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.

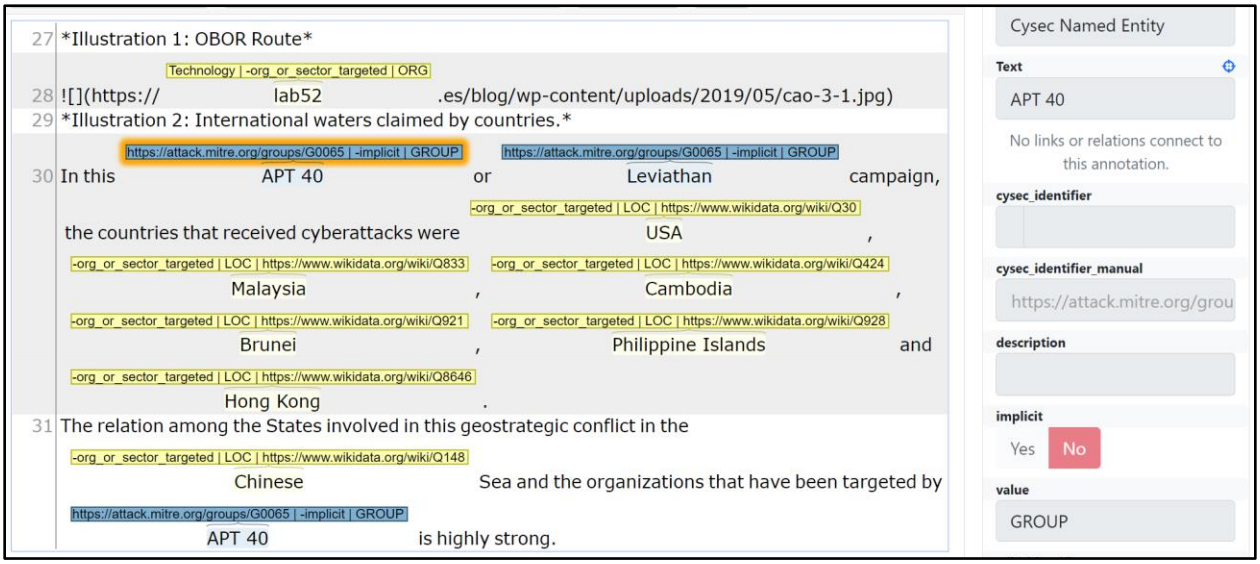
Cyber-security specific concepts, tactics and techniques are marked and linked to the publicly available MITRE ATT&CK framework. In the following, we describe each of the tags and how they are linked.

GROUP

Annotate any APT (Advanced Persistant 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/).



