Reinforcement of cyber deception strategies through simulated user behavior

Federico Pacheco – Diego Staino
National Technological University – University Institute of the Argentinean Federal Police
fpacheco@frba.utn.edu.ar – diegostaino@hotmail.com

Abstract- Cyber deception has emerged as a pivotal defensive strategy in the effort to detect and counter advanced persistent threats and sophisticated attacks. However, the implementation of conventional methods, such as decoy services, frequently lacks the credible support necessary to optimize their effectiveness. Generally, honeypot and honeytoken systems face limitations, including a lack of realism due to static configurations and an absence of human activity. Additionally, the simulation of actions is costly, and profiling is challenging to scale, as well as automation and adaptability. In this paper, we present a tool designed to automate the generation of realistic activities and behaviors of fictitious users, seeking to integrate personalized and coherent patterns of human interaction in cyber deception scenarios to improve the credibility of decoys. Based on the MITRE Engage framework, the tool contributes to the strengthening of defensive operations, addressing a key challenge of cyber deception.

Keywords- cyber deception, user behavior, deception technologies

INTRODUCTION

Cyber deception is an active cyber defense strategy that seeks to overcome the asymmetry between attacker and defender by using techniques like those employed by the adversaries themselves [1]. This strategy is based on human-directed activities, with automated elements, which increase the diversity and complexity of systems, making it difficult for attackers to gather useful information. Implementation of this strategy requires both attraction to instrumented traps and monitoring of systems to detect their presence. The fundamental principle of cyber deception is the deployment of traps and lures within infrastructures or systems that mimic authentic assets. This approach enables the real-time monitoring of attacker interactions, facilitating the collection of tactical data. The scope of these strategies encompasses various elements, including information, servers, folders, files, accounts, workstations, tokens, projects, individuals, domains, services, and fake networks. The overarching objective of these strategies is to create confusion, impede the progress of malicious actors, and expose their actions within controlled computing environments.

When a threat actor performs reconnaissance or lateral movements, they assume that their interactions are legitimate, thereby reducing their effectiveness. That can degrade an attacker's interpretation capacity, forcing them to interact with elements that provide little useful information and generating uncertainty about their previous findings. Cyber deception can reduce an attacker's dwell time in the system, accelerate threat detection, and reduce alert fatigue. These technologies generate reliable metrics, such as Indicators of Compromise (IoC) and Techniques, Tactics, and Procedures (TTPs) with a low false positive rate.

In 2022, MITRE introduced the Engage framework, which organizes cyber deception operations into well-defined phases (prepare, expose, affect, elicit, and understand)[2]. The first two phases focus on strategic planning, while the latter involve direct engagement with the adversary. Engage does not delineate the specifics of implementation, thereby delegating this responsibility to each respective organization. Deception technologies, which are tools that implement cyber deception techniques, are designed to entice and entrap attackers by simulating systems, data, and credentials. These solutions create environments in which it is challenging to maneuver without triggering alerts. These tools offer capabilities such as credential protection, decoy deployment, and integration with security operations to enhance detection and response. Their objective is to be effective in generating early detection and minimizing the impact of attacks by diverting the attacker into controlled environments. The primary challenge in implementing cyber deception strategies lies in the credibility of the lures (commonly called "decoys"). Advanced adversaries can identify and evade decoys that lack a realistic human context. The prevailing tools concentrate on generating fake technical assets, neglecting to simulate human activity, which curtails the efficacy of deception and exacerbates decoy detection. To remediate this, we propose an experimental, open-source tool designed to automate the generation of fictitious profiles, which we designate "honey profiles" to emulate human behavior in digital environments. This concept has been employed in virtual community environments and online social networks, in relation to attracting cyber criminals, yet it has never been applied to internal organizational environments. Although this tool is only just being tested in real-life and in laboratory settings, its objective is to introduce personalized and consistent interaction patterns.

Our approach is based on three fundamental principles. The first is the creation of tailored fictitious profiles, with digital identities that include job roles, usage patterns, and communication habits. The second is the simulation of human activities, through automated generation of interactions on platforms such as terminal interaction, web browsing, and access to data or resources. The third is the alignment with MITRE Engage, for planning, execution, and analysis of operations.

This paper focuses on presenting a tool with innovative features, describing its technical architecture, and applying it in alignment with MITRE Engage. In addition, the challenges of its implementation and strategies to validate its effectiveness in real scenarios are discussed. The paper is structured in three main sections. The first section provides a context and an overview of the state of the art, wherein the approaches to current cyber deception strategies and their limitations are analyzed. The second section presents the problems associated with cyber deception, exposing specific challenges that are sought to be solved. The third section is the proposal and design of the tool itself, where its architecture and operation are described, along with a feasible strategy to validate its effectiveness. Finally, limitations and future work are presented, along with discussion and conclusions.

