摘要 在ATT&CK中使用自然语言文本描述APT攻击，需要从这些描述文本中提取网络攻击关系，以便进行语义匹配，发现威胁并还原上下文信息。设计并实现了一种基于ATT&CK的语义规则模型，使用命名实体识别，关系抽取等自然语言处理技术对ATT&CK中战术和技术的定义文本进行处理，生成123 个 APT 攻击语义规则，涵盖 ATT&CK 的 115 项技术和 12 种战术，最后对。。。。。。。。。。。。。。。。。。。。。。。。

1. **引言**

APT（高级持续威胁）攻击通常由主权国家政府，大型企业或组织发起，具有持续时间长，隐蔽性高，攻击形式复杂，技术手段高级等特点，APT攻击会使用多种网络攻击技术，利用系统中的各种漏洞进行攻击，由于攻击形式多样，方式隐蔽，又经常在攻击完成后隐藏痕迹，故往往在被发现时攻击已经完成。随着APT攻击的不断发展，其对网络安全的威胁也不断升高。在这种背景下，基于威胁情报的安全分析技术逐渐得到重视，威胁情报是安全情报的一种，通过情报之间的信息共享，减少系统漏洞，降低安全风险，从而避免可能的攻击，提高系统安全性。

以新能源为主体的新型电力系统是国家重要会议上提出的概念，是对我国未来电力系统的重要定位。

新型电力系统与传统的电力系统相比，有清洁低碳， 安全可控，灵活高效，智能友好等特点，其能源结构由多数为煤电机组转变为使用更多水电和风电的新能源机组，而在控制系统方面，新型电力系统也会使用更多与5G，物联网等新领域相关的智能设备。虽然这些新技术的使用提高了电力系统的运行效率，但也正因如此，新型电力系统相较于传统电力系统更为复杂多变，存在更多的安全隐患，又因为电力系统作为重要国家基础设施，常常成为APT攻击的攻击对象，因此，对针对新型电力系统的APT攻击进行建模和预测已是当务之急。

目前主流网络攻击模型已经从传统树结构或图结构模型发展为描述攻击生命周期的杀伤链模型（Cyber Kill Chain）和MITRE ATT&CK模型，ATT&CK更加详细，包括许多攻击技术的描述，更适用于对APT攻击的描述。

威胁情报技术是现代网络安全机制的重要组成部分，David Bianco所提出的痛苦金字塔模型由IOCs组成，其中位于最顶部的TTPs（战术，技术和程序）痛苦指数最高，因为其从攻击者行为本身进行防御，这是攻击者最难改变的，但是TTPs情报均使用自然语言描述，处理比较困难，因此相较于IOC(Indicators of Compromise)威胁情报，对TTPs情报的研究和利用较少，而在自然语言处理中，可以使用神经网络模型对文本进行处理，从而提取语义规则，实现对TTPs情报的利用。

本文的主要工作如下：

1. 基于ATT&CK模型，设计针对新型电力系统APT攻击的ATT&CK模型，通过对

APT攻击使用的战术和技术进行分析，去除不常用的战术，以及恶意代码和行为事件等，更加精确地形容APT攻击

1. 使用自然语言处理（NLP）技术，将ATT&CK中对网络攻击战术和技术进行定义的文本转化为语义规则图，从而构建APT攻击语义规则。

1. 研究背景
   1. 新型电力系统

以新能源为主体的新型电力系统在中央财经委员会第九次会议首次被提出，是我国电力行业的基础和新的发展方向，新型电力系统将会减少传统能源的使用，使用更多的新能源并入电网，这也意味着未来电力系统的能源结构更加多元化，电力调配和存储更加复杂，因此，新型电力系统中会使用大量数字化，智能化设备。由于更多新数字技术的加入，新型电力系统更加复杂，其各个环节的脆弱性也更多，新能源与多元负荷形态比例大幅提升。高比例的可再生新能源与电力电子设备的接入，以及供给侧和需求侧的随机性，导致电网遭受的攻击面与攻击机会增大因此，对针对新型电力系统的攻击行为的识别，检测和防御十分重要

* 1. APT攻击

高级可持续威胁（Advanced and Persistent Threat, APT）通常由大型企业，主权国家及由其控制的组织所发起，通常是为了政治或经济利益，对某一特定的组织或机构进行的攻击，由于其大多不是为了短期的利益，而往往注重对系统整体的渗透，因此，APT具有活动周期长，隐蔽性高以及破坏性大等特点，许多APT攻击的受害者在攻击发生后也不会察觉。APT攻击的发起者往往是大型组织或主权国家，其攻击的对象也大多为跨国企业，国际组织和政府部门。电力和能源设施作为一个国家重要的基础设施，在军事和民生领域至关重要，因此也经常成为APT攻击的目标，比如，

近年来，威胁情报技术发展迅速，Gartner 公司将威胁情报定义为一种基于证据的知识, 为威胁响应提供包括相关场景、威胁 指标、含义和可行建议等决策依据[1]。Bianco 提出金字塔模型[2], 将威胁情报数据划分为 6 个层次, 自下 而上依次为哈希、IP 地址、域名、网络或主机特征、 攻击工具、TTPs, 在模型中, 威胁情报收集难度逐层 增加, 网络防御能力逐层增强。在对威胁情报数据的利用方面，IOC（Indicators of Compromise）（妥协指标）的使用更为广泛，IOC editor便是基于IOC规则来识别攻击行为的一款安全软件，但是对于TTPs（Tactics, Techniques and Procedures）威胁情报的利用则比较少，其原因主要在于TTPs情报是使用自然语言描述的，需要人工对其进行情报分析，难以自动化分析和使用。因此，需要使用NLP中语义规则构建技术从攻击技术定义文本中进行知识抽取，将其转化为易于理解和利用的规则，从而实现对TTPs情报的自动化利用。