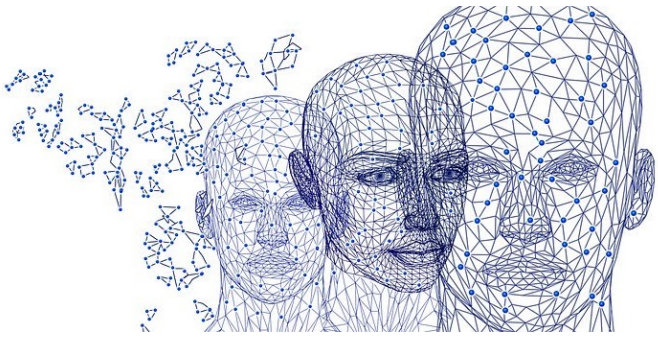
March 2022, Issue Code NL-2022-7

DDD (Digital Data Deception) Technology Watch Newsletter

Table of Contents



"*All warfare is based on deception. Hence, when we are able to attack, we must seem unable; when using our forces, we must appear inactive; when we are near, we must make the enemy believe we are far away; when far away, we must make him believe we are near.*"

— Sun Tzu, *The Art of War*

Editors: Haiyue Yuan, Yichao Wang, Enes Altuncu, Ali Raza, Virginia Franqueira, Shujun Li and Sanjay Bhattacherjee
Affiliation: Institute of Cyber Security for Society (iCSS), University of Kent, UK
Contact Us: ddd-newsletter@kent.ac.uk

Editorial

Unsurprisingly, Digital Data Deception (DDD) has emerged as being very prominent in the conflict between Russia and Ukraine. Synthesised videos involving both parties were released on social media, as reported by the BBC. However, the videos were not technically sophisticated and therefore were easily spotted as *deceptive* by many. While the Russia and Ukraine related videos are not sophisticated, empirical research by Nightingale and Farid [127] indicated that modern machine learning techniques such as generative adversarial networks (GANs) are now mature enough to create synthesised faces that are indistinguishable from, and even regarded by average consumers of online content as more trustworthy than real ones. Another example is the availability of off-the-shelf software tools that can allow a layperson to easily swap faces in real-time, e.g., in video streaming and video calls, which adds more complexity and difficulties in distinguishing synthetic content from real content. Such capabilities in a cyber war context make it more challenging to unpick the provenance of false information campaigns and to enforce accountability, as discussed by Elliott [52].

This final issue of the DDD Technology Watch Newsletter series focuses on four distinct, but often interconnected, themes that have been identified during our work for the previous issues but were not covered or covered only very sparsely. These themes are: (1) *Moving Target Defence as a Deception Strategy* (Section 1), (2) *Deception in Crime-as-a-Service* (Section 2), (3) *Deception in Cyber-Physical Systems* (Section 3), and (4) *Deception with Inaudible Voice Commands* (Section 4). To keep the final issue more open for inspiring future work, we decided to not follow a rigid methodology for selecting papers for this issue. Instead, the editorial team met to agree on the selected themes, the structure of each theme, and general principles on how papers should be identified. Then, each section was assigned to one main editor who led the identification of English research papers and worked with the Chinese team members to identify relevant Chinese papers. Rather than using editorial comment boxes related to different papers, each theme is concluded with a "Research Challenges & Directions" section where we provide more structured and high-level comments beyond a single research paper.

This final issue of the newsletter series follows the six most recent issues that extensively cover several important topics of DDD, i.e., GAN based deceptive methods (NL-2022-1 and NL-2022-2), textual DDD – especially those based on modern Natural Language Generation (NLG) techniques (NL-2022-3, NL-2022-4 and NL-2022-5), and also three other more isolated topics – *Fake Data Injection*, *Fake News, Fake Reviews and Fake Accounts*, and *Deceptive Network Topology* (NL-2022-6).

Following the coverage on defensive deception based on false network topology in the previous issue, we explored this further in this issue. Particularly, defensive deception is at the core of the Moving Target Defence strategy and honeypots, which are covered in two sections of this issue. One promising research direction of defensive deception is the use of cognitive tricks to discourage or slow down progress of cyber attackers [58]. For example, Gutzwiller et al. [67] discussed *human attention allocation* to overload attackers, while Olivola [132] proposed to use the *sunk-cost fallacy* to induce changes in choices of cyber attackers.

We hope you enjoy reading this final issue of the newsletter series. Feedback is always welcome and should be directed to ddd-newsletter@kent.ac.uk. Since the newsletter series will come to an end, we are particularly interested in any ideas for future research problems and collaborations. Please do get in touch if you are interested in discussing DDD with us for possible collaborations.

List of Acronyms

- AE: Adversarial Example
- AI: Artificial Intelligence
- AM: Amplitude-Modulated
- APK: Android Package
- ATM: Automated Teller Machine
- BPH: Bullet-Proof Hosting
- BS: Bayesian Stackelberg
- CaaS: Crime-as-a-Service
- CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart
- CNN: Convolutional Neural Network
- CPC: Controller Placement Camouflage
- CPS: Cyber-Physical System
- CT: Control Theory
- DDoS: Distributed Denial of Service
- DL: Deep Learning
- DNN: Deep Neural Network
- DNS: Domain Name System
- DoS: Denial of Service
- FMCW: Frequency Modulated Continuous-Wave
- FR: Facial Recognition
- GA: Genetic Algorithm
- GAN: Generative Adversarial Network
- GPS: Global Positioning System
- GT: Game Theory
- HTML: HyperText Markup Language
- IIMG: Incomplete Information Markov Game
- IP: Internet Protocol
- IPv6: Internet Protocol version 6
- IPv4: Internet Protocol version 4
- IoT: Internet of Things
- MG: Markov Game
- ML: Machine Learning
- MTD: Moving Target Defence
- NLG: Natural Language Generation
- NTP: Network Time Protocol
- OS: Operating System
- PVA: Phone Verified Account
- RL: Reinforcement Learning
- SCADA: Supervisory Control and Data Acquisition
- SDN: Software-Defined Networking
- SDR: Shuffling, Diversity, and Redundancy
- SEP: Search Engine Optimisation
- SMS: Short Message Service
- SSDP: Simple Service Discovery Protocol
- SVM: Support Vector Machine
- UAV: Unmanned Aerial Vehicle
- UPnP: Universal Plug and Play
- VCS: Voice Controllable System
- VM: Virtual Machine
- XAI: eXplainable Artificial Intelligence

1. Moving Target Defence as a Deception Strategy

1.1. Introduction

Moving Target Defence (MTD) is an adaptive, proactive, and dynamic defence mechanism to prevent attacks by constantly shifting the underlying system configurations. By increasing the uncertainty and complexity for any attacker, the objective is to lower the chances for the attacker to identify system vulnerabilities and increase the cost in launching attacks [7, 32, 156]. Researchers have been investigating the differences between MTD and deception. The major difference identified is that deception techniques are normally more aggressive, meaning that false information is intentionally presented to mislead attackers [32]. However, considering the common objective for both approaches is the same, we consider MTD as one type of deception strategy.

The remainder of this section is structured as follow: Section 1.2 introduces related work on the three design principles for MTD, i.e., (1) *What to move*, (2) *How to move* and (3) *When to move*, and MTD techniques developed based on these principles; Section 1.3 reviews some modelling theories that have been applied to the design MTD; Section 1.4 presents work that describes how MTD techniques have been applied in different domains; Finally, Section 1.5 discusses the challenges and directions for future work.

1.2. MTD design principles

In resource-constrained environments (e.g., Internet of Things (IoT) and wireless networks), the deployment of adaptive and intelligent MTD is essential to extend system lifetime and increase system reliability by adequately allocating defence resources [32]. Many MTD are following the fundamental design principles proposed by Cai et al. [21].

What to move refers to the system configuration parameters that can be dynamically modified to mislead attackers. These parameters can be, e.g., Internet Protocol (IP) addresses [13, 162, 187], instruction sets (e.g., machine instructions of a system, can be hardware or software) [96, 141, 142], and Operating Systems (OS) [103, 174]. The modification of the system parameters will lead to a change on the attack surface, resulting in an increased uncertainty and complexity to the attackers. Sengupta et al. [156] elaborated on this from the perspective of an attacker. At an abstract level, "what to move" represents four surfaces that can be exploited: 1) attack surface; 2) exploration surface; 3) detection surface; and 4) prevention surface – as shown in Figure 1.

How to move refers to how to change those moving parameters to further increase uncertainty and unpredictability [69]. There are three main techniques for that: Shuffling, Diversity, and Redundancy – collectively referred to as SDR [32, 60, 76, 168].

- The *Shuffling* technique randomises or rearranges system configurations (e.g., IP addresses mutation, and dynamic migration time adjustment for Virtual Machines (VM)). Eventually, shuffling-based MTD could delay or prevent attackers from accessing a target system [24, 83].
- The *Diversity* technique deploys system com-

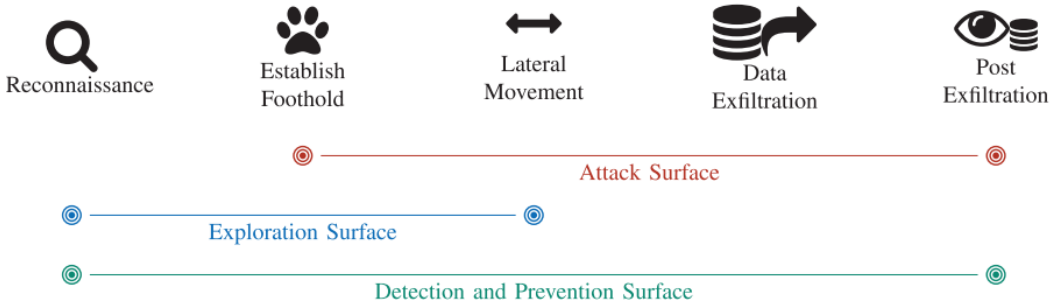


Figure 1: Representation of "what to move" in the form of four surfaces that an attacker can exploit; the detection surface and the prevention surface are represented together (Figure 3 in [156]).

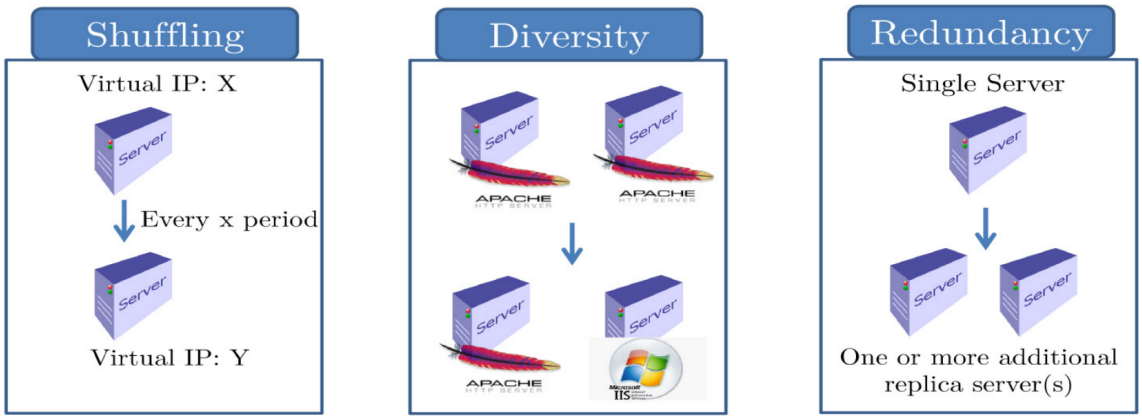


Figure 2: An illustration of three main MTD techniques: Shuffling, Diversity, and Redundancy (SDR) (Figure 1 in [32]).

ponents that perform the same functions but with different implementations to enhance system resilience in the presence of attackers. Examples include network topology diversity [29], software stack diversity [80], and code diversity [20].

- *Redundancy* technique relies on multiple replicas of system/network components with the same functions (e.g., software components redundancy [186], routing path redundancy [4], VM redundancy [89]) to increase system dependability and enhance the system resilience, mainly concerned with denial of service types of attack [76].

Figure 2 illustrates the SDR techniques. They are also often combined to provide hybrid MTDs [5–7, 147, 168], which can enhance security while decreasing defence cost or service interruptions. However, the potential drawback of using hybrid approaches is to introduce a larger attack surface compared to a single MTD technique [32].

When to move refers to the optimal time to change the MTD system's state to invalidate or discard information/progress made by an attacker. Three main approaches have been used to decide "when to move".

- The *time-based approach* [32] periodically changes the attack surface such as port/IP addresses swap [24, 49, 60], OS rotation [103, 174], VM migration ([49, 175]) [24, 33, 155] on a schedule with either fixed or random intervals. Cho et al. [32] also referred to this approach as proactive adaption. It is noticeable that a too-long time interval could allow attackers to have enough time to prepare for an attack, whereas a too-short time interval could wastefully trigger MTD even when there is no attack [32]. Thereby, it is important to determine the correct time interval to design useful and efficient MTD [24, 32, 156]. Thompson et al. [174] evaluated the performance of OS rotation using various time interval ranging from 60-300 seconds, and demonstrated that the time interval of 60 seconds was able to thwart network mapping attacks. In addition, research also has been conducted to derive the time interval in an adaptive way using historical data [43, 133].

- The *event-based approach* [32], also referred to as "on-event switching" [156], executes an MTD operation when a certain event occurs. Due to its property, this approach is also known as reactive adaptation [32]. The key strategy of this approach is to predict potential attacks from incurred events and subsequently trigger appropriate MTD operations. Adaptive MTD implementations have been proposed based on different modelling techniques such as Machine Learning (ML) [37], Game Theory [106, 207], Genetic Algorithms [41, 92], and Control Theory [32, 151], reviewed in the next section.

- The *hybrid approach* combines both proactive and reactive approaches to perform MTD op-

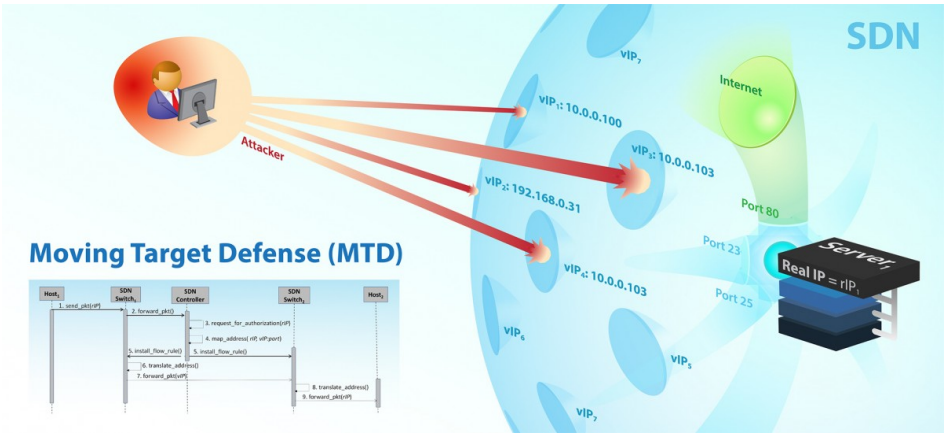


Figure 3: SDN-based MTD [162] according to a U.S. Army illustration [11].

erations in an adaptive way [35, 97, 142, 171, 196, 208, 209].

1.3. Modelling MTD

The design and development of MTD techniques have been facilitated by the adoption of different modelling theories such as *Game Theory*, *Genetic Algorithm*, *ML*, and *Control Theory*. The goal is to help generate the best possible MTD strategies especially for the design principles "when to move" and "how to move". A few examples have been given in the previous section. The remainder of this section reviews existing works that utilise the advantages of these theories to design adaptive and advanced MTD.

1.3.1. Game Theory

The fundamental idea behind MTD is to add another layer of defence via manipulation of the attack surface to enhance system security which, at the same time, could result in extra cost (e.g., system reconfiguration cost and service unavailability) to the users [32]. In the context of gain and loss, Game Theory (GT) is very relevant to model and design MTD as a game between an attacker and a defender. From the defender's perspective, the aim is to identify optimal system configurations in order to effectively and adaptively shift the attack surface. Whereas, the main aim from the attacker's perspective is to launch attacks with minimum effort/time and maximum effect.

One GT-based approach to developing MTD techniques is to use a general game framework, where the assumption is that both the attacker and defender are rational and the common goal is to maximise their utility/payoff respectively. Therefore, the best strategy can be selected based on the estimated gain and loss. For example, Zhu and Başar [207] modelled a game between an attacker and a defender aiming to minimise the risk and maintain service availability by continuously changing the defensive strategies based on information learned dynamically. Carter et al. [25] used GT to derive an optimal migration strategy by analysing temporal platform migration patterns. Neti et al. [126] adopted an anti-coordination game to investigate the scalability of risks for an MTD strategy based on Software-Defined Networks (SDNs). GT is used in a scenario where, if one node in a network is compromised, the remaining nodes can decide whether or not to switch to an alternative software/platform. Figure 3 shows another SDN-based MTD [162]. Lakshminarayana et al. [101] proposed an MTD approach to detect coordinated cyber-physical attacks against power grids, where a zero-sum game is used to identify the best subset of links to disturb and protect against a strategic attacker.

There are other approaches to utilise a Bayesian Stackelberg (BS) game to design MTD techniques. In such a game, there are two players: a leader and a follower. The leader takes an action, and the follower's goal is to optimise the payoff of its action based on the investigation of the impact of the leader's action. Feng et al. [57] studied the strategies of information disclosure by defenders to improve the performance of MTD techniques based on the BS game. Research has also been conducted to use BS games for deriving effective switch mecha-

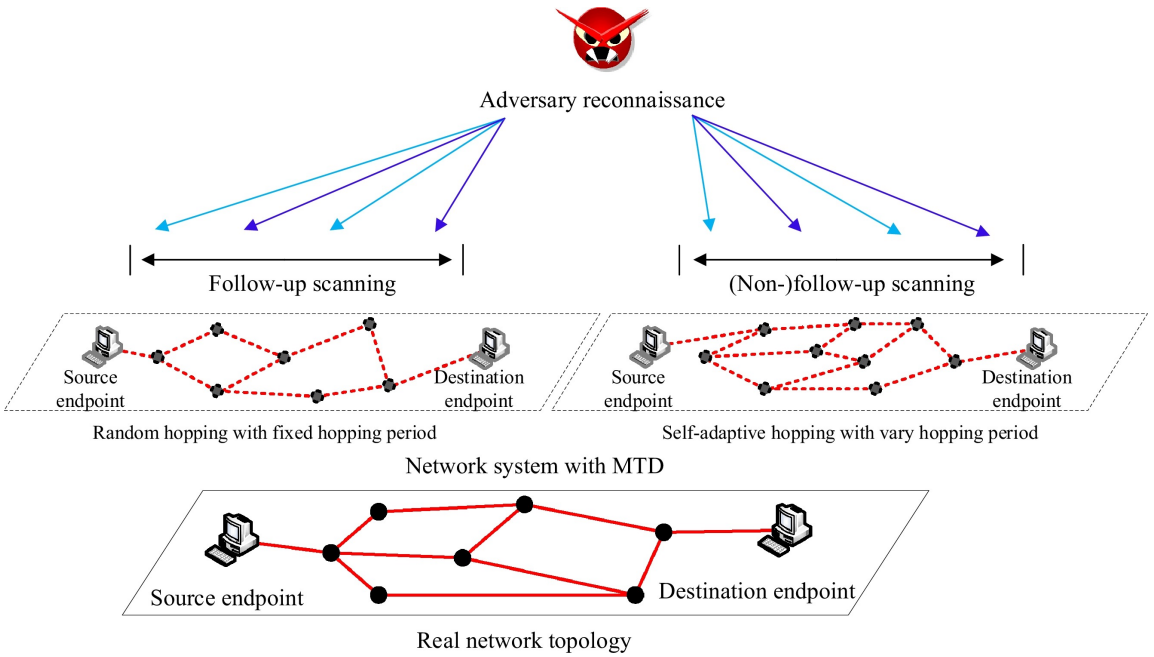


Figure 4: An illustration of Markov games (MG)-MTD and incomplete information Markov games (IIMG)-MTD (Figure 4 in [108]).

nisms for web applications, aiming to maximise security while minimising operational costs from the defender's perspective [157, 179].