CONTEXT AND STATE OF THE ART

The advent of modern technology has precipitated the evolution of cyber deception, thereby facilitating the integration of artificial intelligence and machine learning to generate decoys and adapt strategies in accordance with the movement of attackers. This has enabled the automation and scalability of more realistic deceptions, thereby integrating them more efficaciously into cyber defense practices. Moreover, contemporary solutions empower the customization and simulation of entire networks, thereby enhancing their credibility and effectiveness. Integration with security tools such as intrusion detection and prevention systems (IDS/IPS) enhances threat detection and incident response.

Despite their effectiveness, deception technologies need continuous management to maintain their credibility, requiring specialized manpower time. Attackers continuously develop methods to detect and evade these systems, necessitating constant innovation in defensive strategies. The implementation and maintenance of these strategies can incur substantial expenses, underscoring the need for a cost-effective balance between investment and effectiveness[5].

The complexity and expense of commercial cyber deception solutions often limit their accessibility to corporate and government environments in developed countries. While open-source tools exist for deploying honeypots and other specific elements, they constitute only a fraction of an integral cyber deception strategy[6]. Organizations may elect to develop their own implementations using available open-source tools, along with their own custom scripts. However, this approach is more viable in large environments with adequate resources to deal with costs and technical issues. In-house implementation does not avoid introducing additional complexity into production environments, increasing risks and generating conflicts with technology and infrastructure areas, which may see these projects as an additional burden. The combination of high licensing costs, difficulties in internal implementation, and the inherent complexity of these strategies is a common problem.

As a proactive cybersecurity strategy that seeks to detect, deflect, and gather intelligence on adversaries in digital environments, it requires the creation of fake assets, such as honeypots, honeytokens, and deception networks. These assets are used to confuse attackers and delay their movements while their tactics, techniques, and procedures (TTPs) are analyzed. The benefits of this strategy include the ability to detect threats early, divert attacker resources to spoofed assets, and collect valuable information to enhance defensive systems[7].

The effectiveness of cyber deception is contingent upon the ability of the decoys to mimic real systems in a credible manner. If an attacker detects that an asset is a decoy, its effectiveness is significantly diminished. To address this challenge, the MITRE Engage framework provides structured guidance for implementation. The application of this framework is organized into three phases: planning, in which objectives are defined and deception elements are designed; execution, which involves the implementation and monitoring of decoys; and analysis, focused on evaluating the effectiveness of the tactics employed to improve future strategies. Engage underscores the significance of crafting plausible narratives within deception environments, emphasizing that the mere generation of fake technical assets is insufficient. Instead, it asserts that these decoys must mirror the typical actions of legitimate users to enhance the credibility of these decoys and optimize the efficacy of cyber deception strategies in safeguarding digital infrastructures.

CYBER DECEPTION CHALLENGES

A significant challenge in the realm of cyber deception is the absence of compelling human context in the lures, a deficiency that can be readily identified by sophisticated attackers[4]. This, in turn, reduces the efficacy of traditional traps. In this regard, honeypots and honeytokens, which emulate technological systems, are deficient in their lack of realistic interactions. While a system may appear legitimate on a technical level, the absence of records of consistent human activity, such as access to applications or interactions with other users, can easily lead to its identification as a hoax, prompting the adversary to divert their efforts to other targets. Indeed, advanced operations necessitate the adoption of strategic disinformation measures, such as the controlled dissemination of false information, to confuse and delay the actions of cyber attackers[8].

The aforementioned lack of plausible interactions directly impacts the effectiveness of cyber deception. If attackers detect the trap, they may circumvent it or adjust their tactics, thereby limiting the opportunity to gather intelligence and divert their resources. This phenomenon results in a decrease in the detection of targeted attacks and a weakening of the overall performance of these strategies, hindering their defensive purpose. The majority of current implementations focus on the technical infrastructure, neglecting the simulation of the human factor, leaving a gap for attackers to exploit.

Concurrently, traditional decoys have significant limitations. Most are static and do not simulate any human activity, making them easily detectable by experienced attackers. Anomalies, such as the absence of consistent historical logs or low user interaction alert adversaries to the falsity of the system. In addition, manual simulation of fictitious profiles is costly and not very scalable, making it difficult to implement in large and dynamic environments.

Lack of automation and adaptability is also a challenge. Many current solutions rely on predefined configurations, which limit their ability to evolve in the face of constantly changing adversarial tactics. Additionally, there is no standard integration to incorporate human-like actions. The absence of convincing user profiles reduces the credibility of decoys and facilitates their detection. Another common issue is the design of coherent narratives, which is critical for decoys to be effective. The absence of realistic interactions, such as logins, queries, emails, or web browsing, can generate suspicions[9].

