如何使用CLI搜索导出大量数据。

适用范围

Splunk企业

描述

分步指南

要导出大量数据，可以在搜索头上的CLI提示符处按以下顺序进行搜索：

```
splunk search "index=pan_log sourcetype=pan:* src_ip!=<Internal_IP_Scheme> dest_ip!=<Int
| fillnull value=null
| dump basefilename=<insert_your_chosenfilename> compress=5 fields=_time,app,app:subcate
dest_ip,dest_port,dest_translated_ip,dest_user,dest_zone,direction,duration,filename,htt
packets_out,sig,signature,sourcetype,src_ip,src_translated_ip,src_user,target_id,threat:
```

上述搜索的细分如下：

搜索pan:*源类型的“pan_log”索引，并排除IP范围，可以填充空值，然后生成转储命令，然后继续执行以下操作：

此命令运行指定的搜索查询，并将搜索结果立即导出到位于“$SPLUNK_HOME/var/run/splunk/dispatch/<sid>/dump”的本地磁盘。 它识别输入事件中的一个特殊字段_dstpath，如果设置该字段，将用作附加到dst以计算最终目标路径的路径。

此过程的一个主要缺点是结果的TTL为2分钟，因此必须创建保存令牌文件以确保在获取由过程创建的*.gz文件之前没有获取这些结果。请参阅下面的“相关”链接。

相关信息：

[如何在Windows中创建无扩展名的文件]

调查

调查此问题时，发现它没有获取完整的数据集。

已确定是搜索获取这些数据，并以比预期更快的速度进行剔除。经研究发现这是正常行为。摘要索引搜索和其他搜索可以使用2分钟的搜索项目TTL运行。因此，请在<sid>目录中创建保存令牌文件。对于Windows计算机，这将在本文说明部分的链接中介绍。使用*nix，您可以导航到$SPLUNK_HOME/var/run/splunk/dispatch/<sid>，然后发出命令“touch save”

完成此操作后，就可以绕过对象获取问题[***不要忘记删除保存文件!***]

然后，再进行一些故障排除，获得可生成定义的完整数据集的可用时间范围。由于数据集中的事件数量极大，因此最终需要进行24小时的搜索，以便逐日完成导出。

分辨率

一旦在需要针对24小时时间段运行的数据集上收窄，能够生成24小时延期PAN日志记录的完全转储。另一家TSE最近也使用同样的方法来追溯历史并获取特定信息。

相关链接

Splunk文档：转储命令手册： 搜索引用转储

关于此翻译
思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。