In addition, a number of MTD techniques have been developed based on stochastic games [161], which reflect the complex and dynamic relationship between multi-players at different stages based on probabilistic transitions. A two-player stochastic game model was proposed by Manadhata [117] to derive the optimal MTD strategy based on attack surface diversification. A zero-sum stochastic game model [115] was used to design an SDN CPC (Controller Placement Camouflage) to guide MTD solutions. As the stochastic game extends the Markov decision process, it is also called Markov game [18]. A number of approaches based on Markov games were proposed to address the issue of optimal strategy generation/selection for MTD research, such as incomplete information Markov game (IIMG) [108], Markov robust game model [170], Markov time game model [172], and multi-stage Markov signalling game model [91] (see Figure 4 for an illustration of some of such games).

1.3.2. Genetic Algorithms

Genetic algorithms (GAs) are meta-heuristic algorithms for solving optimisation and search problems inspired by biological operators such as crossover, mutation, and selection. They have been proven useful to find the best or near-optimal MTD solutions [32]. Crouse and Fulp [40] modelled a computer configuration as a chromosome, and used GA to identify a secure computer configuration with high diversity in time and space. Their later works [41, 92] also considered the ageing aspect and mutation of computer configuration to improve the GA-based MTD. A series of other evolutionary algorithmic operations, such as reproduction, mutation and recombination, have been adopted to design MTD techniques to identify secure alternative computer configurations over time [40, 113]. Similarly, Collado et al. [39] adopted GA to generate different low-vulnerability configurations for industry-level servers, and Ge et al. [61] used GA to address the "how to move" design principle of MTD and find the best strategy for network topology shuffling on a decoy-populated IoT network. Despite the fact that GA-based MTD has proven to be useful, it is not a trivial task to design MTD solutions to accommodate multiple conflicting goals using GAs. In addition, GA-based MTD approaches are not popular for resource-constrained environments due to their complexity [32].

1.3.3. Machine Learning

Due to the ability to better capture evolving at-

tack patterns, ML-based MTD techniques have been used in a number of scenarios. One of the challenges of MTD is to find an optimal strategy when facing a resourceful and determined adversary [51]. Colbaugh and Glass [37] proposed an ML-based predictive MTD technique to mitigate the attacker's ability to learn about the defensive mechanism by leveraging a reverse-engineering method. Similarly, Colbaugh and Glass [38] proposed an ML-based method using the co-evolutionary relationship between an attacker and a defender to find an optimal defensive MTD strategy against intelligent and adaptive adversaries. Zhu et al. [206] proposed two iterative Reinforcement Learning (RL) algorithms to identify the optimal defence strategy when there is limited information about the attacker. In addition, multi-agent RL algorithms were also used to derive optimal MTD defence strategies for network and web applications [34, 51, 171].

Deep Learning (DL) has attracted a lot of significant interest from both research and industry communities, especially for computer vision tasks. Different from the normal countermeasures, such as adversarial training, adding an input transformation layer, and gradient masking [167], researchers have been exploring the use of MTD to mitigate adversarial attacks [181] to DL models. For instance, DL models are shown to be vulnerable to adversarial attacks such as small but crafted perturbations, which can be added to the clean examples to make the victim DL model produce incorrect classification results [148].

Sengupta et al. [154] developed an MTD framework, called MTDeep, for Deep Neural Networks (DNNs) against adversarial attacks. The authors showed that MTDeep can maintain high classification accuracy on legitimate datasets while reducing mis-classification on perturbed images for both MNIST [45] and ImageNet [44] datasets. Song et al. [167] proposed DeepMTD to detect and thwart adversarial examples by presenting multiple new deep models after system deployment. Qiu et al. [144] proposed a multi-training based MTD to defend against Trojan attacks on DL models deployed on smart devices. Izmailov et al. [87] proposed an MTD strategy with combinatorial boosting of the number of diversified classifiers, which showed promising results for both network intrusion detection and color image classification. He (何康) et al. [72] proposed an MTD technique to improve ML model security to resist evasion attack to detection algorithm by introducing dynamic transformations in terms of algorithm model, feature selection and result output.

A main concern for using ML-based approaches applied to MTD is the need for a large amount of training data to ensure accuracy. In addition, it is essential to ensure that the environment where MTD is to be deployed has sufficient computing power as some resource-constrained setups cannot afford ML-based MTD [32].

1.3.4. Control Theory

Control Theory (CT) is a branch of applied mathematics that focuses on dynamic systems, in which a controller with a transfer function is adopted to control process variables (i.e., inputs and outputs of the system) to ensure that the system is operating correctly. Considering MTD's complex and dynamic features, CT has been adopted to model and analyse MTD systems [32, 107, 195]. Rowe et al. [151] proposed a diversity transformation MTD technique based on CT, where a range of cyber manoeuvre techniques are provided so that the system can select the most appropriate ones to ensure sufficient shifts of attack surface when an attack is detected. Zheng and Siami Namin [203] developed a CT-based approach to identify optimal security policies for MTD deployment. Meira-Góes and Lafortune [120] proposed an MTD technique based on switched supervisory CT to mitigate sensor deception attacks.

1.4. MTD applications

Due to their proactive and adaptive defence mechanism, MTD techniques have been deployed to a number of application domains. This section presents related work on the deployment of MTD to different application domains in terms of the techniques and attacks that can be mitigated.

1.4.1. MTD Applied to IoT

The development and advancement of IoT technologies have contributed to a plethora of innovative applications in various domains. However, conventional mechanisms to ensure security and privacy have shown limitations when applied to IoT environments [32, 125, 149] due to resource constraints and scalability issues. Therefore, researchers turned to MTD to identify alternative and better solutions.

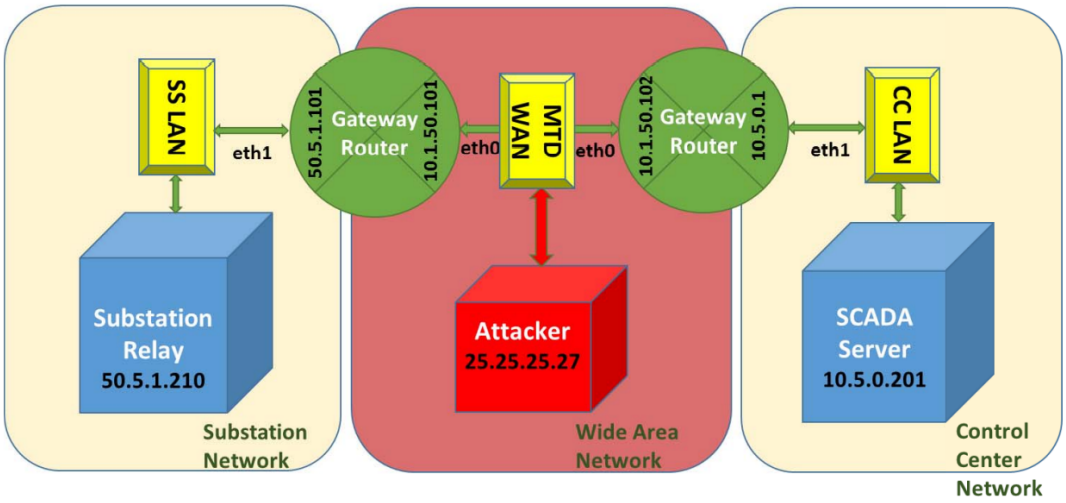


Figure 5: A network diagram of the experimental setup for IP hopping based MTD for power grid SCADA systems (Figure 1 in [178]).

Many MTD solutions for IoT environments were developed based on the shuffling and/or diversity techniques mentioned in Section 1.2; for instance, mutation-based shuffling of cryptosystems [62] and of firmware version for the reconfiguration of devices in an IoT environment [26], network topology shuffling [60], and code partitioning & diversification on IoT devices [152]. MTD techniques based on IP randomisation [164] and rotation of Internet Protocol Version 6 (IPv6) [197, 198] have been shown to be useful to secure IoT from adversary attacks. Study [93] assessed the applicability of IPv4 and IPv6 shuffling based MTD for low-power devices in an IoT environment; results indicated that it is a feasible option to protect Internet-connected embedded systems from attacks. The main attack types considered in the IoT-based MTD techniques include reconnaissance (also known as scanning) attacks [115, 164, 197, 198] and data exfiltration attacks [60].

The deployment of MTD in an IoT environment can stop an attacker at the early stage of the cyber kill chain, which can make it more difficult for attackers to further exploit vulnerabilities and map the devices to launch an attack. However, the network and resources constraints on IoT devices could affect the usefulness and effectiveness of MTD [32].

1.4.2. MTD Applied to Cyber-Physical Systems

Cyber-Physical Systems (CPS) is "a promising new class of systems that deeply embed cyber capabilities in the physical world, either on humans, infrastructure or platforms, to transform interactions with the physical world" [140]. Typical CPS examples can be found in various domains including smart grid, smart home, vehicular systems, and industry control systems [98]. A range of MTD solutions have been proposed to protect CPS in different domains. In the context of Supervisory Control and Data Acquisition (SCADA) systems, examples include: dynamically generated containment systems [31] for threat evasion by generating an individual Docker container for each threat detected, and time-based shuffling (IP addresses) for securing SCADA communications in general [75].

Further work has been done to develop MTD for power grids such as work based on IP address mutation/hopping [134, 178] (see Figure 5). Rahman et al. [146] developed an MTD based on applying controlled randomisation on the power grid system properties that are used in state estimation and the grid topology. Lakshminarayana and Yau [102] proposed an MTD technique for power grid systems that can defend attacks against state estimation using reactant perturbations. Mahmood and Shila [115] developed an MTD technique based on a new concept of CPC to dynamically change the attack surface of SDN controller placements for smart grids.

Furthermore, research has been conducted to develop MTD techniques that are not for a specific type of CPS, but can be applied to different types of CPS, such as the mixed time and event-

triggered architecture based MTD [142], switching-based MTD strategy [112], and deep RL-based MTD (DQ-MOTAG) [27].

Due to the scale and scope of CPS, the associated MTD techniques need to deal with a significant number of attacks such as code injection attacks, eavesdropping and traffic analysis attacks, IP scanning attacks, brute-force login, malicious binary upload, Distributed Denial of Services (DDoS), IP spoofing, relay attacks and false-data-injection attacks [27, 31, 109, 112, 134, 143, 178].

1.4.3. MTD Applied to Other Domains

MTD has also been used in *Cloud Computing*, where different techniques are utilised such as web application stack shuffling [157], VM migration [42, 137, 176] and VM snapshotting [137, 176]. Another popular destination of deploying MTD is *SDN*, which is an emerging technology to decouple the network control plane from the data-forwarding to a networked system. Some common MTD techniques, such as IP shuffling/mutation [78, 114, 162] and network topology shuffling [1, 83], have been used in SDN environments.

1.5. Research Challenges & Directions

Based on state of the art of MTD research, the following items summarise future MTD research challenges and directions.

- Recent surveys [32, 156] provided different classifications of MTD. However, they tend not to map the whole MTD space. Therefore, one of the future challenges is to develop an MTD classification/taxonomy that can capture the multi-dimensions of MTD in a more comprehensive way, consolidating our understanding of the subject.

- Although some research specifically focused on adaptive MTD, the techniques and implementation are still immature [32]. Designing better adaptive MTD techniques will require a multidisciplinary approach involving Psychology, Cyber Security, ML, and Human-Computer Interaction to learn about system vulnerabilities, system security conditions, and behavioural patterns of both attackers and defenders.

- Most studies on MTD techniques focused on a single aspect of the MTD and related security requirements. One research challenge, and an opportunity, is to look at 'the integration and use of full-stack, full-spatiotemporal action space (e.g., VM live migration, OS diversification, hybrid diversity, shuffle, and redundancy actions) in virtualised infrastructure (multiple layers of the software stack) for inherent entropy maximization goal of MTD' [168].

- Research on designing MTD for 5G/6G networks is still scarcely explored. Therefore, another future direction is to design MTD techniques that target 5G/6G. For instance, one topic could be the design and implementation of distributed MTD solutions as 5G/6G network traffic could be processed locally and on the fly at different points [15, 168].

- Current research on MTD do not offer highly lightweight and distributed solutions, which are essential requirements in resource-limited environments, such as military tactical environments and IoT environments [32]. It could be achievable by decreasing the MTD overhead, executing necessary and useful operations (e.g., shuffling and mutations) adaptively according to the system status such as network state, real-time risk and threat analysis.

- Artificial Intelligence (AI), ML and DL based techniques have been heavily adopted to design advanced MTD methods. The classic problems related to AI/ML/DL, such as AI ethics, eXplainable AI (XAI) and their impacts, should also be addressed when designing MTD techniques – this has been largely overlooked so far.

2. Deception in Crime-as-a-Service

2.1. Introduction

As consumers, we always want to get a great service experience, e.g., enjoying a nice dinner at a restaurant. This experience can now be applied to cyber crime. Crime-as-a-Service (CaaS) refers to the sale and provision of tools or information by experienced cyber criminals to others as a service for profit. As the market of CaaS is rapidly growing with a variety of types on offer, law enforcement agencies around the world, such as Europol, have started prioritising action against CaaS between 2022-2025 [53]. This section discusses selected types of CaaS which provide intrinsically deceptive services.

This section is organised in three subsections according to the classification of CaaS, as defined by Akyazi et al. [3] and Huang et al. [79]: *existing services* (Section 2.2), *evolving services* (Section 2.3), and *emerging services* (Section 2.4). Existing services have a well developed and stable business model, and therefore are less likely to change further in the near future. Evolving services refer to services that are currently available on the darkweb/underground forums and very likely to develop rapidly and widely with a new service model due to technology development. Emerging services refer to services that have not been considered as mature services (i.e., it is usually inspired by one or more existing CaaS or a shift from a legit business model) on the darkweb/underground forums. Due to their specialisation and desirability, these services are expected to be widely available and developed. Finally, we discuss future research challenges and directions in Section 2.5.

2.2. Existing Services

Deception-as-a-Service provides vulnerabilities and tools that can be used to generate fake resources (including fraudulent websites, phishing emails, rogue software). Gopal et al. [63] identified four categories of fraudulent websites. (1) *Phishing websites* are responsible for identity theft activities by imitating the structure and interface of legitimate websites. They often contain deceptive URLs and interfaces. (2) *Fake e-commerce websites* provide online shopping stores that do not deliver ordered products, or sell counterfeit goods. (3) *Fake news websites* contain fake or unreliable news. (4) *Piracy websites* often contain a large amount of pirated digital content such as movies, software tools and games. Phishing and scam emails have been used for social engineering attacks [73]. Cyber criminals use fake emails to distribute phishing or malicious software and URLs that trick victims into divulging authentication information. Huang et al. [79] mentioned that when some specific target information is involved, such as a specific company or individual, the attack is called a *targeted attack*. For instance, cyber criminals impersonate the Chief Executive Officer or the Chief Financial Officer and target senior employees within the company to deceive them and successfully obtain employee payroll data [74]. Rogue security software is a popular type of fake software. It typically falsely reports that it found a virus or Trojan on the user's computer and convinces them to pay for a fake malware removal tool. *BraveSentry Variant* [55] is one example; it installs itself in the system via registry keys, making it difficult for users to remove it. Such spyware may also have malicious behaviours such as delivery of ad pop-ups, keylogging, screen capturing, stealing confidential files and downloading malware [205]. Two specific **Deception-as-a-Services**, Reputation-Escalation-as-a-Service for fake reviews, and Phone/SMS-Verification-as-a-Service for fake accounts will be discussed later. Readers are also recommended to read Section 2 of Issue NL-2022-6 of the newsletter series for more about detection and discussion of fake news, fake reviews and fake accounts.

Obfuscation-as-a-Service provides obfuscation services to evade intrusion detection systems or anti-malware solutions for a fee. There are three obfuscation techniques studied by O'Kane et al. [131], namely packers, polymorphism, and metamorphism. Depending on the service provider, they can use one or more mixed technologies [211]. Usually the customer does not need to choose the technology to be used as the obfuscation process is fully automatic. A packer is a tool that can compress, encrypt, and modify malicious file formats. Legitimate software vendors use packagers to bundle executable and support files to manage their software. Therefore, anti-malware software cannot determine whether the software is malware without identifying the packaging algorithm and decompression. Oberheide et al.

[130] noted in their experiments that about 40% of the software could not be unpacked in the 98,801 malware samples they tested. Therefore, the content of that software is hidden. Polymorphism is a cryptographic method that alters static binary code to evade scanning for malware signatures. Furthermore, the malware changes with a different encryption key each time the code runs. Metamorphism is when the malware loads a different action code to maintain its malicious behaviour each time the malware is run. Therefore, traditional signature-based detection methods need to scan a large number of signatures to detect one malware.

With the rapid development of mobile devices, obfuscation services on Android devices have begun to emerge. Its service is in an automated obfuscation platform. The price list ranges from $20 to obfuscate one APK to $850 per month for unlimited obfuscation. This service modifies applications using complex string splitting, decoy strings, and nested junk flow control.

To detect obfuscated source code, there are mainly four different types of code analysis [153]. Pattern matching refers to automated matching to find known code snippets in binary programs. The types of patterns typically range from sample code to regular expression. The signatures of these patterns are created by human analysts and stored in a database to help future queries and comparisons. Static analysis refers to the inspection of executable code but does not run in a virtual machine or real environment. Dynamic analysis is the opposite, which refers to observing the behaviour of software as it runs. Dynamic analysis has been frequently used for malware investigation and forensics [50]. Human-assisted reverse engineering refers to using some automated tools to do the analysis. It should be noted that compared to fully automatic analysis, humans may be subject to some deception, such as being misled by class names and their relationships. Schrittwieser et al. [153] mention that simple, fully automatic obfuscation methods are still effective for non-human analysis methods. Using manual analysis will incur high costs. Obfuscation and analysis of malware turn into an arms race. However, it seems that the existence of such obfuscation services confirms their effectiveness.

Traffic-as-a-Service, which includes **DDoS-as-a-Service** (also known as booters/stress testing service), refers to buyers paying the service provider to attack a specific target through the Internet. Figure 6 shows the process of this service in six steps. (1) The attacker (i.e. the buyer) finds and uses a stresser operator front-end website. (2) Using PayPal or other payment methods, they subscribe or pay for a single stress test service. (3) The attacker uses the front-end website to set a target and to request an attack to start. (4) The attack request is forwarded to back-end servers. (5) The servers send malformed packets with spoofed IP addresses to the amplifiers for traffic amplification. (6) Traffic is directed to the target. Those attacks make use of amplification servers, often mis-configured with the Simple Service Discovery Protocol (SSDP) (as known as Universal Plug and Play (UPnP)), DNS, Network Time Protocol (NTP) and Chargen [95]. They are used to send a large number of spoofed packets to a target.