For deception to be effective, fake assets must reflect the typical behavior of legitimate users and maintain temporal and contextual consistency. Absent these elements, attackers can detect inconsistencies and avoid the trap. Creating profiles with humanized interactions is a manual and costly process, which limits its scalability. The definition of activity patterns for each profile necessitates time and resources, rendering this approach impractical in environments with voluminous data or numerous attackers. Moreover, contemporary tools are deficient in their inability to automatically and adaptively generate human activities. Effective simulation must be varied, consistent, and customized to each user's profile to align with diverse threat contexts.

The lack of credibility in decoys directly influences the efficacy of the implemented solution. Attackers who detect traps avoid them, thereby enabling them to continue their malicious activities on actual systems. This, in turn, undermines the effectiveness of the defense, which is unable to detect and identify adversaries in a timely manner. The current tools' limited customization and adaptability give rise to predictable and vulnerable environments, enabling sophisticated attackers to evade detection by identifying artificial patterns.

To address these challenges, the tool presented in this work has been designed to automate and customize the automatic profiles of humanized behavior in cyber deception. The tool's design involves the generation of fictitious profiles based on specific roles and the simulation of credible activities. The aim is to enhance the credibility of the decoys and augment the effectiveness of these strategies. The tool's alignment with MITRE Engage not only addresses the deficiencies in human activity simulation but also offers a scalable and flexible approach to cyber deception deployment across various sectors and threat contexts.

PROPOSED TOOL

The tool has been designated BUDA, as an acronym for Behavioral User-driven Deceptive Activities. It is presented as an experimental solution aimed at improving the credibility and effectiveness of cyber deception strategies through the management and customization of "honey profiles." Honey profiles are profiles of fictitious people like real users, and this solution involves including human-like behaviors within the decoy environments. This increases their credibility and makes them more difficult for attackers to detect.

The design is predicated on the capability to generate customized fictitious profiles that interact autonomously with a set of defined base rules (context), simulating typical patterns of legitimate users. This enriches the cyber deception narratives and the fortification of the resilience of defensive operations against more advanced adversaries. The proposed framework leads to an examination of the domain of cyber psychology, which integrates cyber behavioral sciences with adaptive environments to enhance cyber deception. It further investigates and identifies gaps, including a scarcity of empirical assessment and the under-examined effects of organizational culture[10].

The objective of the tool is to overcome the limitations of current cyber deception strategies through four fundamental axes: first, the automation of user simulation, allowing the creation and management of fictitious profiles that realistically mimic human behavior; second, personalization and narrative coherence, adapting these profiles to specific contexts to ensure plausible interactions within the deception environment; third, scalability, facilitating the simultaneous generation of multiple profiles and their integration into diverse environments and industries. Finally, the tool integrates with MITRE Engage, aligning its capabilities with the planning, execution, and analysis phases proposed by this strategic framework.

A. Basic functionalities

The tool's fundamental functionalities are explained below, accompanied by a detailed description of each component.

1) Narrative management: Narratives constitute a pivotal component of a cyber deception operation, delineating the operational guidelines that guide the subsequent activities. These narratives are crafted by the cybersecurity team, contingent upon the strategic objectives delineated for the operation. Consequently, each dummy profile activity is designed to reinforce the narrative underpinning the lure, token, or deception activity deployed. To illustrate this point, consider a scenario in which the objective is to emulate the use of a well-known public service that could potentially be vulnerable to a data leak. In such a case, the profiles can be designed to exhibit interactions with these services, thereby guiding attackers to the conclusion that the service is valid within the environment. This, in turn, enables deception-based detection. The planning process, which is aligned with MITRE Engage, mandates capabilities at each stage of the framework. This facilitates the definition of targets for operations and the selection of mock profiles that are appropriate to the context. This alignment aims for a more structured implementation of tactics.

2) Creation of fictitious user profiles: The definition of fictitious users with specific roles and description of behavioral patterns is provided. These profiles are built from real context data or specific definitions from the cybersecurity team, ensuring that each simulated user is consistent with the environment in which the decoys, or any other deception activity, are deployed. To illustrate, a simulated profile of an IT analyst might encompass activities such as accessing server management systems, while a profile of an administrative employee might involve interactions with an intranet or office automation applications. This flexibility in distinguishing profiles enables their adaptation to diverse industries and operational scenarios, thereby achieving their intended qualities.

3) Automation of activity generation: The tool generates and executes typical user actions, according to defined criteria, in an automated and narrative-consistent manner. These actions replicate everyday behaviors, file accesses, web browsing, database queries, and application usage. This automation reduces the operational burden associated with manual decoy management and ensures that interactions are consistent and plausible. The emulation of human activities seeks to replicate everyday actions to increase the believability of the fictitious profiles. To enhance the realism of the simulation, temporal interaction patterns are implemented that simulate daily routines with natural variations, thus avoiding predictable patterns that could be detected by advanced attackers.

