text</div> <div>PsExec</div> <div>No links or relations connect to this annotation.</div> <div>cysec_identifier</div> <div>cysec_identifier_manual</div> <div>https://attack.mitre.org</div> <div>description</div> <div>implicit</div> <div>Yes No</div> <div>value</div> <div>TOOL</div>
181	...		
182	...		
183	From observing the deobfuscated payload strings, we can assess with high confidence that TerraStealer functions as an information stealer.	-implicit CON https://www.wikidata.org/wiki/Q55106975	
184	TerraStealer targets a list of web browsers, e-mail clients, and FTP clients, including popular products such as Google Chrome, Firefox.	Technology +org_or_sector_targeted ORG Google	

Concept (CON)

Annotations of this tag cover mentions of concepts that are relevant to cyber-security but that are not specific to MITRE ATT&CK. They are more generally known and occur in WikiData. They include non-malicious computer software, protocols, hash algorithms or encryption algorithms or other concepts from the cyber security/IT domain. Mentions of CON are linked to WikiData. Set value to CON and enter the wiki_identifier (and cysec_identifier in addition if suitable).

Examples: RemoteAccessTool/Trojan (RAT), Bitcoin, MD5, payload, Ransomware as a Service (RaaS), Discord

181	...		
182	...		
183	From observing the deobfuscated payload strings, we can assess with high confidence that TerraStealer functions as an information stealer.	-implicit CON https://www.wikidata.org/wiki/Q55106975	
184	TerraStealer targets a list of web browsers, e-mail clients, and FTP clients, including popular products such as Google Chrome, Firefox.	Technology +org_or_sector_targeted ORG Google	

value

CON

wiki_identifier

wiki_identifier_manual

https://www.wikidata.org

Explicit vs. Implicit Tactics and Techniques

We use the [MITRE ATT&CK framework](#) to classify mentions of tactics and techniques, i.e., either when they are mentioned explicitly or when a particular action is described implicitly where an attacker made use of a *tactic* or *technique*.

- **Tactics** represent the „why“ of a technique. The adversary’s tactical goal: the reason for performing an action. For example, an adversary may want to achieve credential access.
- Techniques are the technical ways to achieve a tactical goal. For example, if the tactical goal is credential access, the technique to achieve this may be to use the technique *Adversary-in-the-Middle*.

Reconnaissance	Resource Development	Initial Access	Execution	Persistence	Privilege Escalation	Defense Evasion	Credential Access	Discovery	Lateral Movement	Collection	Command and Control
10 techniques	7 techniques	9 techniques	12 techniques	19 techniques	13 techniques	40 techniques	15 techniques	29 techniques	9 techniques	17 techniques	16 techniques
Active Scanning (2) Gather Victim Host Information (4) Gather Victim Identity Information (3) Gather Victim Network Information (6) Gather Victim Org Information (4) Phishing for Information (3) Search Closed Sources (2) Search Open Technical Databases (5) Search Open Websites/Domains (2) Search Victim-Owned Websites	Acquire Infrastructure (6) Compromise Accounts (2) Compromise Infrastructure (6) Develop Capabilities (4) Establish Accounts (2) Obtain Capabilities (6) Stage Capabilities (5) Valid Accounts (4)	Drive-by Compromise Exploit Public-Facing Application External Remote Services Hardware Additions Phishing (3) Replication Through Removable Media Supply Chain Compromise (3) Trusted Relationship Valid Accounts (4)	Command and Scripting Interpreter (8) Container Administration Command Deploy Container Exploitation for Client Execution Inter-Process Communication (2) Native API Scheduled Task/Job (6) Shared Modules Software Deployment Tools System Services (2) User Execution (3) Windows Management Instrumentation	Account Manipulation (4) BITS Jobs Boot or Logon Autostart Execution (15) Boot or Logon Initialization Scripts (5) Browser Extensions Compromise Client Software Binary Create Account (3) Create or Modify System Process (4) Event Triggered Execution (15) Event Triggered Execution (15) External Remote Services Hijack Execution Flow (11) Implant Internal Image Modify Authentication Process (4)	Abuse Elevation Control Mechanism (4) Access Token Manipulation (5) Boot or Logon Autostart Execution (15) Boot or Logon Initialization Scripts (5) Create or Modify System Process (4) Domain Policy Modification (2) Escape to Host Event Triggered Execution (15) Exploitation for Privilege Escalation Hijack Execution Flow (11) Process Injection (11) Scheduled Task/Job (6) Valid Accounts	Abuse Elevation Control Mechanism (4) Access Token Manipulation (5) BITS Jobs Build Image on Host Deobfuscate/Decode Files or Information Deploy Container Direct Volume Access Domain Policy Modification (2) Execution Guardrails (1) Exploitation for Defense Evasion File and Directory Permissions Modification (2) Hide Artifacts (9) Hijack Execution Flow (11) Impair Defenses (9) Indicator Removal on Host (6) Indirect Command Execution Valid Accounts	Adversary-in-the-Middle (2) Brute Force (4) Credentials from Password Stores (5) Exploitation for Credential Access Forced Authentication Forge Web Credentials (2) Input Capture (4) Modify Authentication Process (4) Network Sniffing OS Credential Dumping (6) Steal Application Access Token Steal or Forge Kerberos Tickets (4) Steal Web Session Cookie Two-Factor Authentication	Account Discovery (4) Application Window Discovery Browser Bookmark Discovery Cloud Infrastructure Discovery Cloud Service Dashboard Cloud Service Discovery Cloud Storage Object Discovery Container and Resource Discovery Domain Trust Discovery File and Directory Discovery Group Policy Discovery Network Service Scanning Network Share Discovery Network Sniffing Password Policy Discovery Peripheral Device Discovery Permission Groups Discovery (3)	Exploitation of Remote Services Internal Spearphishing Lateral Tool Transfer Remote Service Session Hijacking (2) Remote Services (6) Replication Through Removable Media Software Deployment Tools Taint Shared Content Use Alternate Authentication Material (4)	Adversary-in-the-Middle (2) Archive Collected Data (3) Audio Capture Automated Collection Browser Session Hijacking Clipboard Data Data from Cloud Storage Object Data from Configuration Repository (2) Data from Information Repositories (3) Data from Local System Data from Network Shared Drive Data from Removable Media Data Staged (2) Email Collection (3)	Application Layer Protocol (4) Communication Through Removable Media Data Encoding (2) Data Obfuscation Dynamic Resolution (3) Encrypted Channel (2) Fallback Channels Ingress Tool Transfer Multi-Stage Channels Non-Application Layer Protocol Non-Standard Port Protocol Tunneling Proxy (4) Remote Access Software