It was reported that the majority of victims of DDoS-as-a-service were distributed across broadband and hosted networks, with broadband victims accounting for 62% and hosted networks accounting for 26% [129]. Education, government and corporate networks accounted for only a small fraction at 12% [129]. This situation appears to be due to the commoditisation of such attacks, which has resulted in attackers already targeting ordinary users. There is evidence that attackers purchase such services in online games in order to disrupt the opponent's gaming experience or availability [81]. Even with existing technologies that can reduce DDoS attacks, such attacks are becoming more common to the public. Several possible interventions have been proposed by Karami et al. [95]. For example: (1) To reduce the scale by restricting the payment methods, which mainly includes cooperation with large payment companies such as PayPal; (2) To reduce attack efficiency by discovering and repairing related amplified servers; (3) To increase costs for service operators by locating and blocking low-cost hosting services (bullet-proof hosting), so operating this service will make less profit than before; (4) To work with law enforcement in major countries to combat such crimes, which serves as a warning to potential criminals and increases the crime risk.

Traffic-Redirection-as-a-Service refers to a service which allows incoming web traffic to a specific address to be redirected to another website address. This service includes malicious optimisation of the target website to gain more traffic or buy fake clicks from the provider. Search engine optimi-

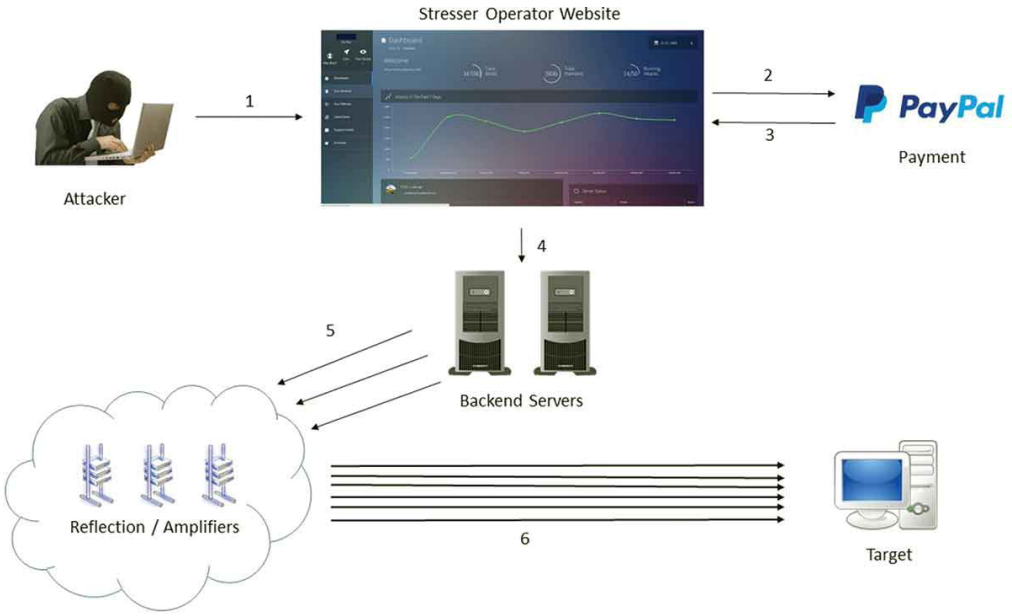


Figure 6: Overview of the process of DDoS-as-a-Service by Hyslip and Holt [84]. (1) The attacker finds and uses a stresser operator front-end website. (2) Using PayPal or other payment methods to subscribe or pay for a single stress test service. (3) The attacker uses the front-end website to set a target and request an attack to start. (4) The attack request is forwarded to the back-end server. (5) Using malformed packets with spoofed IP addresses to amplification servers. (6) Traffic reflected and sent to the target.

sation (SEO) is a set of techniques that can make legitimate websites rank higher for increased visibility. A large number of malicious websites use semantic obfuscation (a blackhat SEO technique that can evade detection) to disguise themselves [190]. Yang et al. [190] summarised the characteristics of malicious SEO pages that use semantic obfuscation as follows. (1) The webpage contains numerous external links to other websites, which may lead to gambling or pornographic websites. (2) Utilising the HTML iframe attribute to hide the original web page and present obfuscated malicious content to human users. (3) Using scripts to determine whether a visitor is a search engine crawler or a human user in order to display different contexts, see Figure 7.

These blackhat SEO techniques are often maliciously applied for search engine poisoning. Wadleigh et al. [180] investigated counterfeit websites on search engines by using 255 queries across 25 brands. Results indicated that, in a popular search engine (e.g., Google), counterfeit websites were found mixed into legitimate search results. This criminal service has been proven to be widely used for malicious activities such as phishing, scam and cryptojacking [77].

Similarly, in addition to the optimisation services mentioned above, service providers may also provide fake traffic (clicks) to target websites. Customers only need to provide a link to their own website. Its charging model is pay-per-click. The estimated price is as low as US$7-$15 per thousand visitors [3].

```
1  <script type="text/javascript" style="display:
       none;">
2      var strRef=document.referrer;
3      var robots=['baidu','google','yahoo','bing
           ','soso','sogou','so','youdao','jike
           ','anquan','360.cn','haosou'];
4      var ishave=false;
5
6      for(var t in robots){
7          if(strRef.indexOf(robots[t])!=-1){
8          ishave=true;
9          if(parent.window.opener){
10             parent.window.opener.location='
                   https://www.tc8806.com/';
11     }
12 </script>
```

Figure 7: An example, provided by [190], of a website using scripts to determine whether a visitor is a search engine crawler or a human user.

Reputation-Escalation-as-a-Service is a type of service which exploits vulnerabilities of current recommendation systems to improve the repu-

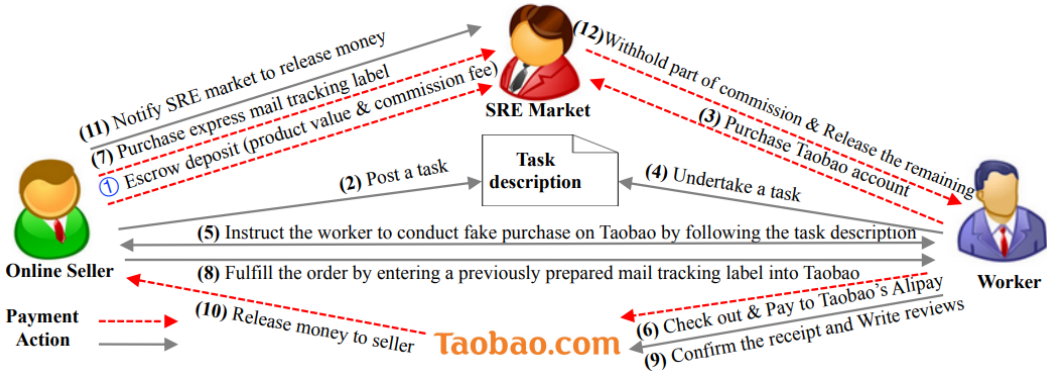


Figure 8: The life cycle of a fake-purchase task on the reputation escalation website, as illustrated in Figure 1 of [185].

tation of e-commerce websites. Figure 8 shows the lifecycle of a fake-purchase task and its escalation effect on the reputation website.

Class	Detail
Goods	Type (physical or virtual), Selling price
Commission	Commission fee offered for this task
Browsing behavior	Search first on Taobao by the keywords given, randomly choose three other stores to browse before finally entering the seller's store.
	Like the store and add it to favorites.
	Stay on the page for 5 minutes and scroll down to the bottom before adding to cart.
	Feign chat with the seller via Taobao's built-in IM program AliWangWang.
Payment method	The worker pays either for herself or using the e-Gift card provided by the seller.
Shipping address	Use the shipping address designated by the seller for the order placed.
Confirmation & reviews	Confirm the delivery and leave good ratings and positive reviews after a predefined waiting time.

Figure 9: A typical task description in Reputation-Escalation-as-a-Service, illustrated by Xu et al. [185]. SRE: Seller-Reputation-Escalation.

Xu et al. [185] concluded that this process is divided into five stages (twelve steps), which are: (i) Task Creation. The online seller (the service buyer) deposits a fee to the operator (Seller-Reputation-Escalation (SRE) market) (step 1) and creates a fake-purchase task on the service operator's website along with some requirements or restrictions (e.g. account registration time in specified e-commerce platform) (step 2). (ii) Task undertaken. If needed, the buyer can even purchase other accounts (often with fake registration information) on the operator's website. This includes step 3 and 4. (iii) Conducting fake purchase. Fake purchases are made on e-commerce sites based on the task descriptions. Figure 9 shows a typical task description by Xu et al. [185]. This includes step 5. (iv) Order fulfilment. The payment method is usually an e-gift card provided by the seller (step 6). For virtual goods, the process can be quickly carried out online. For physical goods, e-commerce platforms usually require information such as courier tracking numbers as a proof that the goods were dispatched. Therefore, the seller would send an empty parcel or buy a tracking label from the reputation escalation service provider. This includes step 7 and 8. (v) Task completed. The buyer leaves a fake good review (step 9). Once the e-commerce platform releases money to the seller (step 10), the commission will be released to the fake buyer (step 11 and 12). The reputation escalation service provider withholds a portion of the fee.

In order to escape detection by e-commerce platforms, several strategies are implemented [185]. Firstly, the service provider limits the frequency of the use of each account and usually requires real-name verification for it. Secondly, depending on the IP address, the shipping address is required to align with the IP. Thirdly, some purchase behaviours are required, such as browsing similar products on e-commerce platforms or browsing other products in the store before purchasing. The timing of writing a positive review often reflects the speed of shipping; therefore, fake buyers should post fake reviews at the right time, depending on location. Potential mitigation strategies include e-commerce platforms working with couriers to detect fake packages, domain names and website hosting to detect such suspicious SRE market domains and take down SRE market sites. A study by [14] shows that, even if a legitimate e-commerce platform increases the cost of this kind of service, it still does not completely prevent shops that benefit from using reputation escalation services. This strategy may also encourage

more shops to use this service. Therefore, intervention methods based on economics need to be studied in the future to tackle the problem represented by such services.

Money-Mule-Recruiting-as-a-Service refers to the service of recruiting people (called money mules) who will be involved in a money-laundering network. Money mules are often required to transfer funds in a variety of ways, including through bank accounts, checks, virtual currency, prepaid debit cards, and more [56]. After the transaction is completed, a commission will be received by the money mules. They are often unaware that they are committing a crime. Criminals lure money mules through promises of employment, such as part-time jobs, working from home, no experience needed, as shown in Figure 10. Money laundering networks for virtual currency are emerging [184].

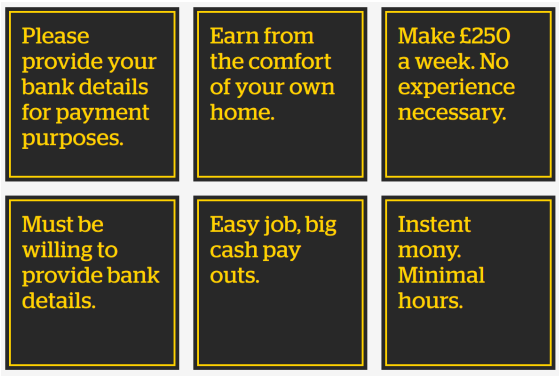


Figure 10: Commonly used words when recruiting money mules, provided by UK Finance and Cifas [177].

Another type of money mule is the so-called reshipping mule scams. Figure 11 shows the process of scams using reshipping mules. Overall, cyber criminals use stolen credit cards to buy physical products at online shops and send such products to the recruited mules (drop). The mules end up forwarding the products to the criminals who then re-sell them on the underground market for cash. Sometimes mules are also victims because the criminals do not pay them a commission, making this kind of crime model more complex. Such crimes often involve multiple countries, making combating them more difficult. A study by Hao et al. [68], which monitored seven reshipping scam sites, revealed that about six thousand packages were forwarded over a nine-month period, generating more than US$7 million in annual revenue.

Bullet-Proof-Hosting-as-a-Service provides hosting services to criminals, i.e., a platform to support, for example, botnets, DDoS attacks and scam distribution services [128]. Noroozian et al. [128] studied the bullet-proof hosting (BPH) platform *MaxiDed*. Interestingly, *MaxiDed* itself is not a Bullet-Proof-Hosting-as-a-Service supplier. However, it attracts upstream suppliers and customers in a particular business model illustrated in Figure 12. *MaxiDed* acts as a middle-man, which allows different abuse and is clearly marked on the sales page. Server packages from those hosting providers are placed on *MaxiDed* by merchants. However, most merchants had no reseller relationship with those upstream hosting providers. The platform collects fees from sales between merchants and customers. Customers need to deposit money on the platform first and then make a purchase. Pre-payment aims to prevent customers from terminating a transaction in the middle, causing loss for the platform. For mitigation measures, other than shutting down the platform itself, it seems difficult to intervene in the supply chain.

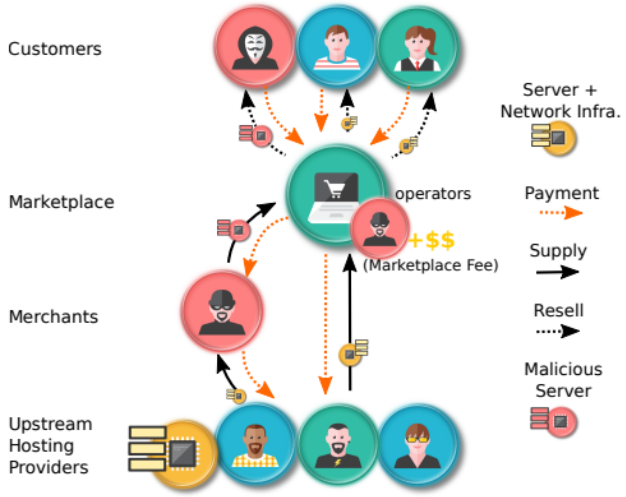


Figure 12: An overview of *MaxiDed*'s business model, illustrated by Noroozian et al. [128]. *MaxiDed* itself is not a Bullet-Proof-Hosting-as-a-Service supplier. However, it is offered as a platform for other suppliers, asking a fee for external merchants.

CAPTCHA-solving-as-a-Service refers to a paid batch-solving CAPTCHA service. CAPTCHA has been widely used to allow humans access to online resources (e.g., websites) while preventing access by automated agents, i.e., bots. There are two modes

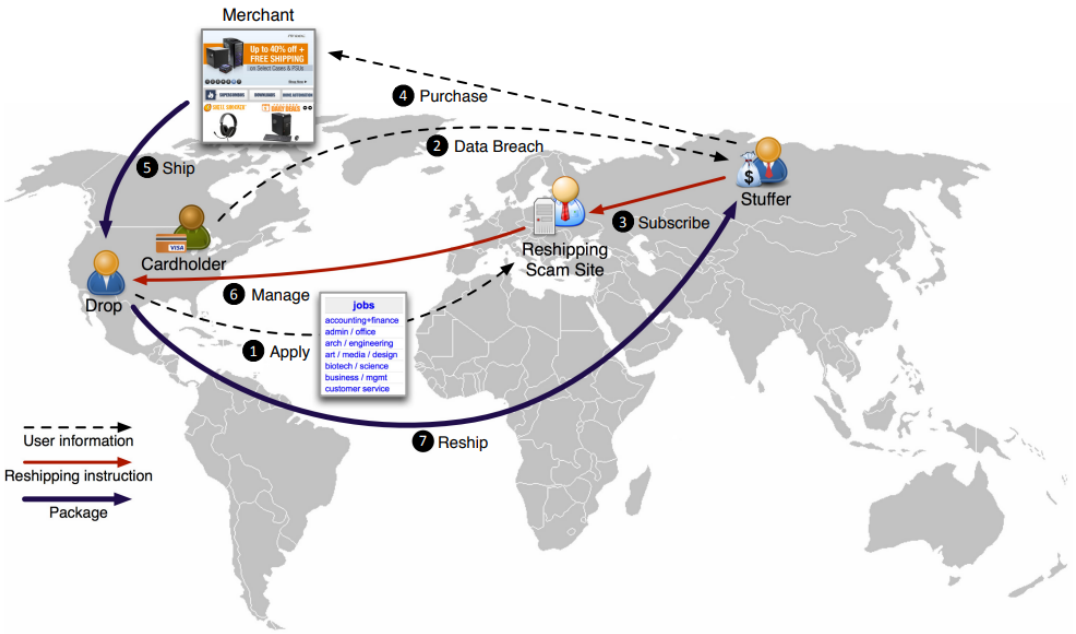


Figure 11: An overview of the reshipping mule scams, illustrated by Hao et al. [68]. (1) Apply: A drop (mule) applies for jobs through the reshipping scam site. (2) Data breach: Getting stolen credit cards from underground forums or darkweb. (3) Subscribe: The stuffer signs up with the reshipping scam site, so the stuffer get the reshipping service. (4) Purchase: Stuffer purchases electronics (e.g. computer, smartphone) at online shops. (5) Ship: The online shop ships the goods. (6) Manage: Reshipping scam site gives the reshipping label to the drop. (7) Reship: the drop ships the goods to stuffer.

of commercial CAPTCHA solving: (1) solve automatically using computer technology, and (2) solve manually using low-cost human labour (prices as low as US$1 per thousand [124]). Service providers usually target customers who, e.g., want to send bulk scam emails but are prevented by a CAPTCHA. The recruiter will hire workers to solve these CAPTCHAs and then forward them back to the mailing interface.

Service	Reg. Cost	PVA Cost	Increase
Google	$80	$100	1.25x
Youtube	$270	$349	1.29x
Youtube	$80	$150	1.875x
Google	$120	$230	1.9x
Google	$80	$500	6.25x
Facebook	$300	$600	2x
Facebook	$70	$350	5x
Facebook	$400	$1800	4.5x
Twitter	$20	$500	25x

Figure 13: Selling prices per thousand phone verified accounts (PVAs) on Google, YouTube, Facebook, and Twitter, aggregated statistics from three different "Phone/SMS Verification as a Service" merchants in 2013, provided by Thomas et al. [173].

In addition to using the CAPTCHA, some websites have started to use secondary verification methods (e.g., mobile verification codes) to distinguish machines from humans. Therefore, registered Phone Verified Accounts (PVA) started to be sold online, called **Phone/SMS-Verification-as-a-Service**. Thomas et al. [173] investigated the selling prices of PVAs on Google, YouTube, Facebook and Twitter in 2013 from three different merchants; this is shown in Figure 13. Abuse of phone registration has been seen as a waste of public resources, as phone verification is not free, and businesses need to pay a carrier for the communication. A large number of illegal PVA may directly disrupt the registration of legitimate users. By combining the above two mentioned services (i.e., CAPTCHA-solving-as-a-Service and Phone/SMS-Verification-as-a-Service), criminals can easily and quickly distribute deceptive information where both services support the underground ecosystem.

2.3. Evolving Services

Reputation-as-a-Service refers to a service that provides merchants with reputation rating in-

formation in darknet markets. Reputation is an essential characteristic in the underground cyber crime ecosystem [192]. Merchants are often active on multiple underground forums and the darkweb at the same time. Existing darknet markets often use third-party databases to display merchants ratings in other markets in order to attract potential customers. Dishonest traders are called "rippers" on the darknet [47]. For example, a ripper will sell invalid stolen credit cards or not deliver promised goods [47]. There is often mistrust among cyber criminals, but they may choose to cooperate because of mutual benefit [193]. In 2016, "Ripper[.]cc", a ripper profiles database, was launched allowing visitors to create a profile, which includes various contact details and information on how the scam happened. More recently (2019), a cross-market merchants' review website –called Kilos– was launched and stood out fast in the darknet [138]. It contains vendor information and reviews by their customers. Customers can easily query which market the vendor is active in and their reviews [138]. These services typically post advertisements for profit or charge the market operator for the number of Application Programming Interface (API) requests. The service fosters cooperation and trust among cyber criminals, enabling further criminal activity [192].