4) Behavioral customization: Each generated profile can be configured to reflect specific daily routines and interactions with decoy assets. Such routines may include work schedules, frequency of access to certain resources, technology preferences, and communication habits. By incorporating this layer of personalization, the aim is to tend to replicate complex human dynamics, thereby increasing the credibility of profiles.

5) Activity and interaction reports: The continuous monitoring and logging of dummy profiles' interactions with the environment facilitates the generation of reports that offer insights into these interactions. These reports empower security teams to adjust their criteria and perpetually enhance their cyber deception operations. The comprehensive logs enable integration with existing security tools, such as SIEMs

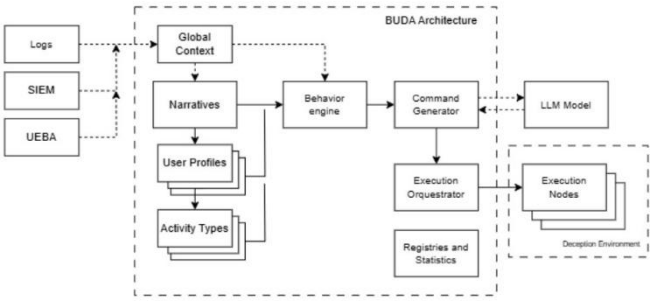
or UEBAs, thereby augmenting their operational efficacy within the established cybersecurity ecosystem.

6) Variability and realism: The capacity to introduce variation based on a "percentage of similarity" definition of the actions and behaviors of dummy profiles in relation to the context and definitions made for each element has been incorporated. This feature thwarts the predictability of activity patterns, thereby mitigating the risk of attackers detecting decoys as fakes. Moreover, it enables the testing of anomaly detection systems based on user behavior. This approach provides an additional use case that adds value to deployed cyber deception operations.

7) Assisted generation of narratives and profiles: This creation capability allows profiles and narratives to be created and configured from a set of context definitions. Each object can then be configured with more detailed attributes, including the attacker profile associated with the narrative, or the basic patterns of the fictitious user profiles to increase their realism and credibility.

B. Technical Architecture

The tool's architecture was designed to be modular, thus facilitating its implementation and adaptation to different environments. The following block diagram illustrates its conceptual structure, which comprises the main configuration modules.



1) Narrative Module

The objective of this module is to facilitate the creation and management of strategic narratives, which serve as the foundation for orchestrating cyber deception operations. These narratives are designed to orchestrate the interactions of fictitious users with decoy assets, aligning the simulation with specific objectives such as detection, dissuasion, or intelligence gathering from the attackers. Integration with cyber deception operations is proposed, ensuring that simulated interactions are consistent and strategically aligned with security objectives. The module's design enables the creation of realistic environments by leveraging attacker profiles, decoy management, and fictitious user allocation. This contributes to the difficulty of adversaries detecting decoys. In addition, the inclusion of visual representations of narratives makes it easier to track and adjust simulations over time and also allows Blue Team operators to optimize their strategies based on the information gathered. The module consists of three key elements, which are discussed in the following sections.

- *Narrative creator*: In the process of designing a narrative, it is mandatory to assign a descriptive and unique name to facilitate its identification within the system. The objectives of the narrative should be clearly defined to ensure alignment with the organization's cybersecurity strategies. These objectives may include the purpose and scope of the scenario. The “percentage of similarity” is defined in this step for the activities executed within the context of this narrative. In this context, 100% implies an execution that is fully compliant with the guidelines, and 0% implies an execution that does not follow the defined context and guidelines at all.

- *Additional definitions*: Before proceeding with the execution of a narrative, it is necessary to establish the profile of the expected attacker and define the decoy assets that will be part of the scenario. The former is configured with key attributes that model the expected actions within the deception environment, such as motivation (financial, espionage, hacktivism, etc.), skill level (categorized as basic, intermediate, or advanced), and expected tactics (network scanning, lateral movement, data exfiltration, etc.). Templates based on known threat actor types are provided for ease of configuration, and decoy assets in the environment are also defined for integration into the execution. Deception activities may include purpose-built dummy files (e.g., financial documents, strategic reports), simulated services and systems, and fake credentials created to attract malicious access attempts. Finally, the narrative must relate to an execution node on which activities designed to leave traces on them will be executed.

- *Definition of time limits*: To ensure the operational control of the narrative, it is needed to establish an end date for its execution. This limit can be set in a fixed manner, with a specific date, thereby constraining the duration of the simulation. Alternatively, it can be initiated by a predetermined condition or series of conditions. The assignment of an end date is instrumental in preventing the activity from continuing indefinitely and facilitates the planning of future simulations, as is state of MITRE Engage.

