ABSTRACT

Cybersecurity experiments/exercises in a testbed require benign traffic to ensure a more effective and rigorous experiments/exercises process. This traffic is usually generated by humans, to camouflage the attack traffic by producing different activities/tasks in the testbed. Some agents can generate this traffic, but unfortunately, most simple agents are predictable, and their actions are easily distinguished from human behavior. We propose distributional models of human activity datasets to mimic replicate specific human actions. These models can be employed in a human agent (i.e., bot) controlled by our orchestrator to generate a realistic human activity generation that is easy to deploy and scale. This paper discusses our modeling of real users web browsing data into well-known distributions and how to fed them into our bot. The results show that our bot can produce realistic web browsing activities similar to real-human behavior.

1 INTRODUCTION

As the complexity of generation scenarios may change dramatically or the size of the exercise environment needs to scale up very quickly, a human-based or hard-coded tools scenario can not achieve scalability or efficiency in producing background traffic. This paper adopted the well-known bot concept from a BotNet (Kokkonen et al. 2015). However, a large number of bots execute complex human activities orchestrated by a single controller. The bot is designed as an independent software component with logic packed (i.e., containerized) for web browsing activity. It is fed with decision-making based on distribution models constructed from real human data to mimic real user’s activity. The summary of contributions are:

1. Model the dataset from real users dataset into well-known distribution of users’ browsing behaviors.
2. Propose a unique web browsing bot implementation to take the model as input.
3. Show proposed bot can generate realistic activity that are very similar to the users’ behavior.

2 REALISTIC WEB BROWSING ACTIVITY GENERATION

To mimic the web browsing activity generated by a human, we model the web browsing activity dataset. We select six features of human web browsing behaviours (i.e., normal inter-keystroke interval, password inter-keystroke interval, page reading rate, probability of first-time visit, funds deposit, and transfer amount) from five datasets, How We Type (Feit et al. 2016), Keystroke Dynamics (Killourhy and Maxion 2009), Natural Stories Corpus (Richard Futrell and Edward Gibson and Hal Tily and Idan Blank and Anastasia Vishnevetsky and Steven T. Plantadosi and Evelina Fedorenko ) and Daily website visitors (Nau 2020), PaySim Mobile Money Simulator (Lopez-Rojas et al. 2016). The overall design of realistic browsing activity generation, including data preparation, measurement results, and comparison, is shown in 1. As
an example of web browsing activity generation, a terracotta bank web application (Cummings 2021) is used with different workflows such as login/logout, account deposit, and balance transfer.

![Diagram of browsing activity generation using human behavior dataset](image)

Figure 1: Overall design of browsing activity generation using human behavior dataset.

3 BOT ACTIVITY MEASUREMENT AND ANALYSIS

We compare the distribution of the original data and generated data for each of the six features using K-S (Kolmogorov-Smirnov) test. Table 1 shows how “realistic” our bot is in generating web browsing activity. Given that the p-values are all greater than 0.45, even at a 45% level, we cannot reject the null hypothesis (i.e., generated data sample came from the distribution of the original datasets).

![Table showing distribution distance between datasets and measurement data](image)

<table>
<thead>
<tr>
<th>Features</th>
<th>K-S Test Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Password Keystroke Interval</td>
<td>0.014372759287635595</td>
<td>0.5650549040480954</td>
</tr>
<tr>
<td>Normal Keystroke Interval</td>
<td>0.01351518257560047</td>
<td>0.6435859689093139</td>
</tr>
<tr>
<td>Reading Time</td>
<td>0.01536088676002937</td>
<td>0.4785529177692344</td>
</tr>
<tr>
<td>Probability of First-Time Visit</td>
<td>0.012407546140967673</td>
<td>0.7448894171799325</td>
</tr>
<tr>
<td>Funds Transfer Amount</td>
<td>0.014301810387064517</td>
<td>0.5714668079068523</td>
</tr>
<tr>
<td>Funds Deposit Amount</td>
<td>0.012958882639986402</td>
<td>0.6948866263149966</td>
</tr>
</tbody>
</table>

REFERENCES