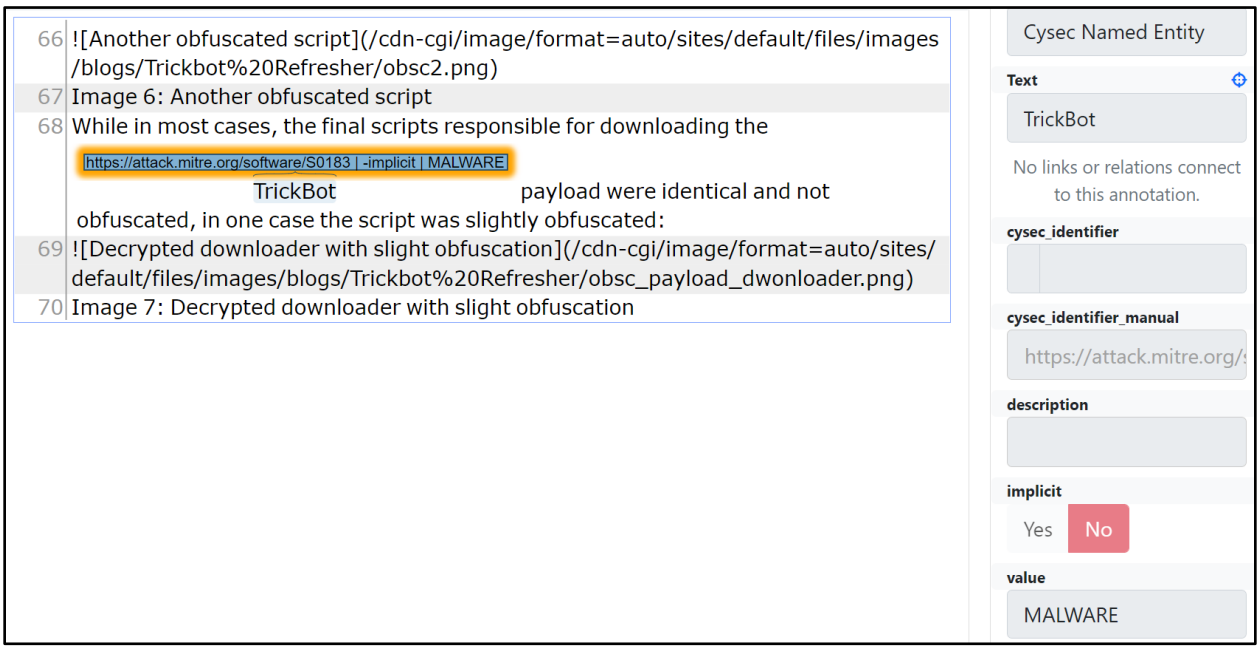
Software: MALWARE, TOOL, CONCEPT

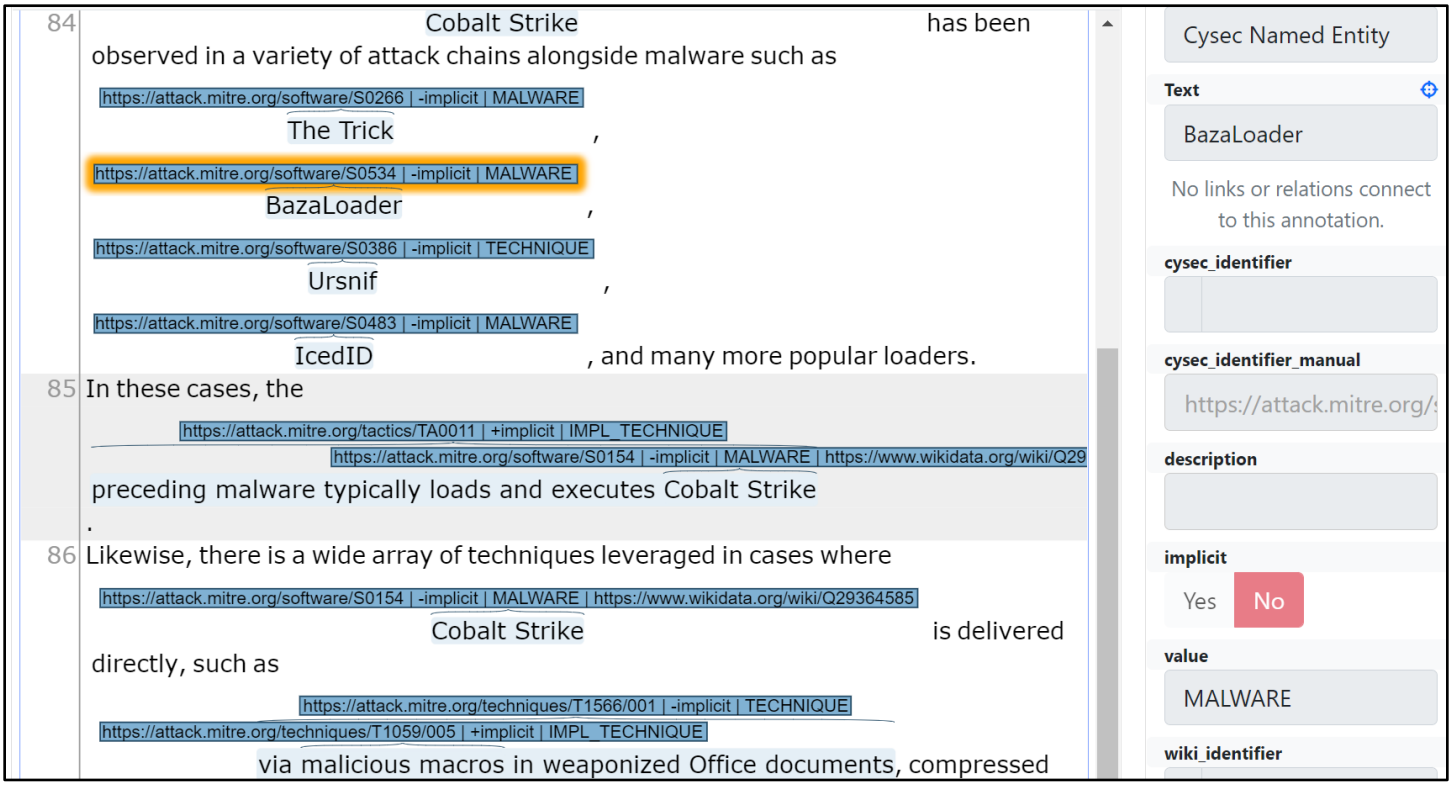
Definition: *Software* is a generic term for custom or commercial code, operating system utilities, open-source software, or other tools. We distinguish two types of software in our annotation: MALWARE, which refers to software that has been written specifically for malicious purposes, and TOOLS, which 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.”