2) User profiles module

The objective of this module is to configure fictitious users to interact with execution nodes based on cyber-deception narratives. The module is composed of four elements, which are detailed below.

- *Profile creator*: This feature enables the design and customization of fictitious users, encompassing fundamental attributes such as their nomenclature and the roles they fulfill within the organizational structure (e.g., Finance Manager, IT Analyst). Additionally, each profile encompasses behavioral attributes that delineate work schedules, including the designated start and end of the workday, designated periods for lunch and breaks, and designated rest periods. Daily routines are also configured, determining common activities such as file access, participation in virtual meetings, and utilization of internal tools. The frequency of execution of these tasks can be adjusted in levels (low, medium, or high)

according to the role and needs of the simulation. To avoid detection by predictable patterns, a margin of randomness in schedules, access, and activities can be configured. Users can be created manually or assisted by querying the LLM module integrated in the platform. Each profile can be assigned to one or more narratives, to allow the mobility of its context to other scenarios, as occurs in reality. This facilitates the adaptation of narratives to different threat situations and optimizes configuration time.

- *Profile library*: This dynamic repository facilitates the management, reuse, and modification of existing profiles, offering an organized view of all available profiles with search options by name, role, or type of activity, thereby facilitating their management in complex scenarios. Each profile can be edited, duplicated, or deleted according to the needs of the simulation, allowing for quick adaptation to different operational contexts. Additionally, profiles can be exported in standard formats (JSON and CSV), facilitating integration with other tools. A version control system is included, which keeps a history of changes in each profile and allows auditing modifications, ensuring the consistency of configurations, and restoring previous versions.

- *Activity simulation*: This automates the actions of the fictitious users, thereby reinforcing the authenticity of the decoys. To this end, it generates defined interactions with execution nodes, including frequent access to files, databases, and internal systems. Profiles can also interact with corporate tools (CRM, ERP, office suites, etc.) by replicating common usage patterns. Other actions may include operating system interactions, such as logon and logoff, reboots, and more.

- *Variability editor*: The " percentage of similarity" is a defined metric that is used to adjust the alignment or randomization of the simulated behaviors. This adjustment is made to propose a shift from the baseline. In order to improve credibility, the editor allows the modification of the margins of the hourly definitions by introducing fluctuations in the activity times and adjusting the use of defined applications to ensure variability in the interactions.

3) Activity automation module

The objective of this module is to execute actions within the systems, with these actions being based on the cyber-delusion narratives and the profiles that have been created. The module is composed of four elements, which are detailed below.

- *Activity manager*: This module enables the management of diverse activities to be executed, with the specific assignment for each user delineating the basis for activity generation. Such activities may include file access, queries to database engines or services, and the utilization of applications such as web browsers or office tools. During the execution of narratives, these types of activities are taken into consideration, with the objective being to limit the interactions of fictitious profiles and to simulate a real user within their work environment. Accesses, modifications, and information queries are exclusively executed with the user's credentials and on the designated execution nodes, without exerting any influence on systems external to the narrative.

- *Command generator*: This feature translates the types of activities into executable commands for the nodes. This is achieved through integration with a large language model via API (e.g., OpenAI, Gemini) or through local deployment of language model deployment platforms (e.g., Ollama, LM Studio). This integration enables the dynamic generation of adapted instructions. To ensure the usability and reliability of the responses generated, prompt formulation strategies are employed that minimize ambiguities and divergence in the interpretation of requests, ensuring that commands are consistent with the narrative. This is achieved through clear structuring of prompts, inclusion of relevant context, and validation of responses prior to execution, thus avoiding unintended actions or inconsistencies.

- *Execution dashboard*: A centralized interface has been developed for the monitoring and control of automated activities in real time. This interface provides a comprehensive view of the active narratives, along with a detailed breakdown of the actions in progress and the fictitious users involved. This functionality enables operators to effectively monitor the evolution of the simulation and make necessary adjustments. The interface is equipped with a manual control system that allows activities to be halted or resumed at any time, and it can also inject additional commands into the ongoing operation. Finally, the system maintains a comprehensive activity history by systematically logging the actions of dummy users and the results of each operation, facilitating auditing and further analysis. Such analysis can include assessing the effectiveness of the deception model, understanding attacker interaction patterns, and improving the simulation strategy in future deployments.

4) Monitoring and reporting

This module facilitates the monitoring and analysis of interactions generated by fictitious users, as well as the response of the attackers. Its primary function is to provide real-time visibility into activities within the environment and to enable the collection of information to assess the effectiveness of the simulation. It comprises a panel with statistics that offers a centralized, real-time perspective on interactions within the execution environment. This panel enables operators to supervise the actions of fictitious users, including but not limited to file access, application use, email sending, and any other automated activity programmed by design. In conjunction with the execution panel, these tools facilitate comprehensive visualization of activities.