Explicit: if the tactic / technique is mentioned explicitly. For example, the technique “[Inter-Process Communication: Dynamic Data Exchange](#)” is mentioned explicitly below.

<p>directly, such as https://attack.mitre.org/techniques/T1059/005 +implicit IMPL TECHNIQUE</p> <p>via malicious macros in weaponized Office documents</p> <p>https://attack.mitre.org/techniques/T1059/001 -implicit TECHNIQUE</p> <p>, compressed executables, PowerShell</p> <p>https://attack.mitre.org/techniques/T1559/002 -implicit TECHNIQUE</p> <p>dynamic data exchange (DDE),</p> <p>https://attack.mitre.org/techniques/T1059/007 +implicit IMPL TECHNIQUE</p> <p>HTA/HTML files , and traffic distribution systems</p> <p>https://attack.mitre.org/software/S0154 -implicit MALWARE https://www.wikidata.org/wiki/Q29364585</p>	<p>cysec_identifier_manual</p> <p>https://attack.mitre.org/t</p> <p>description</p> <p>implicit</p> <p>Yes No</p>
--	---

For more examples of explicit mentions of tactics and techniques, see below.

Implicit: Use this if it can be inferred from the description of an action or intention that a particular tactic or technique is employed. The annotation spans over the (smallest) part of the sentences that the annotator feels conveys or evokes the concept. The value is set to IMPL_TECHNIQUE and the feature “implicit” is set to “Yes”. For example, the expression “document implemented malicious Excel 4.0 Macros (XLM)” implicitly refers to the technique [Command and Scripting Interpreter: Visual Basic](#).

<p>13 The actor used a series of tools in this operation, including KeitaroTDS, https://attack.mitre.org/techniques/T1059/005 -implicit IMPL_TECHNIQUE Technology -org_or_sector_targeted ORG Microsoft a malicious Microsoft Excel spreadsheet document builder and the http://www.bosch.com/node1fpjil0acx6 -implicit MALWARE Zloader banking trojan (aka http://www.bosch.com/node1fpjil0acx6 -implicit MALWARE Terdot).</p> <p>14 Considering the nature of the malspam documents (usually named “Invoice”) and the use of a banking trojan, we assess the intended goal of the attackers was to make unauthorized bank transfers from victim accounts.</p> <p>15 Email details</p>	<p>No links or relations connect to this annotation.</p> <p>cysec_identifier</p> <p>cysec_identifier_manual</p> <p>https://attack.mitre.org/t</p> <p>description</p> <p>implicit</p> <p>Yes No</p> <p>value</p> <p>IMPL_TECHNIQUE</p>
--	--

At times, the decision whether to mark a sentence/phrase/expression that refers to a technique or tactic is explicit or implicit is non-trivial.

