2016 年 6 月 29 日，星期三

检测 DNS 数据泄漏

本博文由 Martin Lee 和 Jaeson Schultz 共同撰写。特别感谢 Warren Mercer 提供的建议。

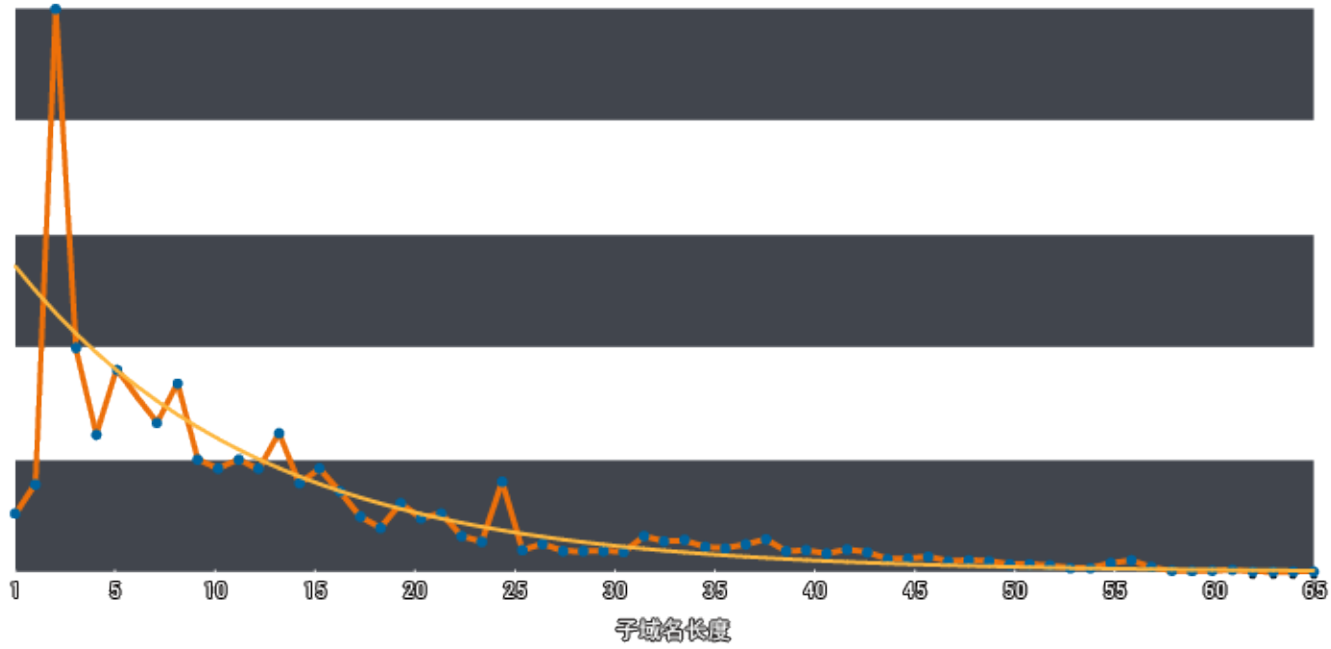
Wekby 和销售点最新发现恶意软件可使用 DNS 请求作为命令和控制信道，这更加突显了将 DNS 视为潜在恶意信道的必要性。虽然专业分析师能够凭借对所在组织正常 DNS 活动的了解，快速发现异常活动，但手动查看 DNS 日志通常是一项耗时又乏味的工作。在不了解恶意 DNS 流量活动表现的情况下，我们该如何识别恶意 DNS 请求呢？

我们都有潜意识，这种思维模式会影响我们对环境的感知，并帮助我们识别异常情况。例如，街坊四邻发生的特别的或异常的事件往往能勾起我们的好奇心，驱使我们一探究竟。我们会比较观察到的情况与我们所熟悉的常态，如果二者不符，我们便想了解原因所在。对于 DNS 日志，我们也可以采用相同的方法。如果我们能构建一个“常态”基准或模式，就可以将观察到的结果与之进行比较，以确定我们看到的实际现象是否与我们的预期存在很大差距。

我们熟悉一般的 DNS 请求（例如请求“*www.cisco.com*”的 IP 地址），但是哪类请求属于异常请求并且需要调查呢？恶意软件会对窃取的数据进行编码，将其作为以攻击者控制的域名服务器所在的域为对象的 DNS 查询的子域名部分。例如，对“long-string-of-exfiltrated-data.example.com”的 DNS 查询会被转发到 *example.com* 的域名服务器，后者将记录“*long-string-of-exfiltrated-data*”并向恶意软件回复一个编码的响应。