Similar reputation mechanisms also include using escrow trading systems and blacklists. The escrow trading systems are usually owned by the operator of the marketplaces/forums. When a transaction occurs, the money is first deposited in the website's account. Once the buyer confirms receipt of the item, the website releases the money into the seller's account. The website will act as a "court" to arbitrate the case when a dispute arises [9]. The blacklist is to ban dishonest users directly. Even though it is easy to register a new account on the darkweb, it takes time to build a reputation from scratch.

Personal-Profile-as-a-Service, also known as **Impersonation-as-a-Service**, aims to bypass authentication systems by collecting user profiles [22]. The diagram in Figure 14 illustrates how the service operates in three main stages. (1) *Profile acquisition* refers to the collection of user credentials and cookies from the victim's browser. Such acquisition could be done by malware infections, such as pay-per-install [19] or exploitation-as-a-service infrastructures [64]. (2) *Profile selection* refers to the sorting and selecting of relevant profiles by service operators for sale in underground markets, usually in cryptocurrencies. Finally, (3) *profile enforcement* refers to the use of those profiles (including browser sessions, behavioural metadata, geographic locations) to bypass risk-based authentication engines for profits (i.e., to cash out). Since the entire service process can be automated, this service is very likely to serve as a cyber criminal infrastructure. It could be further exploited, for example, to log into a victim's social network or an organisation's network for a social engineering attack.

2.4. Emerging Services

E-Whoring-as-a-Service refers to an online fraud service that uses simulated cyber sexual encounters for profit [136]. The service provides resources and tutorials (e.g., how to operate, how to make videos, and how to find potential victims) to customers. Hutchings and Pastrana [82] divided the E-Whoring scam into nine steps. (1) *Preparation*: In underground forums, some actors post free and paid tutorials and packages. Interestingly, the customers are reminded that they should be careful about distributing new types of tutorials, as this directly shares their profits. (2) *Obtain images*: Appealing material may be collected on the open web or purchased from a service provider. "verification templates" are sometimes useful, which are collected from social media, then edited with some special mark. This step may involve the trade of pornographic images/videos and sexual exploitation materials. (3) *Pre-condition*: Register an account on social platforms, forums, payment sites. This process may involve Phone/SMS-Verification-as-a-Service. Customers also need some fake backstory to enrich their fake identity. (4) *Dissemination*: Use false information to camouflage and promote the E-Whoring service. (5) *Negotiation*: Price negotiation. Potential victims may ask for a preview before buying. (6) *Payment*: Receive money, usually PayPal and Amazon gift cards. (7) *Doing*: Send prepared photos or play prepared fake live videos. (8) *Post-condition*: Block victim or continue with previous steps. (9) *Exit*: Retrieve the funds from PayPal or exchange gift cards to cash. There are also possible extras to increase income with the service; examples include attaching malware to images, inducing victims to click on a marketing website in order to earn traffic clicks, and blackmailing [82].

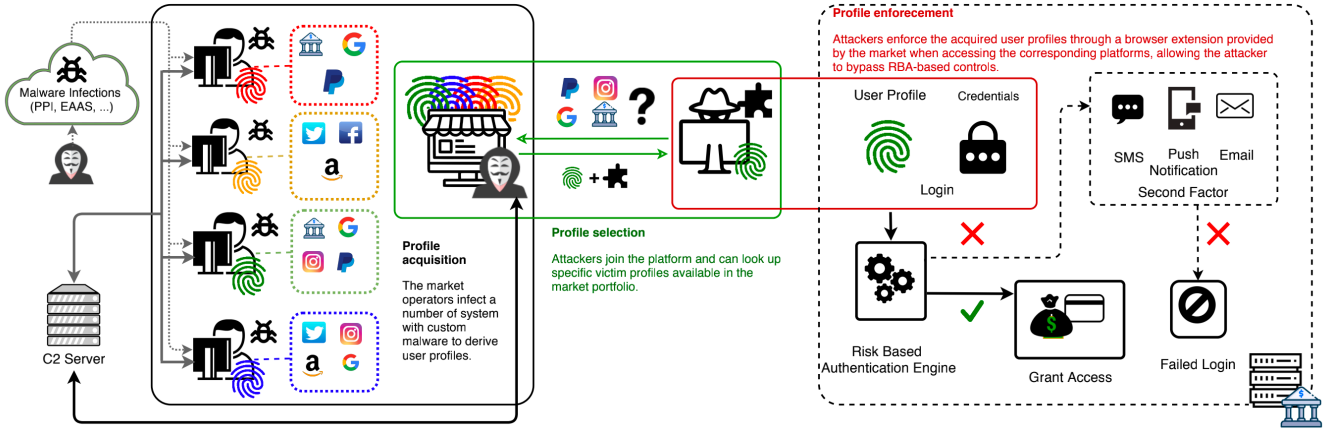


Figure 14: An overview of the Impersonation-as-a-Service operations, provided by Campobasso and Allodi [22]. Left: Profile acquisition. Middle: Profile selection. Right: Profile enforcement.

A similar form of deception enabled by e-whoring-as-a-service is called cyber-enabled romance fraud, which refers to an individual being deceived by someone with a fake romantic relationship [17]. This service also sometimes includes cyber sexual encounters, i.e., when a victim does not pay money to the offender, such private videos or pictures may be used to request a ransom payoff [183]. This category of criminal services has been shown to grow during the global pandemic [17].

Potential methods of intervention include disrupting tutorials, detection of false advertising such as those when criminals are looking for potential customers, and cooperation with PayPal to shut down accounts linked to fraud [82]. In terms of psychological factors, people who choose to trust criminals can be alone, so future interventions can also focus on the victim's side [17].

2.5. Research Challenges & Directions

This section reviewed many types of CaaS under the deception scope; those different types evolve and sometimes complement each other creating complex ecosystems of services, ready to be consumed. Most importantly, CaaS lowers the barriers of entry for new criminals [3]. Despite the progress made in the understanding and disrupting CaaS, there are still open challenges and directions for further research in this space.

Researchers perform an important role by identifying evidence and trends in CaaS development. A better understanding of the CaaS providers and consumers, their *modus-operandi* and underground economy, remains a direction for further development in order to fight cyber crime more effectively. One possible way forward, proposed by Huang et al. [79] and Clark [36], is the use of "honeypots" to mimic different types of service and capture activities and behaviours in the CaaS ecosystem.

Another direction is to promote direct cooperation between defenders, victims and other stakeholders, such as security companies, law enforcement, the financial sector, threat intelligence service providers, and policy makers [79]. We have seen victims of some services not being able to effectively prevent crime and protect themselves (e.g., DDoS in online games) [94]. In fact, mitigation and preventive measures against CaaS are limited for now. It is also important to note, however, that potential interventions are not necessarily limited to technology but may need to involve other disciplines such as Psychology, Sociology, and Economics. On the other hand, because the development of networks and services is difficult to predict, and new types of services will always appear, it becomes a challenge to detect and counter such crimes. Europol has prioritised (2022-2025) to crack down on criminals who provide online professional crime services, money laundering services (including money mules), and produce and disseminate child abuse [53].

3. Deception in Cyber-Physical Systems

3.1. Introduction

Cyber-physical systems (CPS) [46, 104], utilise physical and computational components to validate a process in the real-time world [135] and are employed in many applications, such as smart grids [8], health monitoring [70], and autonomous vehicles [160]. Owing to the growing popularity of such systems, cyber attackers have been actively trying to compromise such systems using various techniques. CPS are comprised of the combination of various different elements (operational technology and information technology infrastructures), which makes it an open target for cyber threats if required cyber security measures are not taken. Therefore, security of CPS is one of the most important challenges to address. Moreover, characteristics of CPS, such as heterogeneous hardware, unique protocols, limited resources and model specific nature, limit the application of current defensive information technology techniques such as cryptography and continuous patching [188]. Defence based on deception, such as honeypots and MTD as discussed in Section 1, are possible solution directions for CPS challenges. Hence, this section reviews selected state-of-the-art methods based on deception applied to CPS from the perspective of defence (Section 3.2) and offence (Section 3.3). Section 3.4 discusses research challenges and directions for deception in CPS.

3.2. Deception for Defence in CPS

Defence using deception in CPS has attracted researchers from academia and industry due to its promising results [139]. One of the promising use cases of deception for defence in CPS is to use honeypots to collect information about emerging attack vectors [59]. Attackers employ sophisticated attack strategies which are complex to anticipate by security managers (such as zero-day attacks). Honeypots can be leveraged to collect information about such emerging attacks, so that the security managers have enough time to mitigate such attacks before the attackers cause any harm to the CPS. A honeypot is a deception mechanism that uses a decoy to lure adversaries away from legitimate targets. Additionally, a honeypot gathers intelligence about adversaries such as the identity, methodology and motivation [23]. Honeypots have evolved to be a more dynamic deception mechanism to act as smart alarm systems. Honeypots can be classified in various sub-classes based on their level of interaction (connection attempts, command execution, etc.) as follows [48, 123].

1. **High-interaction honeypots** are advanced in nature (provide complex design) and their associated risks are high as they involve a real OS. These honeypots provide attackers with real operating systems with which to interact, with no simulation or restrictions. They provide possibilities to collect more information, through analysing and logging various attacks and actions.

2. **Medium-interaction honeypots** are less sophisticated compared to high-interaction honeypots. Unlike high-interaction honeypots they do not have have an OS. However, they offer simulation of complex services. Such honeypots have a high probability that an attacker will find vulnerabilities, but are still less likely to be compromised. Such honeypots provide a better simulation of OS for the attacker to interact with, thereafter more sophisticated attacks can be logged.

3. **Low-interaction honeypots** simulate services which cannot be exploited by the attackers to gain access of the honeypot. Usually they do not use an OS which the attackers can interact with. Thereafter, it minimises the risk of cyber threats associated with such honeypots. Nevertheless, this makes their application very limited. However, low-interaction honeypots can be used to analyze spammers and worms.

Intrusion detection systems have been utilised to detect possible cyber threats in CPS with promising results [122]. However, honeypots have several advantages over the intrusion detection systems, because a honeypot in principle should not get any legitimate requests; any request to the honeypot is likely an intrusion or a probe. Hence, it is much more convenient to detect intrusion attempts in contrast to a real system which is complex, with high levels of legitimate requests. One of the key advantages of honeypots is that they have a low false positive rate,

in contrast to traditional intrusion detection systems which usually produce a large number of false positive alerts. While using a honeypot, security managers will often only see an adversarial request, making it easy to identify.

Honeypots as a deception based mechanism is applicable to various CPS. You et al. [194] proposed a hybrid (semi-virtual and semi-physical) honeypot for industrial control systems. Their method improved honeypots to make them more realistic, attractive and more feasible by greatly reducing the cost of a flexible high-interaction honeypot to capture data related to physical interactions with the honeypots. Figure 15 presents an overview of the honeypot proposed by You et al. [194]. Ananbeh et al. [10] proposed a honeypot and ML based deception method to improve the architecture of SCADA network by adding a honeynet, called *CamouflageNet*, and ML techniques are used to defend against and to collect intelligence about cyber threats. Shahriar et al. [159] proposed a honeypot called DDAF, a deception defence based method to gather information about adversarial attacks in a hierarchical communication network of a CPS. Guarnizo et al. [66] proposed a physical world honeypot (the honeypot is running on a physical machine) platform for IoT devices, which allows a few physical devices among a large number of geographically distributed devices to be exposed to attackers. Using this platform, they collected a large volume of information related to adversarial attacks.

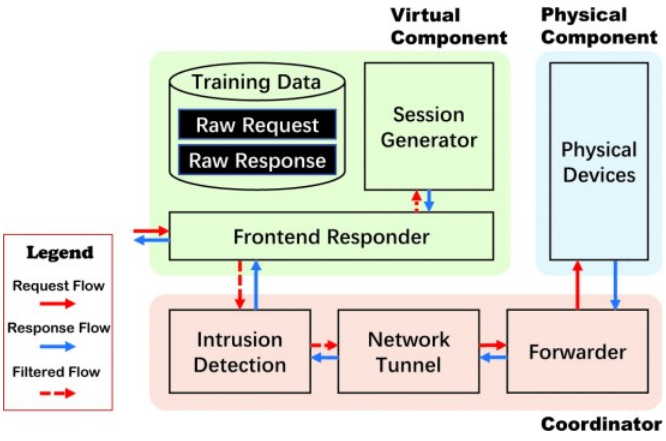


Figure 15: Overview of honeypot proposed by You et al. [194]. The architecture decouples interaction into two different components, and each component addresses a different level of interactions. A coordinator schedules the components to achieve flexibility while they work together.

Although honeypots provide promising features in CPS, it is worth mentioning their limitations. Honeypots cannot protect the real systems when the adversaries do not target them. They cannot replace other cyber security measures [210]. Hence, it is vital to consider security measures to protect the real systems. Additionally, a honeypot may reveal information (e.g., login warning messages and data fields) about the real systems, as they are supposed to mimic the real systems. This can be exploited by an attacker who identifies the honeypot and manipulates the information the honeypot will gather in order to mislead the defenders [210]. Therefore, a honeypot should be used to collect information and to help prioritise measures against adversarial attacks. Another intrinsic challenge in the use of honeypots is the effort required to design, deploy and analyse data collected, which directly impacts on cost. According to Ferguson-Walter [58], the following attributes are required; honeypots have to be "safe, realistic enough, high-fidelity, high-interaction, interesting, and current" to harvest the intended benefits as a defence mechanism to detect or distract attackers.

3.3. Deception for Offence in CPS

Deception has played a vital role in adversarial attacks on CPS, where attackers trick the legitimate users to use fake devices in order to steal information or to cause harm to the physical system. For example, a baseline attack on CPS is skimming devices in an automated teller machine (ATM). The skimmer devices are disguised to look like the legitimated part of an ATM. Such devices collect card numbers and pin codes when a user slides their card into the ATM [145]. This makes the attackers more consequential because the data collected can be used to withdraw cash from ATMs. The skimmer's devices can be attached to ATMs, gas station pumps and any other card transaction machines. With the progress of technology and its applications in sensitive applications such as healthcare or autonomous vehicles [158], such attacks have become more sophisticated [99] than just a simple skimmer device. For example, researchers introduced physical world attacks (an attack that impacts on the physical environment) to deceive autonomous vehicles [54]. The proposed attack actually aims at deception of ML based models in autonomous vehicles, which are responsible for automatically navigating a given route.

Autonomous vehicles use many sensors and built-in ML models to learn about the surrounding environment. One example of such a model is road sign classification. For example, for a given input image the model classifies it as a stop sign or a speed limit sign or something else [16]. To deceive such models, researchers developed an algorithm to generate stickers with a spatially-constrained perturbation to mimic vandalism and art by minimising the likelihood of being detected by a casual observer, but effective enough to deceive a DL-based model. After generating such stickers they stick them on the road signs, and when the vehicle's sensor tries to read such fabricated road signs, it misclassifies the sign, which can result in fatal accidents and loss of human life.

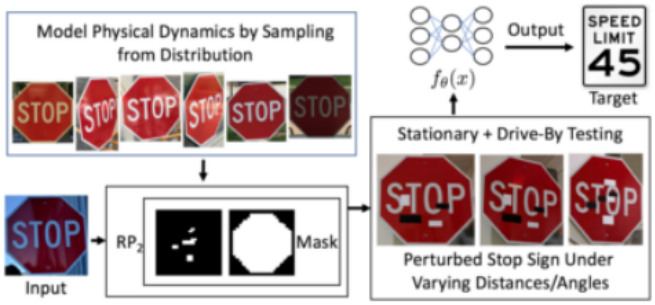


Figure 16: Attack pipeline overview in model proposed by [54], where an attacker tries to compromise the road sign recognition model $f_\theta(x)$.

Examples of emerging attacks on CPS are:

1. **Deception in road sign recognition**: As mentioned earlier, road sign recognition plays a vital role in autonomous vehicles. To deceive such systems, Eykholt et al. [54] proposed an algorithm to generate stickers which were used to deceive road sign recognition systems to misclassify road signs. The proposed algorithm achieved high success rate in deceiving DL-based models for road sign classification (mainly used in autonomous vehicles) by generating real road signs using their proposed algorithm, as shown in Figure 16.

2. **Deception in facial recognition systems**: Facial recognition systems are often used in smart devices, such as smart phones, as a security measure. Such systems usually rely on ML based models trained to recognise a human face. To deceive such systems, researchers have introduced a physical world attack which can generate adversarial stickers which can be attached to a human face to by pass ML-based facial recognition systems. For example, Shen et al. [163] proposed FaceAdv, a physical world attack to generate several adversarial stickers, which were then successfully placed in human faces to bypass ML based facial recognition systems, as shown in Figure 17. Moreover, Wenger et al. [182] evaluated the application of physical back-door attacks (i.e., embedding hidden malicious behaviours inside DNNs or input samples) on facial recognition systems and confirmed that physical backdoor attacks are a serious threat to classification tasks.

3. **Deception in object detection systems**: Object detection systems are mainly used in autonomous vehicles to detect any object (i.e., barriers, people) in order to facilitate a smooth journey and to avoid any possible life threatening incidents. Such systems use ML-based models to detect objects. To deceive such systems, a number of physical world attacks have been proposed by researchers. For example, Yang et al. [191] proposed a method to create physical objects to achieve the adversarial effect on license plates in DNNs based object detection systems.

4. **GPS spoofing in Unmanned Aerial Vehicles (UAVs)**: UAVs are used in many applications in both military and civilian areas. UAVs navigate with the help of signals from the Global Positioning System (GPS). Nevertheless, due to the unencrypted and unauthenticated signals sent from the GPS for civilian use, GPS spoofing [150] is one of the common cyber-physical attacks. In such types of attacks, an attacker generates GPS signals that cannot be distinguished from the original ones to deceive UAVs. Such attacks aim at rendering one or more target UAVs in a disoriented state, which may result in crashes of the target UAVs.

5. **False data injection in radars**: Frequency Modulated Continuous-Waves (FMCW) radars are used in autonomous vehicles. However, FMCW radars are vulnerable to deception attacks such as false data injection attacks. Chauhan [28] exploited a vulnerability

in FMCW radar and designed attacks to deceive the apparent distance, as measured by radar systems. Results of their study showed that it is possible to change the distance of an object arbitrarily, with a high probability of success.

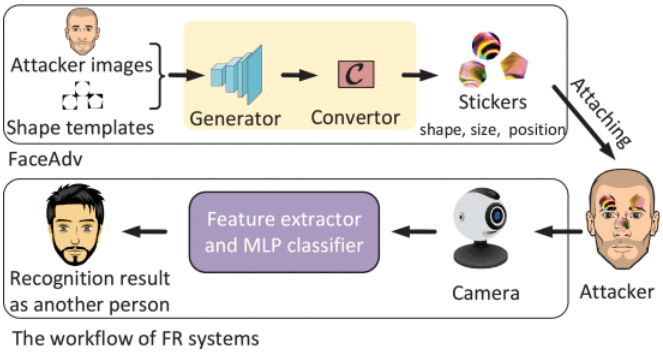


Figure 17: Overview of physical-world attacks on facial recognition (FR) systems using adversarial stickers crafted by FaceAdv, illustrated by [163].