C. Implementation

The subsequent section details the implementation of the tool, including its installation and deployment process, basic use cases, and a proposed methodology for evaluating its effectiveness.

Although the tool can be employed independently and without a contextual framework, a methodological approach to its implementation is suggested to achieve more precise and functional results. This allows the complexity inherent in the application of cyber deception measures to be addressed in a structured manner, while maintaining an agnostic perspective with respect to the technologies used. This aspect is particularly salient given that, in principle, it is feasible to implement cyber deception strategies without resorting to technologically advanced tools. However, a methodological approach demands more effort from the work team, although it offers greater depth in terms of experience generated and knowledge acquired. A systematic approach optimizes the implementation of strategies and contributes to the generation of a reproducible and scalable framework, adapted to the needs of each organization. For those interested in a simplified methodology, we recommend consulting the bibliography[6].

1) Deployment

The implementation of the tool, which was developed in Python language, can be done through a simple *pip* package, which can be installed by executing the command:

```
$ pip install BUDA.
```

Prior to this, it is recommended to create a virtual environment to circumvent dependency issues, utilizing the following command:

```
$ python3 -m venv venv
```

For the documentation of the project, *Read the Docs*, a free open-source software documentation hosting platform that produces documentation written with the Sphinx documentation generator, was used. The documentation can be found at: *https://budaframework.readthedocs.io*. Similarly, the source code is published for free use under GPLv3 license in an open source repository on the GitHub platform[11], which can be found at the web address indicated in the references.

Despite its minimalist approach, there are several technical and operational challenges to its implementation. First, the risk of detection by highly knowledgeable adversaries must be considered, as the simulated activities must be sufficiently realistic to prevent artificial patterns from being identified. Second, the consistency and maintenance of honey profiles requires regular updates to adapt to changes in organizational dynamics, ensuring that profiles maintain their credibility. Third, integration with existing security infrastructure without generating additional operational overhead and ensuring compatibility with existing threat monitoring and analysis tools must be addressed. Finally, scalability in large environments can be complex, as the system must manage multiple profiles simultaneously without compromising network performance or affecting the stability of the environment. Overcoming these challenges is key to maximizing effectiveness in real-world contexts.

2) Use cases

- *Attack deflection and early detection:* The generation of realistic activities on fictitious profiles on productive workstations has been demonstrated to direct attackers toward decoys, thereby facilitating the early detection of such activities and diverting their efforts from compromising critical assets.
- *Validation of behavioral monitoring systems:* The tool utilizes an automated simulation process that replicates the interactions of actual users, as defined by a predefined percentage of similarity. This process enables the evaluation and calibration of monitoring solutions, such as SIEM or UEBA. The tool's functionality includes the refinement of detection parameters and the adjustment of sensitivity to abnormal activities. Consequently, it enhances the capability to identify malicious actions and reduces false positives.
- *Refinement of cyber deception tactics:* The tool can be utilized to experiment with an array of interaction patterns, manipulating the temporal parameters and consistency of the simulated actions. This facilitates the refinement of deception tactics and the adjustment of the "personality" of fake profiles, thereby ensuring their adaptation to novel attacker techniques over time without compromising realism.

3) Evaluation Methodology

Given its status as an experimental tool, a structured methodology is necessary to evaluate and validate its impact on cyber deception strategies. To this end, we propose a three-pronged evaluation approach, encompassing the following domains: First, the credibility of the honey profiles, aimed at ascertaining the extent to which the simulated profiles are perceived as legitimate by attackers, thereby minimizing deception detection. Second, the attack detection and deflection domain, which aims to assess the tool's capacity to identify malicious activities and redirect them to decoys. Third, and finally, the automation and scalability domain, which seeks to examine the system's efficiency in generating and maintaining multiple profiles without the need for significant manual intervention.

To this end, it is recommended to implement a simulated environment with controlled testing prior to deployment in real environments. The test scenario would include an infrastructure that mimics a corporate network with multiple interacting digital assets, honey profiles, and simulated attackers, such as Red Team teams, attempting to compromise the network. Experiments will consist of launching attacks and analyzing how the tool influences attackers' decisions regarding decoys.

Key performance metrics, such as the detection rate (the percentage of attack attempts identified), attacker dwell time (the duration of interaction with decoys before detection of the trap), and the level of interaction with honey profiles (the frequency of attackers interacting with fictitious profiles instead of real systems), can be defined to assess performance. Additionally, detection evasion can be evaluated based on the number of attackers identifying

decoys as fake and the operational load required to maintain the tool's efficiency.