Tactics

Tactics are defined by the MITRE ATT&CK Framework. Tactics are the Goal that the attacker wants to achieve in this phase of the kill chain. For example, the adversary may want to achieve [Command and Control](#). Set the value to TACTIC and enter the URL corresponding to the identifier of the MITRE tactic into the cysec identifier field.

In the following example, the author directly refers to the tactic Command and Control (“The adversary is trying to communicate with compromised systems to control them.”) by using an alias of the technique that is well-known by cyber security professionals (“C2 communication”). Hence, the tactic is marked as being referred to explicitly.

<p>87 After https://attack.mitre.org/software/S0154 -implicit MALWARE https://www.wikidata.org/wiki/Q29364585 Cobalt Strike has been executed and a https://attack.mitre.org/software/S0154 -implicit MALWARE https://www.wikidata.org/wiki/Q29364585 Beacon established for https://attack.mitre.org/tactics/TA0011 -implicit TACTIC C2 communication , actors have been observed attempting to enumerate network connections and https://attack.mitre.org/techniques/T1046 -implicit TECHNIQUE dumping Active Directory credentials as they try to move laterally to a</p>	<p>cysec_identifier_manual</p> <p>https://attack.mitre.org/t</p> <p>description</p> <p>implicit</p> <p>Yes No</p> <p>value</p> <p>TACTIC</p>
--	---

Techniques

Techniques define how the attacker try to achieve his goal in this phase of the kill chain. For example, the technique [Application Layer Protocol: Web Protocols](#) describes the use of HTTP traffic for achieving [Command and Control](#).

In the following example, the technique [Phishing / Spear phishing Attachment](#) is mentioned by the phrase “campaigns distributed email threats with malicious document attachments”.

82

For example, the earliest
<https://attack.mitre.org/software/S0154> -implicit | MALWARE <https://www.wikidata.org/wiki/Q29364585>
Cobalt Strike
<https://attack.mitre.org/techniques/T1566/001> -implicit | TECHNIQUE
campaigns distributed email threats with malicious document attachments to
distribute the malware, but
<https://attack.mitre.org/techniques/T1566/002> -implicit | TECHNIQUE
campaigns distributing malicious URLs directly in the email body have overtaken
attachments as the more frequently utilized threat type.

83

While instances of
<https://attack.mitre.org/software/S0154> -implicit | MALWARE <https://www.wikidata.org/wiki/Q29364585>
Cobalt Strike being sent
directly as an initial payload have dramatically increased, deployment as a second
stage payload remains popular.

84

<https://attack.mitre.org/software/S0154> -implicit | MALWARE <https://www.wikidata.org/wiki/Q29364585>
Cobalt Strike has been

threats with

No links or relations connect
to this annotation.

cysec_identifier

cysec_identifier_manual

https://attack.mitre.org/t

description

implicit

Yes No

value

TECHNIQUE

In the following example, it is not stated that document macros are written in VB, but a domain expert can infer from the context that the technique mentioned here is [Command and Scripting Interpreter / Visual Basic](#).

86

<https://attack.mitre.org/tactics/TA0011> -implicit | IMPL_TECHNIQUE
<https://attack.mitre.org/software/S0154> -implicit | MALWARE <https://www.wikidata.org/wiki/Q29364585>
preceding malware typically loads and executes Cobalt Strike
Likewise, there is a wide array of techniques leveraged in cases where
<https://attack.mitre.org/software/S0154> -implicit | MALWARE <https://www.wikidata.org/wiki/Q29364585>
Cobalt Strike is delivered
<https://attack.mitre.org/techniques/T1566/001> -implicit | TECHNIQUE
<https://attack.mitre.org/techniques/T1059/005> +implicit | IMPL_TECHNIQUE
directly, such as via malicious macros in weaponized Office documents
<https://attack.mitre.org/techniques/T1059/001> -implicit | TECHNIQUE
, compressed executables, PowerShell,
<https://attack.mitre.org/techniques/T1559/002> -implicit | TECHNIQUE
dynamic data exchange (DDE),
<https://attack.mitre.org/techniques/T1059/007> +implicit | IMPL_TECHNIQUE
HTA/HTML files , and traffic distribution systems

cysec_identifier_manual

https://attack.mitre.org/t

description

implicit

Yes No

value

IMPL_TECHNIQUE

wiki_identifier

wiki_identifier_manual