3.4. Research Challenges & Directions

In this subsection we discuss the future challenges and directions related to DDD in CPS. Adversarial examples (AE) [200] are used to deceive DL-based models. It has been a topic of great research interest. Additionally, with the AE in the digital space, cyber-physical adversarial attacks are being considered as a more serious threat to ML-based applications, such as facial recognition in authentication, objection detection in autonomous vehicles [166], and many more [116]. In general, deception in the physical world is more complicated, especially in autonomous vehicles, because of relative positions of objects and because detectors may keep changing. Existing work on physical deception is still very limited in various scenarios [202], e.g., deception in object recognition and facial recognition systems. Although some of the most sophisticated adversarial attacks are limited to academic research, given the ever evolving progress of technology, such attacks can prose life-threatening challenges in the near future. Such attacks should be taken into consideration, especially in CPS, such as applications in healthcare, autonomous vehicles and power grids. Consider autonomous vehicles: if an adversary successfully compromises the road sign recognition system of a car in an area full of people, where the speed limit is 25 km/h and the attacker deceives the autonomous vehicle to read the sign as speed limit 110 km/h. Another example would be the deception of the license plate of a car. Think of an attacker trying to deceive the automatic number plate recognition system to evade law enforcement agencies. Similarly, an attacker can project objects into or out of the headway of an autonomous vehicle, which can cause potential safety hazards [90]. In this case an autonomous vehicle carrying people could hit that object and claim human lives.

As mentioned earlier, the unique architecture of CPS make them more difficult to maintain against emerging cyber threats. Additionally, the direct interaction of CPS with humans (such as autonomous vehicles) makes it an important issue to be addressed. Therefore, proper robust and efficient methods should be developed to mitigate such threats before such attacks cause any harm to human lives. Additionally, proper risk assessment should be performed to minimise the impact of a given attack. There are a number of risk assessment methods and techniques to secure CPS as mentioned by Ashibani and Mahmoud [12]. Risk can be assessed based on possible effects on the CPS. For example, risk can be categorised as: *high impact* if the attack can result in damaging and devastating effects on the CPS; *medium risk* if its impact is less severe, nevertheless it can pose a serious threat against CPS; *low risk* are the attacks which do not cause severe impact nor have a damaging effect, and the effects of such attacks can be mitigated easily.

4. Deception with Inaudible Voice Commands

4.1. Introduction

Voice-enabled systems, including voice assistants, smart home devices (i.e., IoT) and voice controllable systems, are used for daily activities by an increasing number of individuals, smart or autonomous vehicles and organisations. Although these systems provide a high degree of convenience, they also introduce security-related vulnerabilities that, when exploited, can cause adverse impact and harm to users. Inaudible voice commands are one of those classes of security-related vulnerabilities [111], and correspond to voice commands that are not perceived by human ears while being captured by voice-enabled systems. Such voice commands can be used for deceptive purposes, e.g., to take control or misuse such systems and launch attacks, e.g., targeting users' privacy [86]. Section 4.2 explores attacks to manipulate voice-enabled systems with inaudible voice commands, also called *inaudible voice attacks*. Section 4.3 reviews potential ways to detect those attacks, while Section 4.4 focuses on prevention. Finally, Section 4.5 discusses challenges and research directions in this domain.

4.2. Inaudible Voice Attacks

Inaudible voice attacks aim to inject commands into a voice controllable systems (VCS) to perform unauthorised actions without any access to the target (i.e., physical or remote access) and direct interaction with users. From a deception perspective, several scenarios can happen as a result of inaudible voice attacks, including automatic install of malware, initiation of outgoing video/phone calls for spying, injection of fake information, activation of aeroplane mode for disconnecting all wireless communications, and concealing of screens and voice feedback generated from a VCS [189]. A DoS attack can also be launched by using inaudible voice commands, e.g., by disrupting device pairing of IoT [119]. These attacks have been carried out in several ways, i.e., by generating ultrasonic sounds, making use of adversarial examples, manipulating capacitor voltages in electronic devices, and utilising light commands. These different types of attack

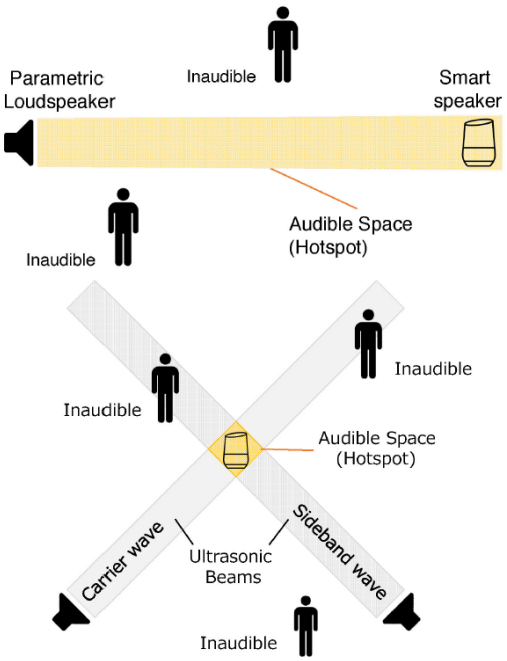


Figure 18: The *Audio Hotspot Attack* proposed by Iijima et al. [85] require three parametric loudspeakers: one is shown at the top (in the figure) and two are shown at the bottom. The sound becomes audible only in the yellow areas. When two parametric loudspeakers are used, each sound beam consists of a "carrier wave" or "sideband wave" with ultrasound frequency; they become audible where the two sound beams cross each other because they become an AM sound wave.

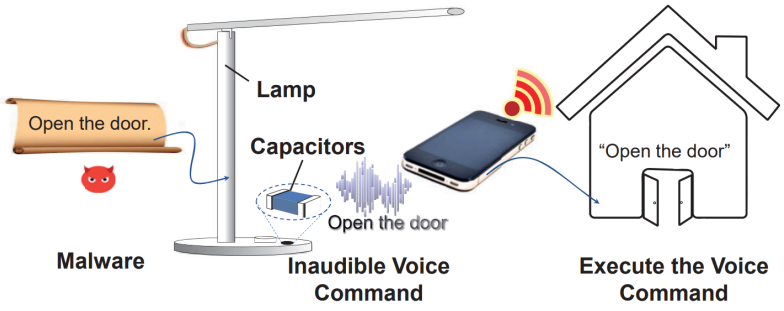


Figure 19: An attack scenario using CapSpeaker, proposed by Ji et al. [88]. The attacker injects malware into a smart lamp to manipulate the voltage across its built-in capacitors, and force it to play malicious inaudible voice commands targeting a smart speaker.

are reviewed in greater details next.

Ultrasonic sounds: This type of attack leverages ultrasonic speakers to generate ultrasound, i.e., sound above human audible frequency range (20 kHz), as a carrier wave for inaudible voice commands. Therefore, it requires a special speaker hardware able to play ultrasonic acoustic signals. For instance, Yan et al. [189] proposed the *DolphinAttack* to modulate voice commands on ultrasonic carriers and achieve inaudibility. Then, the modulated low-frequency audio commands are demodulated, recovered, and interpreted by the voice assistants. The proposed attack was validated on 12 popular voice assistants, e.g., Siri, Google Now, and Alexa, across 25 different models of devices. Although *DolphinAttack* showed the feasibility of this approach in practice, one of its limitations is that it relies on some unique characteristics of the microphones, which could in principle be eliminated. Thus, Iijima et al. [85] proposed the *Audio Hotspot Attack*, an inaudible voice attack based on nonlinearity in the air. It uses directional sound beams generated from parametric loudspeakers (i.e., speakers that can generate directional sound using ultrasound). The beams emit amplitude-modulated (AM) ultrasound which will be demodulated into audible sound in the air. In this way, the sound can only be heard in the audible space (called hotspot) which the adversary can adjust by using one or more parametric loudspeakers, as shown in Figure 18. The *Audio Hotspot Attack* has a range of 3.5 meters in a small room, and 12 meters in a long hallway.

Adversarial examples: This type of attack uses adversarial AI techniques to generate audio imperceptible by humans. The approach relies on adding a small perturbation to the original audio to manipulate a target VCS. For example, Chen et al. [30] presented *Metamorph*, a system which generates imperceptible audio that can survive over-the-air transmission targeting the Neural Network of a speech recognition system. *Metamorph* achieved a success rate over 90% at a distance up to 6 meters.

Manipulating capacitor voltages: This is a type of attack which utilises the capacitors in electronic devices; it was proposed recently by Ji et al. [88]. Capacitors can emit acoustic noises since the voltage across a capacitor causes it to vibrate at the same frequency as the voltage signal. Therefore, a capacitor can generate sounds in a similar way as a speaker. This can be used to play inaudible voice commands targeting a voice-enabled system. As shown in Figure 19, the main idea of the attack is to inject malware into an electronic device (e.g., a smart LED lamp) to induce the right voltage across the capacitors so that the device, called *CapSpeaker*, can play the malicious voice command targeting a smart speaker while remaining inaudible for humans around it. The main drawback of this approach is that it worked only at a distance up to 10.5cm.

Light commands: Sugawara et al. [169] introduced light-based audio injection attacks targeting VCS. This type of attack exploits the fact that microphones often unintentionally respond to light as if it was sound. Therefore, the attack injects sound into microphones by modulating the amplitude of a laser light. Validation of the attack showed that, while 5mW of laser power is sufficient to control many smart home devices, phones and tablets can only be controlled with 60mW of laser power. However, a significant advantage of the use of laser light, compared to existing attacks, is that it increases the attacking range up to circa 110 meters.

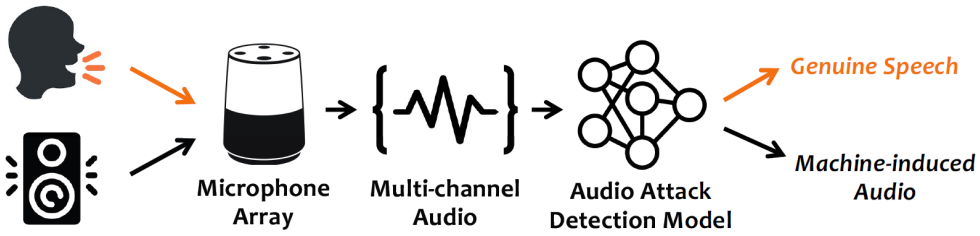


Figure 20: Overview of the detection approach proposed by Li et al. [110]. A DL model is used to differentiate genuine speech from machine-induced speech, based on the differences captured by microphone arrays.

4.3. Inaudible Voice Attack Detection

In the literature, various techniques have been proposed for detecting inaudible voice attacks. The proposed methods fall into several categories in terms of main technique used, including signal processing, microphone arrays, pop noise, and DL.

Signal Processing is used to detect ultrasonic-based inaudible voice attacks. For example, Mao et al. [118] proposed a detection approach for smart home devices, relying on the fact that the ultrasound has a high centre frequency (i.e., central frequency between the upper and lower cutoff frequencies), which is uncommon in smart home scenarios. The main idea of this approach is to capture all environmental ultrasound, and determine if the centre frequency of the received ultrasound falls within the suspicious attack frequency range. If this criteria holds, the ultrasonic signal is demodulated with the centre frequency to obtain the base-band signal, which is, then, used for malicious voice detection. The main limitation of this detection is that it is incapable of differentiating noise from malicious voice commands. This can cause jamming attacks to the detection device to keep on alerting. Similarly, He et al. [71] proposed an attack detection which emits an inverted ultrasound to cancel malicious voice commands. However, this approach can cause several health issues as it requires to constantly emit ultrasound [199].

Microphone Arrays are available on most off-the-shelf VCS. They have been used to detect inaudible voice attacks because of their rich sensing capability. To illustrate the feasibility of this approach, Li et al. [110] leveraged multi-channel microphone arrays to detect machine-induced voice attacks such as replay attacks, inaudible voice attacks, and synthesis attacks. Figure 20 shows the proposed approach which uses a DL model to distinguish machine-induced attack audio from genuine speech by leveraging different patterns in signal frequency and directivity (i.e., directional characteristic of a sound source). As another example, Zhang et al. [199] proposed *EarArray*, a system to detect DolphinAttack, as well as the attacker's direction, by estimating the attenuation rate of the command signals via built-in microphone arrays, as depicted in Figure 21. This method relies on the fact that ultrasound signals attenuate faster than audible sound signals. The authors observed that EarArray can detect DolphinAttack with 99% accuracy, and the attacker's direction with 97.89% accuracy.

Pop Noise is the sound generated by human breathing when speaking close to a microphone. Zhou et al. [204] suggested to leverage it to identify if the received voice command comes from a live user rather than from a speaker. The proposed detection strategy relies on the observation that pop noise has high energy in the low frequencies; e.g., in 0-100 Hz where it lasts for 20-100 ms. Pop noise locations are found in the input signal, and a two-class SVM was used to classify if the received signal was generated by the live user, or the attacker. This approach achieved promising results in detecting some existing inaudible voice attacks, such as the *Dolphin Attack* mentioned in Section 4.2.

Deep Learning based methods can be useful, especially, for the detection of audio adversarial examples. For example, Kokalj-Filipovic et al. [100] showed that DNNs can detect inaudible voice attacks, as a result of their preliminary studies. Guan et al. [65] proposed a detection system for autonomous vehicles by leveraging in-vehicle camera images. More precisely, the proposed system extracts features from the camera images with a CNN, and the voice commands with a Multilayer Percep-

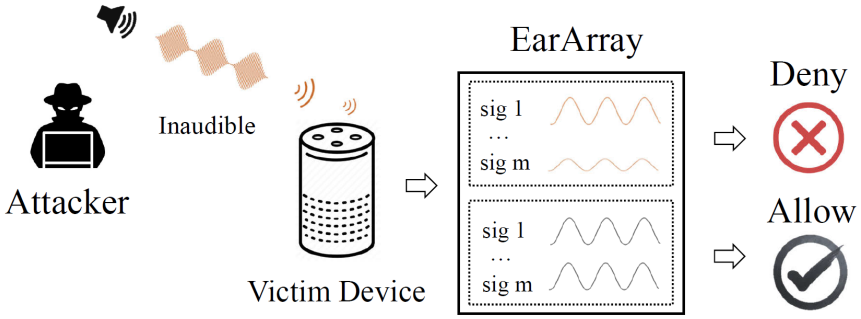


Figure 21: EarArray, proposed by Zhang et al. [199], measures the attenuation properties of the incoming sound to determine if it is an inaudible voice command. In this diagram, "sig i" in the EarArray box represents the voice signal captured by the i-th microphone.

tron. Then, it utilises multiple sensor fusion algorithms to integrate the extracted features, and determine if a voice command and a camera image are semantically consistent, e.g., to determine if a received "stop" voice command correlates with an identified STOP traffic sign. The proposed system achieved 89.2% classification accuracy. The source code of their work is publicly available.

4.4. Inaudible Voice Attack Prevention

Inaudible voice attack prevention techniques aim to defend VCS against inaudible voice attacks. The main approach uses liveness detection to differentiate the live human voice and the machine-induced voice which is utilised in such attacks. Liveness detection can be based on different items, including the user and the environment. Related techniques are further elaborated below.

Liveness detection based on users: User's motions and direction can be helpful for liveness detection. For instance, Zhang and Das [201] leveraged inaudible acoustic signals generated from a known hand gesture to defend voice assistants against several attacks, including inaudible voice attacks. The proposed approach, called *HandLock*, was suggested to be used as a second-factor authentication for certain sensitive operations, such as for confirming purchases. The authors reported that *HandLock* achieved 96.51% true positive rate, and 0.82% false acceptance rate. As another example, Lee et al. [105] presented a sonar-based liveness detection system to check if the user's direction is the same as the direction of the received voice command to protect smart speakers against remote attackers. The proposed system managed to deny remote voice attacks with an average accuracy of 95.5% within a range of 2 meters. Shi et al. [165] proposed *WearID*, a training-free voice authentication system which utilises aerial voice in the vibration domain via motion sensors of the user's wearable device. The captured aerial voice is verified with the captured voice in the audio domain for liveness detection. Other than the user's behaviour, the voice itself can also be used for liveness detection. For example, Ahmed et al. [2] introduced a voice liveness detection system, which leverages the differences in spectral power between live human voice and voices replayed through speakers. While the proposed system can avoid several voice attacks, it achieved 100% detection rate against ultrasonic-based inaudible voice attacks, covered in Section 4.2.

Liveness detection based on the environment: The environment where the target device is located can give some clues for liveness detection. To illustrate this approach, Meng et al. [121] proposed *WSVA*, a device-free liveness detection system utilising the wireless signals generated by Wi-Fi devices in an IoT environment. The wireless signals are utilised to capture the voice signal and the corresponding mouth motions, and to check the consistency between them. *WSVA* achieved 99% detection accuracy and 1% false acceptance rate.

4.5. Research Challenges & Directions

Inaudible voice commands are getting increasing attention from the research community. Although a number of approaches have been proposed, inaudible voice attacks mostly utilise ultrasound. In addition, attack distances are quite important for inaudible voice attacks. Compared to other existing attacks, light-based attacks provide the most promising results in terms of the attack distance. Nevertheless,

they also introduce some new challenges, such as opaque obstacles within the attack range impacting their effectiveness.

Detecting inaudible voice attacks is not as straightforward as differentiating audible sound from inaudible sound. Considering attacks where the produced sound is in the audible frequency range only for the target device [85], or the produced inaudible sound is transmitted through audible sounds [30], more holistic approaches are needed for a reliable detection mechanism. Among the existing detection methods, there is an increasing interest in utilising microphone arrays as smart devices contain multiple microphones [110, 199]. Furthermore, the proposed detection methods mostly focus on ultrasonic-based inaudible voice attacks, and overlook adversarial examples, as discussed in Section 4.3. Although ML is used in many detection methods, standalone DL approaches for inaudible voice attack detection seem to be only emerging.

Despite various techniques that can be applied to prevent some specific inaudible voice attacks, e.g., by developing microphones that are more robust to attacks such as *DolphinAttack*, more comprehensive approaches that are able to prevent VCS from a range of attacks are needed. Liveness detection, covered in Section 4.4 is one of those approaches, and provides promising results in many cases.