The effectiveness of the tool can be evaluated using different approaches. In the controlled attack simulation, a Red Team will attempt to compromise the network by interacting with decoy assets and honey profiles. The results will be compared with scenarios without their use to determine their impact on early threat detection. Through log analysis and telemetry, activity logs will be collected to identify evasion patterns and differences in network traffic between environments with and without it. Additionally, cybersecurity experts will assess the credibility of the generated profiles, verifying whether they are indistinguishable from real users, which will allow validating the quality of the narratives. To quantify the added value, a comparison will be made with traditional cyber deception strategies. A comparative analysis can be conducted to assess the effectiveness of standard honeypot and decoy systems, as well as those that integrate generated honey profiles. This analysis will facilitate the determination of whether the incorporation of simulated actions enhances the credibility of decoys and their capacity to detect and deflect threats. While the experimental approach provides valuable insights, it is important to acknowledge its limitations due to its current stage of development. As it has not been validated in real environments, laboratory tests may not fully reflect the behavior of attackers in productive scenarios. The adaptability of the experimental design to different contexts remains uncertain, as its effectiveness may vary according to the specific sector or infrastructure where it is implemented. In addition, the experimental design could include biases if the selection of simulated attackers does not accurately represent the diversity of tactics employed in genuine attacks. To mitigate these risks, it is recommended to move towards testing in operational environments, exposing it to unknown actors and more complex scenarios. The results of the initial experiment seek to confirm several key hypotheses about the potential. The experiment will attempt to demonstrate that user simulation significantly increases the credibility of decoys, making them more difficult for advanced attackers to detect. It will also seek to validate whether it can detect threats at an early stage before they compromise real systems, which would represent an advance in incident prevention and response. Finally, the experiment will evaluate whether automation allows for scalable and efficient implementation in large and complex environments, ensuring its operational viability in organizations of different sizes and sectors. These evaluations will ascertain the impact in the field of cyber deception.

4) Expected benefits

The objective of the tool is twofold: first, to provide enhanced defensive coverage, and second, to augment threat intelligence gathering capability. This is achieved without imposing a substantial operational burden, thereby circumventing the limitations of current methodologies. The following benefits can be anticipated:

- *Increased credibility of decoys*: The use of honey profiles has been demonstrated to introduce human-like activity patterns into deception environments, thereby impeding the capacity of attackers to discern between decoys and authentic systems. This augmentation in the perceived authenticity is attributable to the simulation of digital interactions.

- *Improved attack detection*: The identification of suspicious activities occurs at an early stage, as attackers interact with profiles and simulated data that motivate their tactics, techniques, and procedures (TTPs) to be detected before compromising real assets.

- *Scalability and reduced operational costs*: The implementation of automation has the potential to eliminate the need for manual creation of realistic decoys, thereby enabling the concurrent deployment of multiple profiles with reduced operational expenditures. Additionally, its capacity to integrate with existing cybersecurity tools, such as SIEMs and threat intelligence platforms, contributes to the simplification of deployment processes and the facilitation of adaptability.

LIMITATIONS

The practical implementation of the tools is subject to some limitations. First, the possibility exists for attackers to identify common patterns in profiles, which could lead to the development of evasion mechanisms to ignore these decoys. Some studies have shown that advanced attackers are able to identify common patterns in automatically generated profiles [12]. To address this challenge, it is imperative to implement continuous diversification of honey profiles and enhancements to the activity generation engine. These measures are crucial to ensure that interactions remain difficult to predict, even in the presence of sophisticated adversaries.

Second, simulating sustainable narratives poses some challenges. The creation of profiles that appear authentic at both the initial stage and over time requires consistency in the interactions. These must be coherent with the specific environment where they are employed, avoiding any inconsistency that could compromise the authenticity of the simulation. The integration of artificial intelligence models during other phases of the process can enhance the authenticity of the profiles; however, this introduces new challenges related to their training and optimization.

Finally, the incorporation of these profiles into real environments poses a significant obstacle, as their integration into production infrastructures necessitates rigorous testing to ensure compatibility with existing systems and security over potential exploitation of the service itself. Furthermore, given the unique characteristics of each organization, including network architecture, industry sector, and threat profile, additional adjustments may be necessary to address the specific needs of each case, thereby increasing operational and configuration complexity. These challenges and constraints underscore the necessity of an iterative and flexible approach to tool development and deployment..

FUTURE WORK

In order to establish itself as a viable instrument in the realm of cybersecurity, a series of research and development

initiatives are proposed. Primarily, the implementation of evaluations in authentic organizational environments, with the objective of assessing its efficacy in the face of authentic attacks and the aggregation of data concerning the performance of honey profiles in cyber-breach operations.