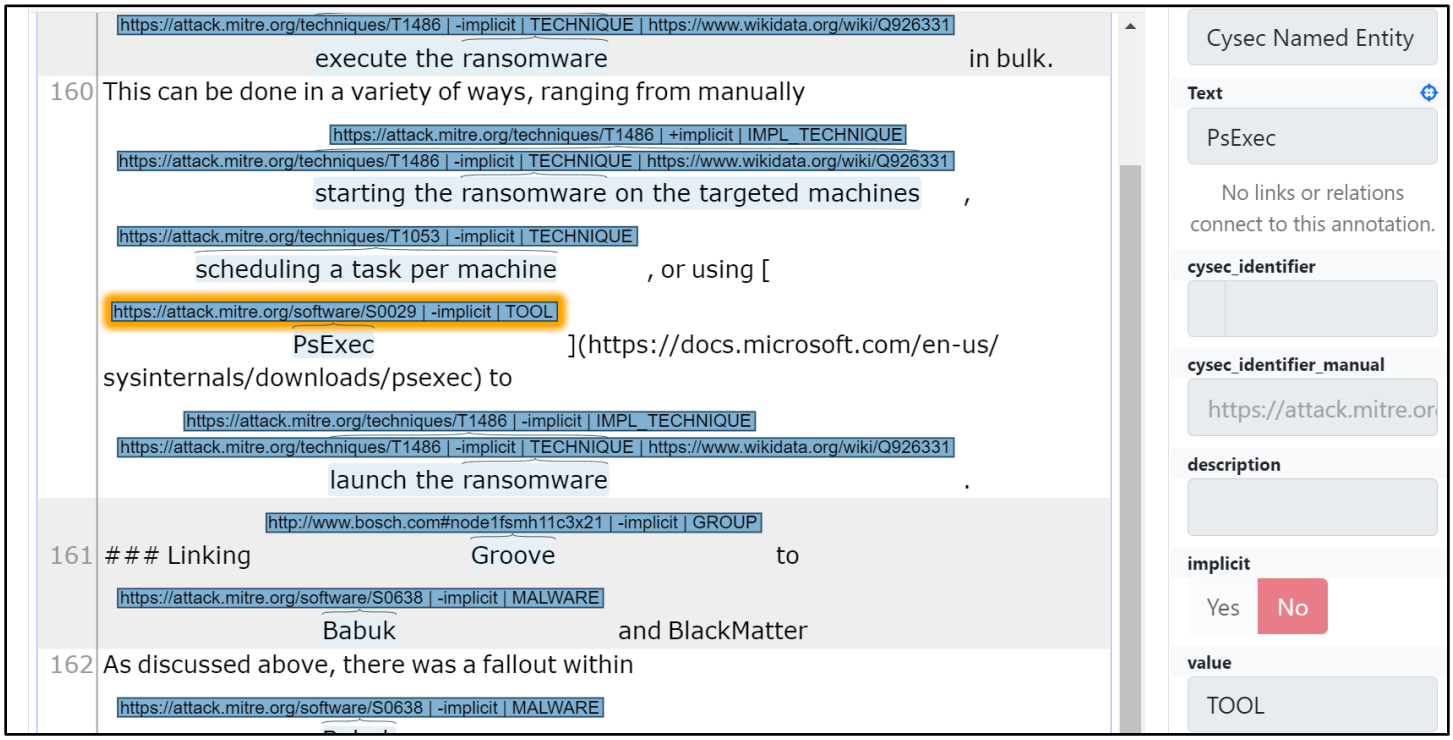




TOOL

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**.

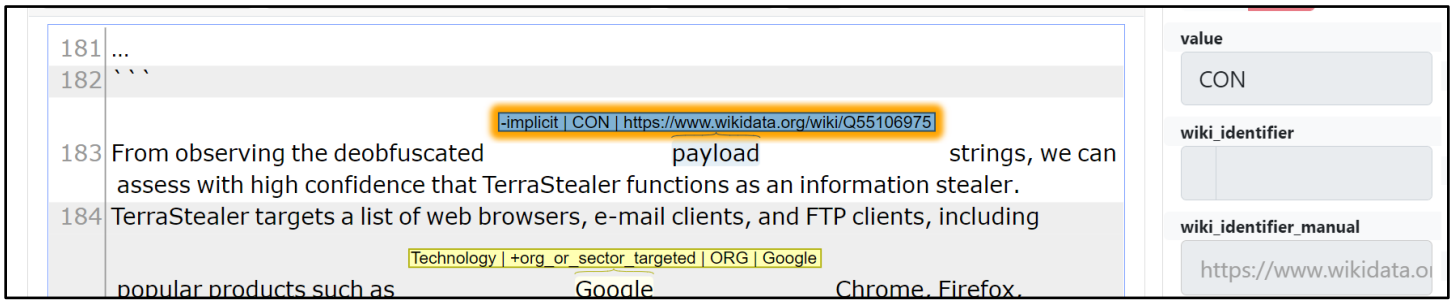
Examples: Netcat, wget, PsExec but also Mimikatz, Metasploit, PsExec…



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).