References

[1] Stefan Achleitner, Thomas La Porta, Patrick McDaniel, Shridatt Sugrim, Srikanth V. Krishnamurthy, and Ritu Chadha. 2016. Cyber Deception: Virtual Networks to Defend Insider Reconnaissance. In *Proceedings of the 8th ACM CCS International Workshop on Managing Insider Security Threats*. ACM, 57–68. https://doi.org/10.1145/2995959.2995962

[2] Muhammad Ejaz Ahmed, Il-Youp Kwak, Jun Ho Huh, Iljoo Kim, Taekkyung Oh, and Hyoungshick Kim. 2020. Void: A Fast and Light Voice Liveness Detection System. In *Proceedings of the 29th USENIX Security Symposium*. USENIX Association, 2685–2702. https://www.usenix.org/conference/usenixsecurity20/presentation/ahmed-muhammad

[3] Ugur Akyazi, Michel van Eeten, and Carlos H. Gañán. 2021. Measuring Cybercrime as a Service (CaaS) Offerings in a Cybercrime Forum. In *Proceedings of the 2021 Workshop on the Economics of Information Security*. 14 pages. https://weis2021.econinfosec.org/wp-content/uploads/sites/9/2021/06/weis21-akyazi.pdf

[4] Sami S. Al-Wakeel and Saad A. AL-Swailem. 2007. PRSA: A Path Redundancy Based Security Algorithm for Wireless Sensor Networks. In *Proceeding of the 2007 IEEE Wireless Communications and Networking Conference*. IEEE, 4156–4160. https://doi.org/10.1109/WCNC.2007.759

[5] Hooman Alavizadeh, Jin B. Hong, Julian Jang-Jaccard, and Dong Seong Kim. 2018. Comprehensive Security Assessment of Combined MTD Techniques for the Cloud. In *Proceedings of the 5th ACM Workshop on Moving Target Defense*. ACM, 11–20. https://doi.org/10.1145/3268966.3268967

[6] Hooman Alavizadeh, Jin B. Hong, Dong Seong Kim, and Julian Jang-Jaccard. 2021. Evaluating the effectiveness of shuffle and redundancy MTD techniques in the cloud. *Computers & Security* 102 (2021), 18 pages. https://doi.org/10.1016/j.cose.2020.102091

[7] Hooman Alavizadeh, Julian Jang-Jaccard, and Dong Seong Kim. 2018. Evaluation for Combination of Shuffle and Diversity on Moving Target Defense Strategy for Cloud Computing. In *Proceedings of the 2018 17th IEEE International Conference on Trust, Security and Privacy In Computing and Communications and the 12th IEEE International Conference on Big Data Science and Engineering*. IEEE, 573–578. https://doi.org/10.1109/TrustCom/BigDataSE.2018.00087

[8] Mahmoud Amin, Fayez F. M. El-Sousy, Ghada A. Abdel Aziz, Khaled Gaber, and Osama A. Mohammed. 2021. CPS Attacks Mitigation Approaches on Power Electronic Systems With Security Challenges for Smart Grid Applications: A Review. *IEEE Access* 9 (2021), 38571–38601. https://doi.org/10.1109/ACCESS.2021.3063229

[9] Analyst1. 2021. Dark Web - Justice League. https://analyst1.com/blog/dark-web-justice-league

[10] Obieda Ananbeh, Rund Alomari, and Austin Daniell. 2022. Improving ICS Security through Honeynets and Machine Learning Techniques. Research Square preprint. https://doi.org/10.21203/rs.3.rs-1333285/v1

[11] ARL Public Affairs. 2018. Army research takes proactive approach to defending computer systems. News release. https://www.army.mil/article/210787/army_research_takes_proactive_approach_to_defending_computer_systems

[12] Yosef Ashibani and Qusay H. Mahmoud. 2017. Cyber physical systems security: Analysis, challenges and solutions. *Computers & Security* 68 (2017), 81–97. https://doi.org/10.1016/j.cose.2017.04.005

[13] Nahid Bandi, Hesam Tajbakhsh, and Morteza Analoui. 2021. FastMove: Fast IP switching Moving Target Defense to mitigate DDoS Attacks. In *Proceedings of the 2021 IEEE Conference on Dependable and Secure Computing*. IEEE, 7 pages. https://doi.org/10.1109/DSC49826.2021.9346278

[14] Lijiang Bao (鲍立江), Weijun Zhong (仲伟俊), and Shu'e Mei (梅姝娥). 2021. The Influence of "Click Farming" on the Sellers' Competition in E-commerce Platform / 电子商务平台中刷单行为对商家间竞争的影响. *Systems Engineering - Theory & Practice* /《系统工程理论与实践》 41, 11 (2021), 2876–2886. https://doi.org/10.12011/SETP2020-0902

[15] Chafika Benzaïd and Tarik Taleb. 2020. AI for Beyond 5G Networks: A Cyber-Security Defense or Offense Enabler? *IEEE Network* 34, 6 (2020), 140–147. https://doi.org/10.1109/MNET.011.2000088

[16] Amal Bouti, Med Adnane Mahraz, Jamal Riffi, and Hamid Tairi. 2020. A robust system for road sign detection and classification using LeNet architecture based on convolutional neural network. *Soft Computing* 24, 9 (2020), 6721–6733. https://doi.org/10.1007/s00500-019-04307-6

[17] David Buil-Gil and Yongyu Zeng. 2021. Meeting you was a fake: investigating the increase in romance fraud during COVID-19. *Journal of Financial Crime* 29, 2 (2021), 460–475. https://doi.org/10.1108/JFC-02-2021-0042

[18] Andriy Burkov, Laëtitia Matignon, and Brahim Chaib-Draa. 2013. Stochastic Games. In *Markov Decision Processes in Artificial Intelligence*. John Wiley & Sons, Ltd, Chapter 8, 229–276. https://doi.org/10.1002/9781118557426.ch8

[19] Juan Caballero, Chris Grier, Christian Kreibich, and Vern Paxson. 2011. Measuring Pay-per-Install: The Commoditization of Malware Distribution. In *Proceedings of the 20th USENIX Conference on Security* (San Francisco, CA) *(SEC'11)*. USENIX Association, USA, 13. https://dl.acm.org/doi/10.5555/2028067.2028080

[20] Javier Cabrera Arteaga, Orestis Floros, Oscar Vera Perez, Benoit Baudry, and Martin Monperrus. 2021. CROW: Code Diversification for WebAssembly. In *Proceedings of the 2021 Workshop on Measurements, Attacks, and Defenses for the Web*. Internet Society, 11 pages. https://doi.org/10.14722/madweb.2021.23004

[21] Gui-lin Cai, Bao-sheng Wang, Wei Hu, and Tian-zuo Wang. 2016. Moving target defense: state of the art and characteristics. *Frontiers of Information Technology & Electronic Engineering* 17, 11 (2016), 1122–1153. https://doi.org/10.1631/FITEE.1601321

[22] Michele Campobasso and Luca Allodi. 2020. Impersonation-as-a-Service: Characterizing the Emerging Criminal Infrastructure for User Impersonation at Scale. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1665–1680. https://doi.org/10.1145/3372297.3417892

[23] Jianhong Cao, Wei Li, Jianjun Li, and Bo Li. 2017. DiPot: A Distributed Industrial Honeypot System. In *Smart Computing and Communication: Second International Conference, SmartCom 2017, Shenzhen, China, December 10-12, 2017, Proceedings*. Springer, 300–309. https://doi.org/10.1007/978-3-319-73830-7_30

[24] Thomas E. Carroll, Michael Crouse, Errin W. Fulp, and Kenneth S. Berenhaut. 2014. Analysis of Network Address Shuffling as a Moving Target Defense. In *Proceedings of the 2014 IEEE International Conference on Communications*. IEEE, 701–706. https://doi.org/10.1109/ICC.2014.6883401

[25] Kevin M. Carter, James F. Riordan, and Hamed Okhravi. 2014. A Game Theoretic Approach to Strategy Determination for Dynamic Platform Defenses. In *Proceedings of the 2014 1st ACM Workshop on Moving Target Defense*. ACM, 21–30. https://doi.org/10.1145/2663474.2663478

[26] Valentina Casola, Alessandra De Benedictis, and Massimiliano Albanese. 2013. A Moving Target Defense Approach for Protecting Resource-constrained Distributed Devices. In *Proceeding of the 2013 IEEE 14th International Conference on Information Reuse Integration*. IEEE, 22–29. https://doi.org/10.1109/IRI.2013.6642449

[27] Xinzhong Chai, Yasen Wang, Chuanxu Yan, Yuan Zhao, Wenlong Chen, and Xiaolei Wang. 2020. DQ-MOTAG: Deep Reinforcement Learning-based Moving Target Defense Against DDoS Attacks. In *Proceeding of the 2020 IEEE 5th International Conference on Data Science in Cyberspace*. IEEE, 375–379. https://doi.org/10.1109/DSC50466.2020.00065

[28] Ruchir Chauhan. 2014. *A Platform for False Data Injection in Frequency Modulated Continuous Wave Radar*. Utah State University. https://doi.org/10.26076/6adb-d066

[29] Lin Chen and Raphaël C.-W. Phan. 2021. Network Reconfiguration via Diversity: Theoretical Foundation and Algorithm Design. In *Proceedings of the 2021 IEEE 94th Vehicular Technology Conference*. IEEE, 5 pages. https://doi.org/10.1109/VTC2021-Fall52928.2021.9625508

[30] Tao Chen, Longfei Shangguan, Zhenjiang Li, and Kyle Jamieson. 2020. Metamorph: Injecting Inaudible Commands into Over-the-air Voice Controlled Systems. In *Proceedings of the 2020 Conference on Network and Distributed Systems Security Symposium*. Internet Society, 17 pages. https://doi.org/10.14722/ndss.2020.23055

[31] Tommy Chin and Kaiqi Xiong. 2016. Dynamic Generation Containment Systems (DGCS): A Moving Target Defense Approach. In *Proceeding of the 2016 3rd International Workshop on Emerging Ideas and Trends in Engineering of Cyber-Physical Systems*. IEEE, 11–16. https://doi.org/10.1109/EITEC.2016.7503690

[32] Jin-Hee Cho, Dilli P. Sharma, Hooman Alavizadeh, Seunghyun Yoon, Noam Ben-Asher, Terrence J. Moore, Dong Seong Kim, Hyuk Lim, and Frederica F. Nelson. 2020. Toward Proactive, Adaptive Defense: A Survey on Moving Target Defense. *IEEE Communications Surveys & Tutorials* 22, 1 (2020), 709–745. https://doi.org/10.1109/COMST.2019.2963791

[33] Ankur Chowdhary, Adel Alshamrani, Dijiang Huang, and Hongbin Liang. 2018. MTD Analysis and Evaluation Framework in Software Defined Network (MASON). In *Proceedings of the 2018 ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization*. ACM, 43–48. https://doi.org/10.1145/3180465.3180473

[34] Ankur Chowdhary, Dijiang Huang, Abdulhakim Sabur, Neha Vadnere, Myong Kang, and Bruce Montrose. 2021. SDN-based Moving Target Defense using Multi-agent Reinforcement Learning. In *Proceedings of the 2021 1st International Conference on Autonomous Intelligent Cyber Defense Agents*. AICA, 10 pages. https://www.aicaconference.org/wp-content/uploads/2021/11/AICA2021_Chowdhary_Huang_Sabur.pdf

[35] Andrew Clark, Kun Sun, Linda Bushnell, and Radha Poovendran. 2015. A Game-Theoretic Approach to IP Address Randomization in Decoy-Based Cyber Defense. In *Decision and Game Theory for Security: 6th International Conference, GameSec 2015, London, UK, November 4-5, 2015, Proceedings*. Springer, 3–21. https://doi.org/10.1007/978-3-319-25594-1_1

[36] David D. Clark. 2012. Control Point Analysis. In *Proceedings of the 2012 TRPC Conference*. SSRN. https://doi.org/10.2139/ssrn.2032124

[37] Richard Colbaugh and Kristin Glass. 2012. Predictability-Oriented Defense against Adaptive Adversaries. In *Proceeding of the 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2721–2727. https://doi.org/10.1109/ICSMC.2012.6378159

[38] Richard Colbaugh and Kristin Glass. 2013. Moving Target Defense for Adaptive Adversaries. In *Proceeding of the 2013 IEEE International Conference on Intelligence and Security Informatics*. IEEE, 50–55. https://doi.org/10.1109/ISI.2013.6578785

[39] Ernesto Serrano Collado, Pedro A. Castillo, and Juan Julián Merelo Guervós. 2020. Using Evolutionary Algorithms for Server Hardening via the Moving Target Defense Technique. In *Applications of Evolutionary Computation: 23rd European Conference, EvoApplications 2020, Held as Part of EvoStar 2020, Seville, Spain, April 15–17, 2020, Proceedings*. Springer, 670–685. https://doi.org/10.1007/978-3-030-43722-0_43

[40] Michael Crouse and Errin W. Fulp. 2011. A Moving Target Environment for Computer Configurations Using Genetic Algorithms. In *Proceeding of the 2011 4th Symposium on Configuration Analytics and Automation*. IEEE, 7 pages. https://doi.org/10.1109/SafeConfig.2011.6111663

[41] Michael Crouse, Errin W. Fulp, and Daniel Canas. 2012. Improving the Diversity Defense of Genetic Algorithm-based Moving Target Approaches. In *Proceedings of the 2012 3rd National Symposium on Moving Target Research*. 8 pages. https://michaelbcrouse.com/papers/nsmtr12_1.pdf

[42] Boris Danev, Ramya Jayaram Masti, Ghassan O. Karame, and Srdjan Capkun. 2011. Enabling Secure VM-VTPM Migration in Private Clouds. In *Proceedings of the 27th Annual Computer Security Applications Conference*. ACM, 187–196. https://doi.org/10.1145/2076732.2076759

[43] Saptarshi Debroy, Prasad Calyam, Minh Nguyen, Allen Stage, and Vladimir Georgiev. 2016. Frequency-minimal Moving Target Defense using Software-Defined Networking. In *Proceeding of the 2016 International Conference on Computing, Networking and Communications*. IEEE, 6 pages. https://doi.org/10.1109/ICCNC.2016.7440635

[44] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *Proceeding of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[45] Li Deng. 2012. The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Processing Magazine* 29, 6 (2012), 141–142. https://doi.org/10.1109/MSP.2012.2211477

[46] Patricia Derler, Edward A. Lee, and Alberto Sangiovanni Vincentelli. 2012. Modeling Cyber–Physical Systems. *Proc. IEEE* 100, 1 (2012), 13–28. https://doi.org/10.1109/JPROC.2011.2160929

[47] Digital Shadows Analyst Team. 2017. Reducing the risk of Ripper fraud. http://www.digitalshadows.com/blog-and-research/innovation-in-the-underworld-reducing-the-risk-of-ripper-fraud

[48] Michael Dodson, Alastair R. Beresford, and Mikael Vingaard. 2020. Using Global Honeypot Networks to Detect Targeted ICS Attacks. In *Proceedings of the 2020 12th International Conference on Cyber Conflict*. NATO CCDCOE Publications, 275–291. https://doi.org/10.23919/CyCon49761.2020.9131734

[49] Matthew Dunlop, Stephen Groat, William Urbanski, Randy Marchany, and Joseph Tront. 2011. MT6D: A Moving Target IPv6 Defense. In *Proceeding of the 2011 Military Communications Conference*. IEEE, 1321–1326. https://doi.org/10.1109/MILCOM.2011.6127486

[50] Manuel Egele, Theodoor Scholte, Engin Kirda, and Christopher Kruegel. 2008. A Survey on Automated Dynamic Malware-Analysis Techniques and Tools. *ACM Comput. Surv.* 44, 2, Article 6 (mar 2008), 42 pages. https://doi.org/10.1145/2089125.2089126

[51] Taha Eghtesad, Yevgeniy Vorobeychik, and Aron Laszka. 2020. Adversarial Deep Reinforcement Learning Based Adaptive Moving Target Defense. In *Decision and Game Theory for Security: 11th International Conference, GameSec 2020, College Park, MD, USA, October 28–30, 2020, Proceedings*. Springer, 58–79. https://doi.org/10.1007/978-3-030-64793-3_4

[52] Vittoria Elliott. 2022. In Ukraine' s cyber-war with Russia, who is a civilian and what is a war crime? https://restofworld.org/2022/in-ukraines-cyber-war-with-russia-who-is-a-civilian-and-what-is-a-war-crime/ Published by Rest of World.

[53] Europol. 2022. EU Policy Cycle - EMPACT - EMPACT 2022 + Fighting crime together. https://www.europol.europa.eu/crime-areas-and-trends/eu-policy-cycle-empact

[54] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. 2018. Robust Physical-World Attacks on Deep Learning Visual Classification. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 1625–1634. https://doi.org/10.1109/CVPR.2018.00175

[55] F-Secure. 2006. Rogue:W32/Bravesentry. Web page. https://www.f-secure.com/sw-desc/rogue_w32_bravesentry.shtml

[56] FBI. 2020. Money mules. https://www.fbi.gov/scams-and-safety/common-scams-and-crimes/money-mules

[57] Xiaotao Feng, Zizhan Zheng, Derya Cansever, Ananthram Swami, and Prasant Mohapatra. 2017. A Signaling Game Model for Moving Target Defense. In *Proceedings of the 2017 IEEE Conference on Computer Communications*. IEEE, 9 pages. https://doi.org/10.1109/INFOCOM.2017.8057200

[58] Kimberly Ferguson-Walter. 2022. Imposing a Cyber Penalty Against Attackers with Cyber Deception. Online article, *;login:* Online. https://www.usenix.org/publications/loginonline/imposing-cyber-penalty-against-attackers-cyber-deception

[59] Javier Franco, Ahmet Aris, Berk Canberk, and A. Selcuk Uluagac. 2021. A Survey of Honeypots and Honeynets for Internet of Things, Industrial Internet of Things, and Cyber-Physical Systems. *IEEE Communications Surveys & Tutorials* 23, 4 (2021), 2351–2383. https://doi.org/10.1109/COMST.2021.3106669

[60] Mengmeng Ge, Jin-Hee Cho, Dongseong Kim, Gaurav Dixit, and Ing-Ray Chen. 2021. Proactive Defense for Internet-of-Things: Moving Target Defense With Cyberdeception. *ACM Transactions on Internet Technology* 22, 1, Article 24 (2021), 31 pages. https://doi.org/10.1145/3467021

[61] Mengmeng Ge, Jin-Hee Cho, Dongseong Kim, Gaurav Dixit, and Ing-Ray Chen. 2021. Proactive Defense for Internet-of-Things: Moving Target Defense With Cyberdeception. *ACM Transactions on Internet Technology* 22, 1, Article 24 (2021), 31 pages. https://doi.org/10.1145/3467021

[62] Mengmeng Ge, Jin B. Hong, Simon Enoch Yusuf, and Dong Seong Kim. 2018. Proactive defense mechanisms for the software-defined Internet of Things with non-patchable vulnerabilities. *Future Generation Computer Systems* 78 (2018), 568–582. https://doi.org/10.1016/j.future.2017.07.008

[63] Ram D. Gopal, Afrouz Hojati, and Raymond A. Patterson. 2022. Analysis of third-party request structures to detect fraudulent websites. *Decision Support Systems* 154, Article 113698 (2022). https://doi.org/10.1016/j.dss.2021.113698

[64] Chris Grier, Lucas Ballard, Juan Caballero, Neha Chachra, Christian J. Dietrich, Kirill Levchenko, Panayiotis Mavrommatis, Damon McCoy, Antonio Nappa, Andreas Pitsillidis, Niels Provos, M. Zubair Rafique, Moheeb Abu Rajab, Christian Rossow, Kurt Thomas, Vern Paxson, Stefan Savage, and Geoffrey M. Voelker. 2012. Manufacturing Compromise: The Emergence of Exploit-as-a-Service. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security* (Raleigh, North Carolina, USA) *(CCS '12)*. Association for Computing Machinery, New York, NY, USA, 821—832. https://doi.org/10.1145/2382196.2382283

[65] Jiwei Guan, Xi Zheng, Chen Wang, Yipeng Zhou, and Alireza Jolfaei. 2021. Robust Sensor Fusion Algorithms Against Voice Command Attacks in Autonomous Vehicles. In *Proceedings of the IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, 895–902. https://doi.org/10.1109/TrustCom53373.2021.00126

[66] Juan David Guarnizo, Amit Tambe, Suman Sankar Bhunia, Martín Ochoa, Nils Ole Tippenhauer, Asaf Shabtai, and Yuval Elovici. 2017. SIPHON: Towards Scalable High-Interaction Physical Honeypots. In *Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security*. ACM, 57–68. https://doi.org/10.1145/3055186.3055192

[67] Robert Gutzwiller, Kimberly Ferguson-Walter, Sunny Fugate, and Andrew Rogers. 2018. "Oh, Look, A Butterfly!" A Framework For Distracting Attackers To Improve Cyber Defense. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62, 1 (2018), 272–276. https://doi.org/10.1177/1541931218621063