我们可以认为在正常情况下，此类请求的子域名部分会比一般的请求长很多。所以我们可以使用 DNS 请求中的子域名长度的分布数据来构建一个描述其常态分布的数学模型，然后使用这个模型与我们的观察结果进行比较，从而识别异常。

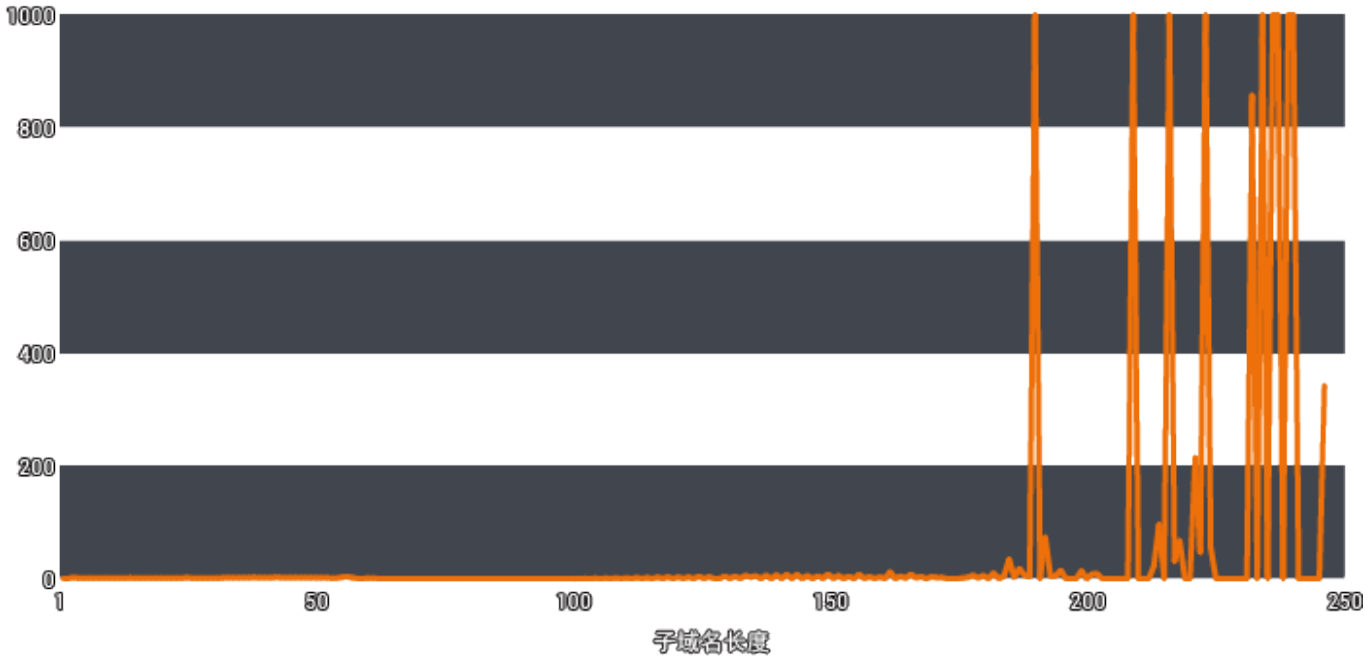
我们选择了一批 DNS 请求样本，通过删除域名和域扩展名来计算子域名长度的出现频率，所得到的结果如下图所示：



从橙色线可以看出，子域名长度的分布范围介于 1 个字符到 65 个字符之间。虽然此分布范围明显与黄色线显示的平滑指数曲线不完全匹配，但二者非常接近。我们可将该曲线用作常态模型，将观测值与该曲线比较，以检测异常。

我们可以直观地发现，长度为 3 个字符的子域名比我们所预计的要频繁得多。但是可以理解，这个长度与“*www*”的长度相对应，这是一个十分常见的子域名字符串。要衡量此观测值的频率比我们的预期高出多少，我们可以用观测值除以曲线中的预测值，计算此观测值的异常指数。

