

XML漏洞和Excel文件

如果您的代码 `.xlsx` 提取的文件来自您没有绝对信任的来源，请注意 `.xlsx` 文件由XML组成，因此容易受到XML漏洞的影响。

xlrd使用ElementTree来解析XML，但正如您所看到的那样，有许多不同的ElementTree实现。您可以在此处找到您应该担心的漏洞的一个很好的摘要：[XML漏洞](#)。

为清楚起见，xlrd将尝试从以下来源导入ElementTree。该列表按优先级顺序排列，列表中较早的列表优先于列表中较晚的列表：

1. `xml.etree.cElementTree`
2. `cElementTree`
3. `lxml.etree`
4. `xml.etree.ElementTree`
5. `elementtree.ElementTree`

为了防止出现这些问题，您应该考虑使用[defusedxml](#) 项目，该项目可以按如下方式使用：

```
import defusedxml
from defusedxml.common import EntitiesForbidden
from xlrd import open_workbook
defusedxml.defuse_stdlib()

def secure_open_workbook(**kwargs):
    try:
        return open_workbook(**kwargs)
    except EntitiesForbidden:
        raise ValueError('Please use a xlsx file without XEE')
```