[68] Shuang Hao, Kevin Borgolte, Nick Nikiforakis, Gianluca Stringhini, Manuel Egele, Michael Eubanks, Brian Krebs, and Giovanni Vigna. 2015. Drops for Stuff: An Analysis of Reshipping Mule Scams. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1081–1092. https://doi.org/10.1145/2810103.2813620

[69] Jassim Happa, Thomas Bashford-Rogers, Alastair Janse Van Rensburg, Michael Goldsmith, and Sadie Creese. 2021. Deception in Network Defences Using Unpredictability. *Digital Threats: Research and Practice* 2, 4, Article 29 (2021), 26 pages. https://doi.org/10.1145/3450973

[70] Shah Ahsanul Haque, Syed Mahfuzul Aziz, and Mustafizur Rahman. 2014. Review of Cyber-Physical System in Healthcare. *International Journal of Distributed Sensor Networks* 10, 4, Article 217415 (2014), 20 pages. https://doi.org/10.1155/2014/217415

[71] Yitao He, Junyu Bian, Xinyu Tong, Zihui Qian, Wei Zhu, Xiaohua Tian, and Xinbing Wang. 2019. Canceling Inaudible Voice Commands Against Voice Control Systems. In *Proceedings of the 25th Annual International Conference on Mobile Computing and Networking*. ACM, Article 28, 15 pages. https://doi.org/10.1145/3300061.3345429

[72] Kang He (何康), Yuefei Zhu (祝跃飞), Long Liu (刘龙), Bin Lu (芦斌), and Bin Liu (刘彬). 2020. Improve the Robustness of Algorithm under Adversarial Environment by Moving Target Defense / 敌对攻击环境下基于移动目标防御的算法稳健性增强方法. *Chinese Journal of Network and Information Security* /《网络与信息安全学报》6, 4 (2020), 67–76. http://www.infocomm-journal.com/cjnis/CN/10.11959/j.issn.2096-109x.2020052

[73] Ryan Heartfield and George Loukas. 2015. A Taxonomy of Attacks and a Survey of Defence Mechanisms for Semantic Social Engineering Attacks. *Comput. Surveys* 48, 3, Article 37 (2015), 39 pages. https://doi.org/10.1145/2835375

[74] Alex Hern. 2016. Snapchat Leaks employee pay data after CEO email scam. news reports, The Guardian. https://www.theguardian.com/technology/2016/feb/29/snapchat-leaks-employee-data-ceo-scam-email

[75] Vahid Heydari. 2018. Moving Target Defense for Securing SCADA Communications. *IEEE Access* 6 (2018), 33329–33343. https://doi.org/10.1109/ACCESS.2018.2844542

[76] Jin B. Hong, Simon Yusuf Enoch, Dong Seong Kim, Armstrong Nhlabatsi, Noora Fetais, and Khaled M. Khan. 2018. Dynamic security metrics for measuring the effectiveness of moving target defense techniques. *Computers & Security* 79 (2018), 33–52. https://doi.org/10.1016/j.cose.2018.08.003

[77] Geng Hong (洪赓), Sen Yang (杨森), Han Ye (叶瀚), Zhemin Yang (杨哲慜), and Min Yang (杨珉). 2021. Detection and Analysis Technology of Cybercrime / 网络犯罪的检测分析技术. *Journal of Computer Research and Development* /《计算机研究与发展》 58, 10 (2021), 2120–2139. https://doi.org/10.7544/issn1000-1239.2021.20210855

[78] Ruiqin Hu (胡瑞钦), Jinglei Tan (谭晶磊), Xinhe Peng (彭心荷), and Hongqi Zhang (张红旗). 2022. Dynamic Hopping Technology of Double Virtual IP Address for SDN Data Layer / 面向 SDN 数据层的双虚假 IP 地址动态跳变技术. *Netinfo Security* /《信息网络安全》 1, 2 (2022), 76–85. http://netinfo-security.org/CN/abstract/abstract7373.shtml

[79] Keman Huang, Michael Siegel, and Stuart Madnick. 2018. Systematically Understanding the Cyber Attack Business: A Survey. *Comput. Surveys* 51, 4 (2018), 36 pages. https://doi.org/10.1145/3199674

[80] Yih Huang and Anup K. Ghosh. 2011. Introducing Diversity and Uncertainty to Create Moving Attack Surfaces for Web Services. In *Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats*. Springer, 131–151. https://doi.org/10.1007/978-1-4614-0977-9_8

[81] Jack Hughes, Ben Collier, and Alice Hutchings. 2019. From playing games to committing crimes: A multi-technique approach to predicting key actors on an online gaming forum. In *Proceedings of the 2019 APWG Symposium on Electronic Crime Research*. IEEE, 12 pages. https://doi.org/10.1109/eCrime47957.2019.9037586

[82] Alice Hutchings and Sergio Pastrana. 2019. Understanding eWhoring. In *Proceedings of the 2019 IEEE European Symposium on Security and Privacy*. IEEE, 201–214. https://doi.org/10.1109/EuroSP.2019.00024

[83] M. F. Hyder, Waseemullah, M. U. Farooq, U. Ahmed, and W. Raza. 2021. Towards Enhancing the Endpoint Security using Moving Target Defense (Shuffle-based Approach) in Software Defined Networking. *Engineering, Technology & Applied Science Research* 11, 4 (2021), 7483–7488. https://doi.org/10.48084/etasr.4316

[84] Thomas S. Hyslip and Thomas J. Holt. 2019. Assessing the Capacity of DRDoS-For-Hire Services in Cybercrime Markets. *Deviant Behavior* 40, 12 (2019), 1609–1625. https://doi.org/10.1080/01639625.2019.1616489

[85] Ryo Iijima, Shota Minami, Yunao Zhou, Tatsuya Takehisa, Takeshi Takahashi, Yasuhiro Oikawa, and Tatsuya Mori. 2021. Audio Hotspot Attack: An Attack on Voice Assistance Systems Using Directional Sound Beams and its Feasibility. *IEEE Transactions on Emerging Topics in Computing* 9, 4 (2021), 2004–2018. https://doi.org/10.1109/TETC.2019.2953041

[86] Yasha Iravantchi, Karan Ahuja, Mayank Goel, Chris Harrison, and Alanson Sample. 2021. PrivacyMic: Utilizing Inaudible Frequencies for Privacy Preserving Daily Activity Recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Article 198, 13 pages. https://doi.org/10.1145/3411764.3445169

[87] Rauf Izmailov, Peter Lin, Sridhar Venkatesan, and Shridatt Sugrim. 2021. *Combinatorial Boosting of Classifiers for Moving Target Defense Against Adversarial Evasion Attacks*. ACM, 13–21. https://doi.org/10.1145/3474370.3485661

[88] Xiaoyu Ji, Juchuan Zhang, Shui Jiang, Jishen Li, and Wenyuan Xu. 2021. CapSpeaker: Injecting Voices to Microphones via Capacitors. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1915–1929. https://doi.org/10.1145/3460120.3485389

[89] Quan Jia, Huangxin Wang, Dan Fleck, Fei Li, Angelos Stavrou, and Walter Powell. 2014. Catch Me If You Can: A Cloud-Enabled DDoS Defense. In *Proceeding of the 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*. IEEE, 264–275. https://doi.org/10.1109/DSN.2014.35

[90] Yunhan Jia Jia, Yantao Lu, Junjie Shen, Qi Alfred Chen, Hao Chen, Zhenyu Zhong, and Tao Wei Wei. 2019. Fooling Detection Alone is Not Enough: First Adversarial Attack against Multiple Object Tracking. arXiv preprint arXiv:1905.11026. https://doi.org/10.48550/arXiv.1905.11026

[91] Lü Jiang (蒋侣), Jin-dong Wang (王晋东), and Heng-wei Zhang (张恒巍). 2021. A Markov Signaling Game-theoretic Approach to Moving Target Defense Strategy Selection / 基于多阶段 Markov 信号博弈的移动目标防御最优决策方法. *Acta Electronica Sinica* /《电子学报》 49, 3 (2021), 527–535. http://www.ejournal.org.cn/EN/10.12263/DZXB.20191070

[92] David J. John, Robert W. Smith, William H. Turkett, Daniel A. Cañas, and Errin W. Fulp. 2014. Evolutionary Based Moving Target Cyber Defense. In *Proceedings of the Companion Publication of the 2014 Annual Conference on Genetic and Evolutionary Computation*. ACM, 1261–1268. https://doi.org/10.1145/2598394.2605437

[93] Aljosha Judmayer, Georg Merzdovnik, Johanna Ullrich, Artemios G. Voyiatzis, and Edgar Weippl. 2018. A Performance Assessment of Network Address Shuffling in IoT Systems. In *Proceeding of the 2017 International Conference on Computer Aided Systems Theory*. Springer, 197–204. https://doi.org/10.1007/978-3-319-74718-7_24

[94] Mohammad Karami and Damon McCoy. 2013. Rent to pwn: Analyzing commodity booter DDoS services. *Usenix login* 38, 6 (2013), 20–23. http://mason.gmu.edu/~mkarami/papers/login13.pdf

[95] Mohammad Karami, Youngsam Park, and Damon McCoy. 2016. Stress Testing the Booters: Understanding and Undermining the Business of DDoS Services. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1033–1043. https://doi.org/10.1145/2872427.2883004

[96] Gaurav S. Kc, Angelos D. Keromytis, and Vassilis Prevelakis. 2003. Countering Code-Injection Attacks with Instruction-Set Randomization. In *Proceedings of the 10th ACM Conference on Computer and Communications Security*. ACM, 272–280. https://doi.org/10.1145/948109.948146

[97] Angelos D. Keromytis, Roxana Geambasu, Simha Sethumadhavan, Salvatore J. Stolfo, Junfeng Yang, Azzedine Benameur, Marc Dacier, Matthew Elder, Darrell Kienzle, and Angelos Stavrou. 2012. The MEERKATS Cloud Security Architecture. In *Proceeding of the 2012 32nd International Conference on Distributed Computing Systems Workshops*. IEEE, 446–450. https://doi.org/10.1109/ICDCSW.2012.42

[98] Siddhartha Kumar Khaitan and James D. McCalley. 2015. Design Techniques and Applications of Cyberphysical Systems: A Survey. *IEEE Systems Journal* 9, 2 (2015), 350–365. https://doi.org/10.1109/JSYST.2014.2322503

[99] Faiq Khalid, Semeen Rehman, and Muhammad Shafique. 2020. Overview of Security for Smart Cyber-Physical Systems. In *Security of Cyber-Physical Systems*. Springer, 5–24. https://doi.org/10.1007/978-3-030-45541-5_2

[100] Silvija Kokalj-Filipovic, Morriel Kasher, Michael Zhao, and Predrag Spasojevic. 2020. Detecting Acoustic Backdoor Transmission of Inaudible Messages Using Deep Learning. In *Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning*. ACM, 80–85. https://doi.org/10.1145/3395352.3402629

[101] Subhash Lakshminarayana, E. Veronica Belmega, and H. Vincent Poor. 2021. Moving-Target Defense Against Cyber-Physical Attacks in Power Grids via Game Theory. *IEEE Transactions on Smart Grid* 12, 6 (2021), 5244–5257. https://doi.org/10.1109/TSG.2021.3095083

[102] Subhash Lakshminarayana and David K. Y. Yau. 2021. Cost-Benefit Analysis of Moving-Target Defense in Power Grids. *IEEE Transactions on Power Systems* 36, 2 (2021), 1152–1163. https://doi.org/10.1109/TPWRS.2020.3010365

[103] Robert Larkin, Steven Jensen, Daniel Koranek, Barry Mullins, and Mark Reith. 2021. Towards Dynamically Shifting Cyber Terrain With Software-Defined Networking and Moving Target Defense. In *Proceedings of the 2021 International Conference on Cyber Warfare and Security*. Academic Conferences International Limited, 535–540. https://www.proquest.com/conference-papers-proceedings/towards-dynamically-shifting-cyber-terrain-with/docview/2505728994/se-2

[104] Jay Lee, Behrad Bagheri, and Hung-An Kao. 2015. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters* 3 (2015), 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001

[105] Yeonjoon Lee, Yue Zhao, Jiutian Zeng, Kwangwuk Lee, Nan Zhang, Faysal Hossain Shezan, Yuan Tian, Kai Chen, and XiaoFeng Wang. 2020. Using Sonar for Liveness Detection to Protect Smart Speakers against Remote Attackers. In *Proceedings of the 2020 ACM on Interactive, Mobile, Wearable, and Ubiquitous Technologies*, Vol. 4. ACM, Article 16, 28 pages. https://doi.org/10.1145/3380991

[106] Cheng Lei, Duo-He Ma, and Hong-Qi Zhang. 2017. Optimal Strategy Selection for Moving Target Defense Based on Markov Game. *IEEE Access* (2017), 156–169. https://doi.org/10.1109/ACCESS.2016.2633983

[107] Cheng Lei, Hong-Qi Zhang, Jing-Lei Tan, Yu-Chen Zhang, and Xiao-Hu Liu. 2018. Moving Target Defense Techniques: A Survey. *Security and Communication Networks* 2018, Article 3759626 (2018), 26 pages.

[108] Cheng Lei, Hong-Qi Zhang, Li-Ming Wan, Lu Liu, and Duo-he Ma. 2018. Incomplete information Markov game theoretic approach to strategy generation for moving target defense. *Computer Communications* 116 (2018), 184–199. https://doi.org/10.1016/j.comcom.2017.12.001

[109] Yu Li, Rui Dai, and Junjie Zhang. 2014. Morphing Communications of Cyber-Physical Systems Towards Moving-Target Defense. In *Proceedings of the 2014 IEEE International Conference on Communications*. IEEE, 592–598. https://doi.org/10.1109/ICC.2014.6883383

[110] Zhuohang Li, Cong Shi, Tianfang Zhang, Yi Xie, Jian Liu, Bo Yuan, and Yingying Chen. 2021. Robust Detection of Machine-Induced Audio Attacks in Intelligent Audio Systems with Microphone Array. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1884–1899. https://doi.org/10.1145/3460120.3484755

[111] Jie Lien, Md Abdullah Al Momin, and Xu Yuan. 2022. Attacks on Voice Assistant Systems. In *Security, Data Analytics, and Energy-Aware Solutions in the IoT*. IGI Global, Chapter 4, 61–77. https://doi.org/10.4018/978-1-7998-7323-5.ch004

[112] Hao Liu, Shaodong Wang, and Yuzhe Li. 2022. Event-Triggered Control and Proactive Defense for Cyber-Physical Systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2022), 9 pages. https://doi.org/10.1109/TSMC.2022.3144337

[113] Brian Lucas, Errin W. Fulp, David J. John, and Daniel Cañas. 2014. An Initial Framework for Evolving Computer Configurations as a Moving Target Defense. In *Proceedings of the 9th Annual Cyber and Information Security Research Conference*. ACM, 69–72. https://doi.org/10.1145/2602087.2602100

[114] Douglas C. MacFarland and Craig A. Shue. 2015. The SDN Shuffle: Creating a Moving-Target Defense Using Host-Based Software-Defined Networking. In *Proceedings of the 2nd ACM Workshop on Moving Target Defense*. ACM, 37–41. https://doi.org/10.1145/2808475.2808485

[115] Kaleel Mahmood and Devu Manikantan Shila. 2016. Moving Target Defense for Internet of Things Using Context Aware Code Partitioning and Code Diversification. In *Proceedings of the 2016 IEEE 3rd World Forum on Internet of Things*. IEEE, 329–330. https://doi.org/10.1109/WF-IoT.2016.7845457

[116] Magdi S. Mahmoud, Mutaz M. Hamdan, and Uthman A. Baroudi. 2019. Modeling and control of cyber-physical systems subject to cyber attacks: A survey of recent advances and challenges. *Neurocomputing* 338 (2019), 101–115. https://doi.org/10.1016/j.neucom.2019.01.099

[117] Pratyusa K. Manadhata. 2013. Game Theoretic Approaches to Attack Surface Shifting. In *Moving Target Defense II: Application of Game Theory and Adversarial Modeling*. Springer, 1–13. https://doi.org/10.1007/978-1-4614-5416-8_1

[118] Jian Mao, Shishi Zhu, Xuan Dai, Qixiao Lin, and Jianwei Liu. 2020. Watchdog: Detecting Ultrasonic-Based Inaudible Voice Attacks to Smart Home Systems. *IEEE Internet of Things Journal* 7, 9 (2020), 8025–8035. https://doi.org/10.1109/JIOT.2020.2997779

[119] Jian Mao, Shishi Zhu, and Jianwei Liu. 2020. An inaudible voice attack to context-based device authentication in smart IoT systems. *Journal of Systems Architecture* 104, Article 101696 (2020), 15 pages. https://doi.org/10.1016/j.sysarc.2019.101696

[120] Rômulo Meira-Góes and Stéphane Lafortune. 2020. Moving Target Defense based on Switched Supervisory Control: A New Technique for Mitigating Sensor Deception Attacks. *IFAC-PapersOnLine* 53, 4 (2020), 317–323. https://doi.org/10.1016/j.ifacol.2021.04.031 (15th IFAC Workshop on Discrete Event Systems WODES 2020 — Rio de Janeiro, Brazil, 11-13 November 2020).