通过继续使用此计算方法对所有长度值进行计算并绘图，我们得到了一个反映每个子域名长度的出现频率与我们预期的常态之间的实际偏差的图示：



显然，有多个子域名长度的出现频率比我们的预期高很多。事实上，观测值与预期值的偏差非常大，以至于我们不得不将一些值取为 1000。

集中显示这些异常值可减少检查日志集所需的手动操作。许多特别长的子域名最终证明是合法的云服务或内容分发网络。但是，子域名长度为 231 和 233 个字符的几个域名看起来特别可疑。

log.nu6timjqgq4dimbuhe.3ikfsb---redacted---cg3.7s3bnxqmavqy7sec.dojfgj.com
log.nu6timjqgq4dimbuhe.otlz5y---redacted---ivc.v55pgwcschs3cbee.dojfgj.com
lll.nu6toobygq3dsnjrgm.snksjg---redacted---dth.ejitjtk4g4lwvbos.amouc.com
lll.nu6timrshe4timrxhe4a.7vmq---redacted---hit.w6nwon3hnifbe4hy.amouc.com
ooo.nu6tcnbug4ytkobxhe4q.zrk2---redacted---hxw.tdl2jg64pl5roeek.beevish.com
ooo.nu6tgnzvgm2tmmbzgq4a.rkgo---redacted---tw5.5z5i6fjnugmxfowy.beevish.com

尽管每个域的域名服务器托管在不同的网络上，但这些域名有几个相同的异常特性：每个域有数百个子域，但每个唯一子域到目前为止均只访问过一次。虽然这不一定是异常情况，但每个 DNS 查询都返回“*192.168.0.1*”这一点很值得注意。

Dojfgj.com 是一个众所周知的恶意域名，Multigrain 恶意软件就是通过它输出窃取的信用卡号码。这三个域名之间明确的相似性说明之前未知的 *amouc.com* 和 *beevish.com* 域名与 *dojfgj.com* 相关。

Multigrain 恶意软件使用 base32 编码从受感染的计算机输出数据。尽管与使用更广泛的 base64 编码技术相比，base32 编码的空间效率较低，但它使用包含字符 a 到 z 和数字 2 至 7 的字母编码字符集。由于数字 0 和 1 与字母 O 和 I 较为相似，因此 base32 中没有采用这两个数字。此编码技术的优势在于任何字符都可在 DNS 查询中使用，并且无需保持大写。

Multigrain DNS 请求的大部分内容是加密的，但用于受感染计算机标识符编码的第一部分是可读的。例如，以下示例中从 nu6t 开始的部分：

ooo.nu6tgnzvgm2tmmbzgq4a.rkgo---redacted---tw5.5z5i6fjnugmxfowy.beevish.com

可解码为：*m=3753560948*

DNS 请求的任何特性（例如域名长度、子域名数量等）都可用来构建用于比对观测值的预期行为模型。
这些模型可识别类似以下模式的域名：

4-9-8-2-2-3-8-5-4-6-2-9-2-3-8-8---redacted---7-.0-0-0-0-0-0-0-0-0-0-0-0-0-49-0-0-0-0-0-0-0-0-0-0-0-0-0.info
5-2-4-6-3-2-2-7-4-8-3-6-7-1-2-3---redacted---0-.0-0-0-0-0-0-0-0-0-0-0-0-0-49-0-0-0-0-0-0-0-0-0-0-0-0-0.info
6-t-y-s-8-l-l-p-6-6-x-q-2-l-2-9-x-7---redacted---a-.0-0-0-0-0-0-0-0-0-0-0-0-0-45-0-0-0-0-0-0-0-0-0-0-0-0-0.info
7-8-5-4-1-2-7-2-7-8-4-5-1-5-0-7---redacted---0-.0-0-0-0-0-0-0-0-0-0-0-0-0-28-0-0-0-0-0-0-0-0-0-0-0-0-0.info

我们已经知道这些模式与 W32/AutoRun 蠕虫相关。还有更有趣的：

77newyourcomputerhaveaseriousproblemcallon18883142770tollfree.yourcomputerhaveaseriousproblempleasecallon18883142770tollfree.yourcomputerhaveaseriousproblempleasecallon18883142770.windows-has-detected-some-suspicious-activity-fromyourcomputer.com

这与网络钓鱼相关。

监控日志（特别是 DNS 日志）是一种非常好的攻击检测方法。但是如果数据多得无法用人眼识别，使用简单的技术对数据进行建模，可帮助我们确定需要额外注意的条目。而能否发现这些需要额外注意的条目，往往决定了您的网络安全与否。

发布者：MARTIN LEE 发布时间：上午 11:30