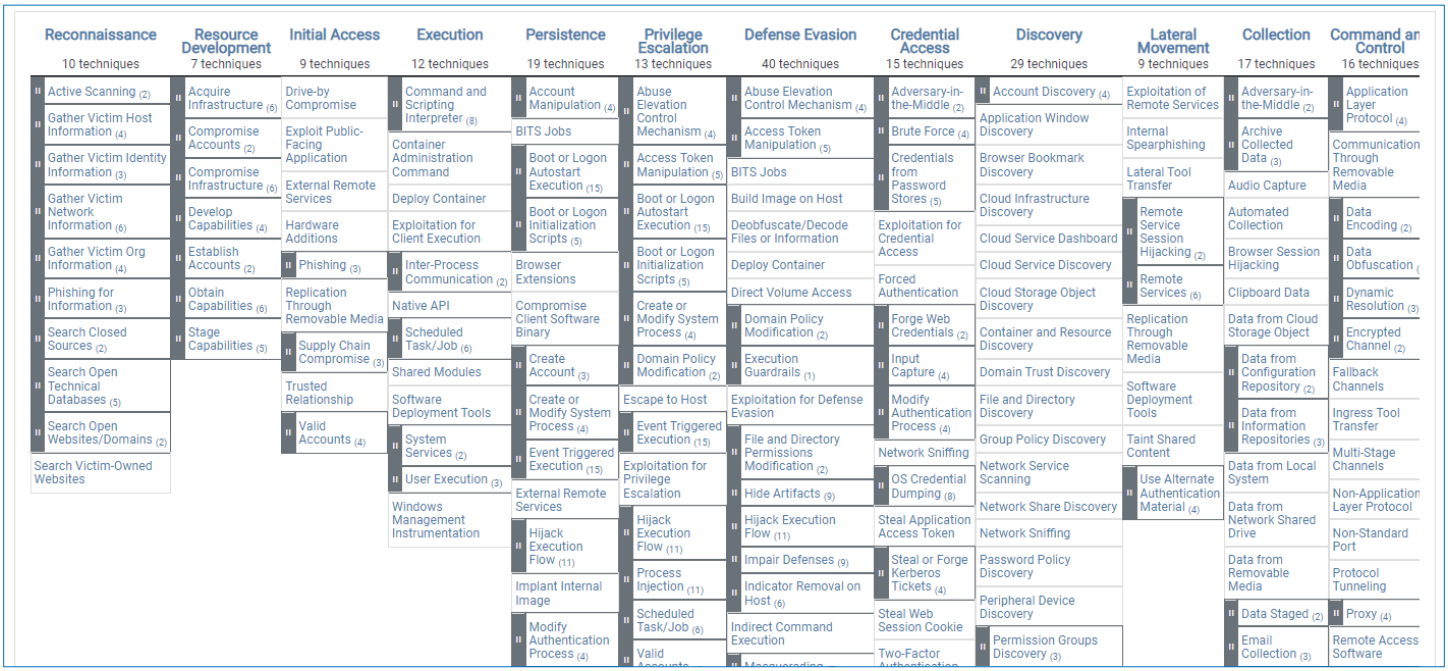
Examples: RemoteAccessTool/Trojan (RAT), Bitcoin, MD5, payload, Ransomware as a Service (RaaS), Discord



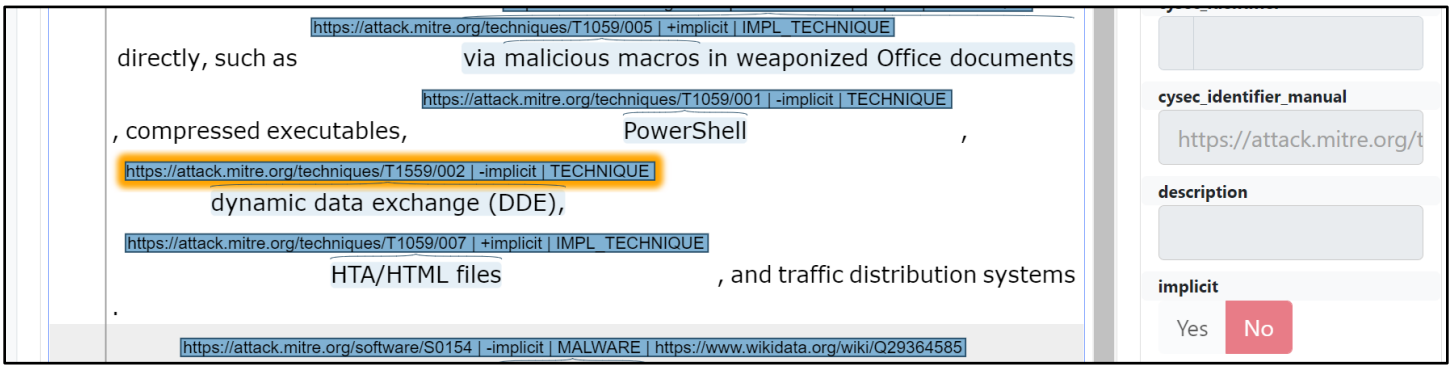
Explicit vs. Implicit Tactics and Techniques

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*.