Second, the necessity to enhance the simulation of human behavior is underscored. This enhancement is to be achieved by the integration of artificial intelligence and machine learning models, which are designed to generate adaptive profiles. Additionally, there is a call for the incorporation of interactions based on real data analysis, a measure that will render the profiles less susceptible to prediction.

Third, the optimization of performance and scalability is proposed, entailing the design of a distributed architecture capable of supporting multiple profiles in large networks without degrading the system. Additionally, the impact of this architecture in environments with different levels of traffic and activity is to be evaluated. Finally, the integration of the proposed system with incident response strategies can be explored, analyzing how to contribute to automate the identification and mitigation of threats in real time, as well as linking with early warning systems and digital forensics..

DISCUSSION AND CONCLUSIONS

In the contemporary landscape of cybersecurity, the paradigm of cyber deception has emerged as a pivotal component within active defense strategies, particularly in the context of countering sophisticated threats. The efficacy of this approach, however, is contingent upon the credibility of the decoys employed. The proposed tool signifies a novel approach aimed at mitigating the disparity between digital asset simulation and the scarcity of authentic interaction within deception environments. Though still in its nascent stages of development, this work lays the foundation for the creation of new tools that seamlessly integrate automated human behavior into defensive strategies.

Despite the long-standing development of cyber deception technologies and the inherent reluctance of organizations to adopt them, the current state of the art allows for experimentation with these technologies. Our tool offers an advantage over commercial tools that often require significant financial investment only to start experimenting, and also enables the testing of techniques and the exploration of cyber deception as a gateway to acquiring knowledge on a high-impact topic in today's world. This work encourages the development of cyber deception strategies that are independent of commercial products, using simplified methodologies and open-source tools.

In conclusion, the objective of this work is to contribute to the proposal of an architecture that integrates automatic interactions on honey profiles to add realism to the operation of cyber deception. This is achieved through a modular architecture design, which includes generating fictitious profiles, simulating human-like activities, and analyzing interactions with attackers based on the behaviors of existing users. This is part of a growing trend in cybersecurity, which is leveraging artificial intelligence and automation to improve active defense.

REFERENCES

[1] F. Pacheco, "Active cyber defense: working model for defensive strategies based on the adversary's error".

[2] M. Morovitz, G. Raymond, S. Barr, and L. Anderson, "MITRE Engage Framework," Feb. 2022. [Online]. Available at: https://engage.mitre.org

[3] M. A. Wani, S. Jabin, G. Yazdani, and N. Ahmadd, "Sneak into Devil's Colony- A study of Fake Profiles in Online Social Networks and the Cyber Law," on22 March 2018, *arXiv*: arXiv:1803.08810. doi: 10.48550/arXiv.1803.08810.

[4] A. Gurtu and D. Lim, "Use of Artificial Intelligence (AI) in Cybersecurity," in *Computer and Information Security Handbook*, Elsevier, 2025, pp. 1617-1624. doi: 10.1016/B978-0-443-13223-0.00101-6.

[5] D. Liebowitz *et al.*, "Deception for Cyber Defence: Challenges and Opportunities," in *2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, Atlanta, GA, USA: IEEE, Dec. 2021, pp. 173-182. doi: 10.1109/TPSISA52974.2021.00020.

[6] F. Pacheco and D. Staino, "Proposal for implementation of minimalist cyber deception strategies."

[7] T. J. Shimeall and J. M. Spring, "Deception Strategies," in *Introduction to Information Security*, Elsevier, 2014, pp. 61-79. doi: 10.1016/B978-1-59749-969-9.00004-3.

[8] Z. Aradi and A. Bánáti, "The Role of Honeypots in Modern Cybersecurity Strategies," in *2025 IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Stará Lesná, Slovakia: IEEE, Jan. 2025, pp. 000189-000196. doi: 10.1109/SAMI63904.2025.10883300.

[9] L. Zhang and V. L. L. L. Thing, "Three Decades of Deception Techniques in Active Cyber Defense -- Retrospect and Outlook," *Comput. Secur.* vol. 106, p. 102288, Jul. 2021, doi: 10.1016/j.cose.2021.102288.

[10] K. Ferguson-Walter, S. Fugate, C. Wang, and T. Patel, "Introduction to the Cyber Deception and Cyberpsychology for Defense Minitrack."

[11] *BUDA (Behavioral User-driven Deceptive Activities)*. [Online]. Available at: https://github.com/Base4Security/BUDA

[12] K. Ferguson-Walter, M. Major, D. Van Bruggen, S. Fugate, and R. Gutzwiller, "The World (of CTF) is Not Enough Data: Lessons Learned from a Cyber Deception Experiment," in *2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC)*, Los Angeles, CA, USA: IEEE, Dec. 2019, pp. 346-353. doi: 10.1109/CIC48465.2019.00048.