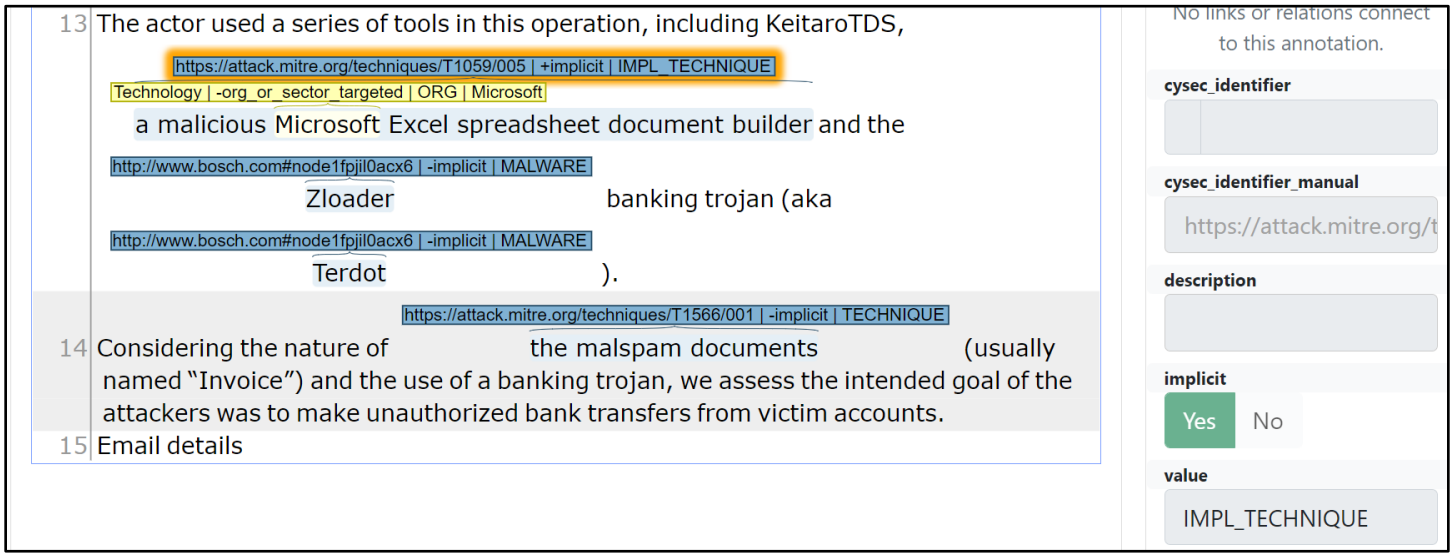


Explicit: if the tactic / technique is mentioned explicitly. For example, the technique “Inter-Process Communication: Dynamic Data Exchange” is mentioned explicitly below.



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](#).

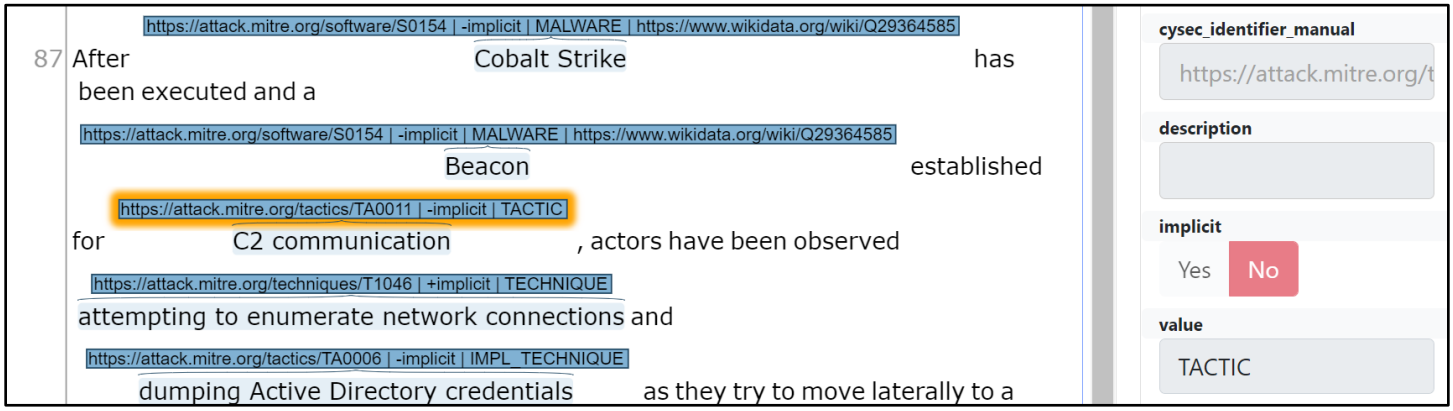


At times, the decision whether to mark a sentence/phrase/expression that refers to a technique or tactic is explicit or implicit is non-trivial.

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.

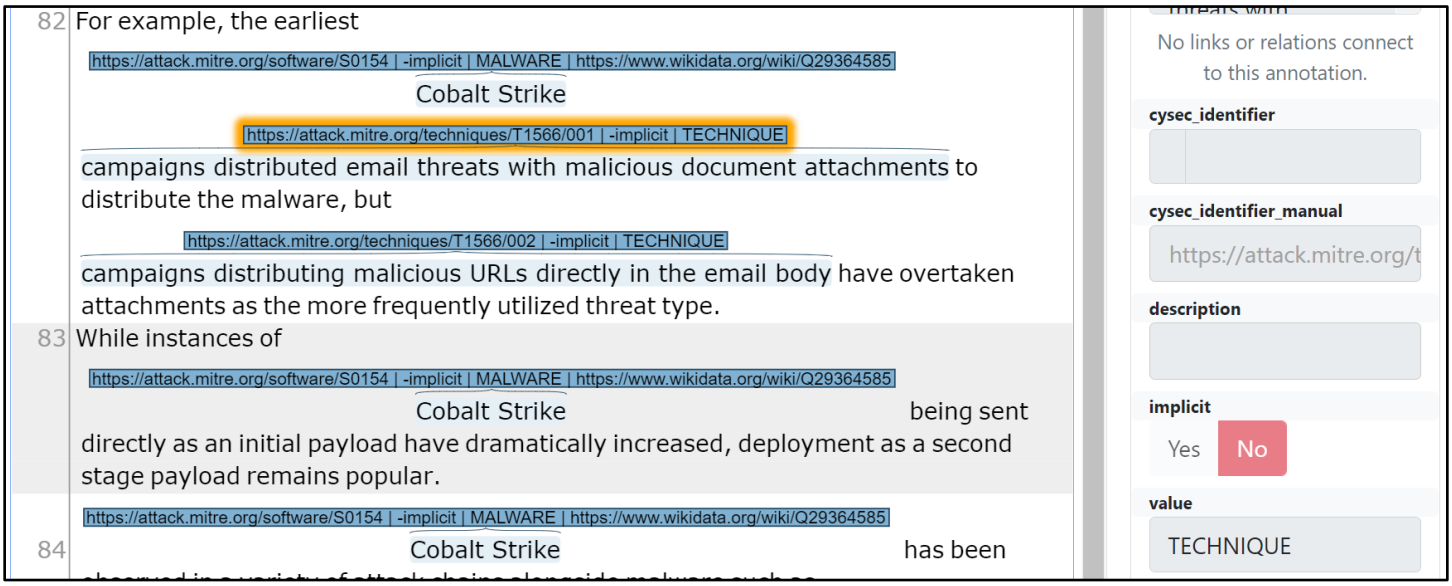
In the following example, the author directly refers to the tactic Command and Control ("The adversary is trying to communicate with compromised systems to control them.)" by using an alias of the technique that is well-known by cyber security professionals ("C2 communication"). Hence, the tactic is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



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”.



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.