[121] Yan Meng, Haojin Zhu, Jinlei Li, Jin Li, and Yao Liu. 2021. Liveness Detection for Voice User Interface via Wireless Signals in IoT Environment. *IEEE Transactions on Dependable and Secure Computing* 18, 6 (2021), 2996–3011. https://doi.org/10.1109/TDSC.2020.2973620

[122] Hossein Mohammadi Rouzbahani, Hadis Karimipour, Abolfazl Rahimnejad, Ali Dehghantanha, and Gautam Srivastava. 2020. Anomaly Detection in Cyber-Physical Systems Using Machine Learning. In *Handbook of Big Data Privacy*. Springer, 219–235. https://doi.org/10.1007/978-3-030-38557-6_10

[123] Iyatiti Mokube and Michele Adams. 2007. Honeypots: Concepts, Approaches, and Challenges. In *Proceedings of the 45th Annual Southeast Regional Conference*. ACM, 321–326. https://doi.org/10.1145/1233341.1233399

[124] Marti Motoyama, Kirill Levchenko, Chris Kanich, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. 2010. Re: CAPTCHAs—Understanding CAPTCHA-Solving Services in an Economic Context. In *Proceedings of the 19th USENIX Conference on Security*. USENIX Association, 28–46. https://www.usenix.org/conference/usenixsecurity10/re-captchas%E2%80%94understanding-captcha-solving-services-economic-context

[125] Renzo E. Navas, Frédéric Cuppens, Nora Boulahia Cuppens, Laurent Toutain, and Georgios Z. Papadopoulos. 2021. MTD, Where Art Thou? A Systematic Review of Moving Target Defense Techniques for IoT. *IEEE Internet of Things Journal* 8, 10 (2021), 7818–7832. https://doi.org/10.1109/JIOT.2020.3040358

[126] Saran Neti, Anil Somayaji, and Michael E Locasto. 2012. Software Diversity: Security, Entropy and Game Theory. In *Proceeding of the 2012 7th USENIX Workshop on Hot Topics in Security*. USENIX Association, 6 pages. https://www.usenix.org/conference/hotsec12/workshop-program/presentation/Neti

[127] Sophie J. Nightingale and Hany Farid. 2022. AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences (PNAS)* 119, 8 (2022). https://doi.org/10.1073/pnas.2120481119

[128] Arman Noroozian, Jan Koenders, Eelco Van Veldhuizen, Carlos H. Ganan, Sumayah Alrwais, Damon McCoy, and Michel Van Eeten. 2019. Platforms in Everything: Analyzing Ground-Truth Data on the Anatomy and Economics of Bullet-Proof Hosting. In *Proceedings of the 28th USENIX Security Symposium*. USENIX Association, 1341–1356. https://www.usenix.org/system/files/sec19-noroozian.pdf

[129] Arman Noroozian, Maciej Korczyński, Carlos Hernandez Gañan, Daisuke Makita, Katsunari Yoshioka, and Michel van Eeten. 2016. Who Gets the Boot? Analyzing Victimization by DDoS-as-a-Service. In *Research in Attacks, Intrusions, and Defenses*. Springer, 368–389. https://doi.org/10.1007/978-3-319-45719-2_17

[130] Jon Oberheide, Michael Bailey, and Farnam Jahanian. 2009. PolyPack: An Automated Online Packing Service for Optimal Antivirus Evasion. In *Proceedings of the 3rd USENIX Conference on Offensive Technologies* (Montreal, Canada) *(WOOT'09)*. USENIX Association, USA, 9. https://jon.oberheide.org/files/woot09-polypack.pdf

[131] Philip O'Kane, Sakir Sezer, and Kieran McLaughlin. 2011. Obfuscation: The Hidden Malware. *IEEE Security & Privacy* 9, 5 (2011), 41–47. https://doi.org/10.1109/MSP.2011.98

[132] Christopher Y. Olivola. 2018. The Interpersonal Sunk-Cost Effect. *Psychological Science* 29, 7 (2018), 1072–1083. https://doi.org/10.1177/0956797617752641

[133] Venkata N. Padmanabhan and Daniel R. Simon. 2003. Secure Traceroute to Detect Faulty or Malicious Routing. *ACM SIGCOMM Computer Communication Review* 33, 1 (2003), 77–82. https://doi.org/10.1145/774763.774775

[134] Aswin Chidambaram Pappa, Aditya Ashok, and Manimaran Govindarasu. 2017. Moving Target Defense for Securing Smart Grid Communications: Architecture, Implementation, and Evaluation. In *Proceeding of the 2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference*. IEEE, 5 pages. https://doi.org/10.1109/ISGT.2017.8085954

[135] Hossein Parastvand, Octavian Bass, Mohammad A. S. Masoum, Airlie Chapman, and Stefan Lachowicz. 2020. Cyber-Security Constrained Placement of FACTS Devices in Power Networks From a Novel Topological Perspective. *IEEE Access* 8 (2020), 108201–108215. https://doi.org/10.1109/ACCESS.2020.3001308

[136] Sergio Pastrana, Alice Hutchings, Daniel Thomas, and Juan Tapiador. 2019. Measuring eWhoring. In *Proceedings of the 2019 Internet Measurement Conference*. ACM, 463–477. https://doi.org/10.1145/3355369.3355597

[137] Wei Peng, Feng Li, Chin-Tser Huang, and Xukai Zou. 2014. A Moving-Target Defense Strategy for Cloud-based Services with Heterogeneous and Dynamic Attack Surfaces. In *Proceeding of the 2014 IEEE International Conference on Communications*. IEEE, 804–809. https://doi.org/10.1109/ICC.2014.6883418

[138] Photon Research Team. 2020. Dark Web Search Engine Kilos: Tipping The Scales In Favor Of Cybercrime. https://www.digitalshadows.com/blog-and-research/dark-web-search-engine-kilos/

[139] Diego G.S. Pivoto, Luiz F.F. de Almeida, Rodrigo da Rosa Righi, Joel J.P.C. Rodrigues, Alexandre Baratella Lugli, and Antonio M. Alberti. 2021. Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review. *Journal of Manufacturing Systems* 58 (2021), 176–192. https://doi.org/10.1016/j.jmsy.2020.11.017

[140] Radha Poovendran. 2010. Cyber–Physical Systems: Close Encounters Between Two Parallel Worlds [Point of View]. *Proc. IEEE* 98, 8 (2010), 1363–1366. https://doi.org/10.1109/JPROC.2010.2050377

[141] Georgios Portokalidis and Angelos D. Keromytis. 2011. Global ISR: Toward a Comprehensive Defense Against Unauthorized Code Execution. In *Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats*. Springer, 49–76. https://doi.org/10.1007/978-1-4614-0977-9_3

[142] Bradley Potteiger, Abhishek Dubey, Feiyang Cai, Xenofon Koutsoukos, and Zhenkai Zhang. 2022. Moving target defense for the security and resilience of mixed time and event triggered cyber–physical systems. *Journal of Systems Architecture* 125, Article 102420 (2022), 11 pages. https://doi.org/10.1016/j.sysarc.2022.102420

[143] Bradley Potteiger, Zhenkai Zhang, and Xenofon Koutsoukos. 2018. Integrated Instruction Set Randomization and Control Reconfiguration for Securing Cyber-Physical Systems. In *Proceedings of the 5th Annual Symposium and Bootcamp on Hot Topics in the Science of Security*. ACM, Article 5, 10 pages. https://doi.org/10.1145/3190619.3190636

[144] Yihao Qiu, Jun Wu, Shahid Mumtaz, Jianhua Li, Anwer Al-Dulaimi, and Joel J. P. C. Rodrigues. 2021. MT-MTD: Muti-Training based Moving Target Defense Trojaning Attack in Edged-AI network. In *Proceedings of the 2021 IEEE International Conference on Communications*. IEEE, 6 pages. https://doi.org/10.1109/ICC42927.2021.9500545

[145] Michail G. Rachavelias. 2019. Online financial crimes and fraud committed with electronic means of payment—a general approach and case studies in Greece. *ERA Forum* 19, 3 (2019), 339–355. https://doi.org/10.1007/s12027-018-0519-2

[146] Mohammad Ashiqur Rahman, Ehab Al-Shaer, and Rakesh B. Bobba. 2014. Moving Target Defense for Hardening the Security of the Power System State Estimation. In *Proceedings of the 1st ACM Workshop on Moving Target Defense*. ACM, 59–68. https://doi.org/10.1145/2663474.2663482

[147] Gayathri Rajakumaran and Neelanarayanan Venkataraman. 2021. Performance assessment of hybrid MTD for DoS mitigation in public cloud. *International Journal of Intelligent Networks* 2 (2021), 140–147. https://doi.org/10.1016/j.ijin.2021.09.003

[148] Kui Ren, Tianhang Zheng, Zhan Qin, and Xue Liu. 2020. Adversarial Attacks and Defenses in Deep Learning. *Engineering* 6, 3 (2020), 346–360. https://doi.org/10.1016/j.eng.2019.12.012

[149] Rodrigo Roman, Jianying Zhou, and Javier Lopez. 2013. On the features and challenges of security and privacy in distributed internet of things. *Computer Networks* 57, 10 (2013), 2266–2279. https://doi.org/10.1016/j.comnet.2012.12.018

[150] Guo Rong-xiao, Tian Ji-wei, Wang Bu-hong, and Shang Fu-te. 2020. Cyber-Physical Attack Threats Analysis for UAVs from CPS Perspective. In *Proceedings of the 2020 International Conference on Computer Engineering and Application*. IEEE, 259–263. https://doi.org/10.1109/ICCEA50009.2020.00063

[151] Jeff Rowe, Karl N. Levitt, Tufan Demir, and Robert Erbacher. 2012. Artificial Diversity as Maneuvers in a Control Theoretic Moving Target Defense. In *Proceeding of the 2012 National Symposium on Moving Target Research*. 10 pages. https://web.cs.ucdavis.edu/~rowe/papers/MTRSymposium2012_final.pdf

[152] Mohamed Samir, Mohamed Azab, and Effat Samir. 2021. SD-CPC: SDN Controller Placement Camouflage based on Stochastic Game for Moving-target Defense. *Computer Communications* 168 (2021), 75–92. https://doi.org/10.1016/j.comcom.2020.11.019

[153] Sebastian Schrittwieser, Stefan Katzenbeisser, Johannes Kinder, Georg Merzdovnik, and Edgar Weippl. 2016. Protecting Software through Obfuscation: Can It Keep Pace with Progress in Code Analysis? *Comput. Surveys* 49, 1, Article 4 (2016), 37 pages. https://doi.org/10.1145/2886012

[154] Sailik Sengupta, Tathagata Chakraborti, and Subbarao Kambhampati. 2018. MTDeep: Boosting the Security of Deep Neural Nets Against Adversarial Attacks with Moving Target Defense. In *Proceeding of the Workshops of the 32nd AAAI Conference on Artificial Intelligence*. AAAI, 376–383. https://www.aaai.org/ocs/index.php/WS/AAAIW18/paper/view/16930

[155] Sailik Sengupta, Ankur Chowdhary, Dijiang Huang, and Subbarao Kambhampati. 2018. Moving Target Defense for the Placement of Intrusion Detection Systems in the Cloud. In *Proceeding of the 2018 9th International Conference on Decision and Game Theory for Security*. Springer, 326–345. https://doi.org/10.1007/978-3-030-01554-1_19

[156] Sailik Sengupta, Ankur Chowdhary, Abdulhakim Sabur, Adel Alshamrani, Dijiang Huang, and Subbarao Kambhampati. 2020. A Survey of Moving Target Defenses for Network Security. *IEEE Communications Surveys & Tutorials* 22, 3 (2020), 1909–1941. https://doi.org/10.1109/COMST.2020.2982955

[157] Sailik Sengupta, Satya Gautam Vadlamudi, Subbarao Kambhampati, Adam Doupé, Ziming Zhao, Marthony Taguinod, and Gail-Joon Ahn. 2017. A Game Theoretic Approach to Strategy Generation for Moving Target Defense in Web Applications. In *Proceeding of the 2017 6th International Conference on Autonomous Agents and Multiagent Systems*, Vol. 1. ACM, 178–186. https://doi.org/10.5555/3091125.3091155

[158] Dimitrios Serpanos. 2018. The Cyber-Physical Systems Revolution. *Computer* 51, 3 (2018), 70–73. https://doi.org/10.1109/MC.2018.1731058

[159] Md Hasan Shahriar, Mohammad Ashiqur Rahman, Nur Imtiazul Haque, and Badrul Chowdhury. 2021. DDAF: Deceptive Data Acquisition Framework against Stealthy Attacks in Cyber-Physical Systems. In *Proceedings of the IEEE 45th Annual Computers, Software, and Applications Conference*. IEEE, 725–734. https://doi.org/10.1109/COMPSAC51774.2021.00104

[160] Reza Shakeri, Mohammed Ali Al-Garadi, Ahmed Badawy, Amr Mohamed, Tamer Khattab, Abdulla Khalid Al-Ali, Khaled A. Harras, and Mohsen Guizani. 2019. Design Challenges of Multi-UAV Systems in Cyber-Physical Applications: A Comprehensive Survey and Future Directions. *IEEE*

Communications Surveys & Tutorials 21, 4 (2019), 3340–3385. https://doi.org/10.1109/COMST.2019.2924143

[161] L. S. Shapley. 1953. Stochastic Games. *Proceedings of the National Academy of Sciences* 39, 10 (1953), 1095–1100. https://doi.org/10.1073/pnas.39.10.1095

[162] Dilli Prasad Sharma, Dong Seong Kim, Seunghyun Yoon, Hyuk Lim, Jin-Hee Cho, and Terrence J. Moore. 2018. FRVM: Flexible Random Virtual IP Multiplexing in Software-Defined Networks. In *Proceedings of the 2018 17th IEEE International Conference on Trust, Security and Privacy In Computing and Communications and the 12th IEEE International Conference on Big Data Science and Engineering*. IEEE, 579–587. https://doi.org/10.1109/TrustCom/BigDataSE.2018.00088

[163] Meng Shen, Hao Yu, Liehuang Zhu, Ke Xu, Qi Li, and Jiankun Hu. 2021. Effective and Robust Physical-World Attacks on Deep Learning Face Recognition Systems. *IEEE Transactions on Information Forensics and Security* 16 (2021), 4063–4077. https://doi.org/10.1109/TIFS.2021.3102492

[164] Matthew Sherburne, Randy Marchany, and Joseph Tront. 2014. Implementing Moving Target IPv6 Defense to Secure LoWPAN in the Internet of Things and Smart Grid [Extended Abstract]. In *Proceedings of the 9th Annual Cyber and Information Security Research Conference*. ACM, 37–40. https://doi.org/10.1145/2602087.2602107

[165] Cong Shi, Yan Wang, Yingying Chen, Nitesh Saxena, and Chen Wang. 2020. WearID: Low-Effort Wearable-Assisted Authentication of Voice Commands via Cross-Domain Comparison without Training. In *Proceedings of the 2020 Annual Computer Security Applications Conference*. ACM, 829–842. https://doi.org/10.1145/3427228.3427259

[166] Chawin Sitawarin, Arjun Nitin Bhagoji, Arsalan Mosenia, Mung Chiang, and Prateek Mittal. 2018. DARTS: Deceiving Autonomous Cars with Toxic Signs. arXiv preprint arXiv:1802.06430. https://doi.org/10.48550/arXiv.1802.06430

[167] Qun Song, Zhenyu Yan, and Rui Tan. 2021. DeepMTD: Moving Target Defense for Deep Visual Sensing against Adversarial Examples. *ACM Transactions on Sensor Networks* 18, 1, Article 5 (oct 2021), 32 pages. https://doi.org/10.1145/3469032

[168] Wissem Soussi, Maria Christopoulou, George Xilouris, and Gürkan Gür. 2021. Moving Target Defense as a Proactive Defense Element for Beyond 5G. *IEEE Communications Standards Magazine* 5, 3 (2021), 72–79. https://doi.org/10.1109/MCOMSTD.211.2000087

[169] Takeshi Sugawara, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu. 2020. Light Commands: Laser-Based Audio Injection Attacks on Voice-Controllable Systems. In *Proceedings of the 29th USENIX Security Symposium*. USENIX Association, 2631–2648. https://www.usenix.org/conference/usenixsecurity20/presentation/sugawara

[170] Jing-lei Tan, Cheng Lei, Hong-qi Zhang, and Yu-qiao Cheng. 2019. Optimal strategy selection approach to moving target defense based on Markov robust game. *Computers & Security* 85 (2019), 63–76. https://doi.org/10.1016/j.cose.2019.04.013

[171] Jinglei Tan, Hengwei Zhang, Hongqi Zhang, Hao Hu, Cheng Lei, and Zhenxiang Qin. 2021. Optimal temporospatial strategy selection approach to moving target defense: A FlipIt differential game model. *Computers & Security* 108, Article 102342 (2021), 15 pages. https://doi.org/10.1016/j.cose.2021.102342

[172] Jinglei Tan (谭晶磊), Hengwei Zhang (张恒巍), Hongqi Zhang (张红旗), Hui Jin (金辉), and Cheng Lei (雷程). 2020. Optimal Strategy Selection Approach of Moving Target Defense based on Markov Time Game / 基于 Markov 时间博弈的移动目标防御最优策略选取方法. *Journal on Communications* /《通信学报》 41, 1 (2020), 42–52. https://doi.org/10.11959/j.issn.1000-436x.2020003

[173] Kurt Thomas, Dmytro Iatskiv, Elie Bursztein, Tadek Pietraszek, Chris Grier, and Damon McCoy. 2014. Dialing Back Abuse on Phone Verified Accounts. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 465–476. https://doi.org/10.1145/2660267.2660321

[174] Michael Thompson, Nathaniel Evans, and Victoria Kisekka. 2014. Multiple OS Rotational Environment: An Implemented Moving Target Defense. In *Proceedings of the 2014 7th International Symposium on Resilient Control Systems*. IEEE, 6 pages. https://doi.org/10.1109/ISRCS.2014.6900086

[175] Matheus Torquato, Paulo Maciel, and Marco Vieira. 2021. *Analysis of VM Migration Scheduling as Moving Target Defense against Insider Attacks*. ACM, 194–202. https://doi.org/10.1145/3412841.3441899

[176] Matheus Torquato and Marco Vieira. 2020. Moving target defense in cloud computing: A systematic mapping study. *Computers & Security* 92, Article 101742 (2020), 11 pages. https://doi.org/10.1016/j.cose.2020.101742

[177] UK Finance and Cifas. 2021. if it sounds to good to be true, it probably is. - money mules. https://www.moneymules.co.uk/

[178] Jacob Ulrich, Jacob Drahos, and Manimaran Govindarasu. 2017. A Symmetric Address Translation Approach for a Network Layer Moving Target Defense to Secure Power Grid Networks. In *Proceeding of the 2017 Resilience Week*. IEEE, 163–169. https://doi.org/10.1109/RWEEK.2017.8088667

[179] Satya Gautam Vadlamudi, Sailik Sengupta, Marthony Taguinod, Ziming Zhao, Adam Doupé, Gail-Joon Ahn, and Subbarao Kambhampati. 2016. Moving Target Defense for Web Applications using Bayesian Stackelberg Games (Extended Abstract). In *Proceedings of the 2016 international Conference on Autonomous Agents & Multiagent systems*. ACM, 1377–1378. https://dl.acm.org/doi/10.5555/2936924.2937168

[180] John Wadleigh, Jake Drew, and Tyler Moore. 2015. The E-Commerce Market for "Lemons": Identification and Analysis of Websites Selling Counterfeit Goods. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1188–1197. https://doi.org/10.1145/2736277.2741658

[181] Bing Wang (王滨), Liang Chen (陈靓), Yaguan Qian (钱亚冠), Yankai Guo (郭艳凯), Qiqi Shao (邵琦琦), and Jiamin Wang (王佳敏). 2021. Moving target defense against adversarial attacks / 面向对抗样本攻击的移动目标防御. *Chinese Journal of Network and Information Security* /《网络与信息安全学报》7, 1 (2021), 113–120. http://www.infocomm-journal.com/cjnis/CN/10.11959/j.issn.2096-109x.2021012

[182] Emily Wenger, Josephine Passananti, Arjun Nitin Bhagoji, Yuanshun Yao, Haitao Zheng, and Ben Y Zhao. 2021. Backdoor Attacks against Deep Learning Systems in the Physical World. In *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 6206–6215. https://doi.org/10.1109/CVPR46437.2021.00614

[183] Monica T. Whitty. 2013. The Scammers Persuasive Techniques Model: Development of a Stage Model to Explain the Online Dating Romance Scam. *The British Journal of Criminology* 53, 4 (04 2013), 665–684. https://doi.org/10.1093/bjc/azt009

[184] Yun Wu (吴云), Hongjiao Xue (薛宏蛟), Wei Zhu (朱玮), and Fan Luo (罗璠). 2021. A Study of Money Laundering Through Virtual Currencies / 虚拟货币洗钱问题研究: 固有风险、类型分析与监管应对. *Financial Regulation Research* /《金融监管研究》 118,

10 (2021), 1–19. https://jrjg.cbpt.cnki.net/WKB2/WebPublication/paperDigest.aspx?paperID=9a8e940f-fac9-4e7a-9aca-6df6b01482ca

[185] Haitao Xu, Daiping Liu, Haining Wang, and Angelos Stavrou. 2015. E-Commerce Reputation Manipulation: The Emergence of Reputation-Escalation-as-a-Service. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1296–1306. https://doi.org/10.1145/2736277.2741650

[186] Junjie Xu. 2021. Research on Cyberspace Mimic Defense based on Dynamic Heterogeneous Redundancy Mechanism. *Journal of Computer and Communications* 9, 7 (2021), 1–7. https://www.scirp.org/pdf/jcc_2021070215332857.pdf

[187] Xiaoyu Xu, Hao Hu, Yuling Liu, Hongqi Zhang, and Dexian Chang. 2021. An Adaptive IP Hopping Approach for Moving Target Defense Using a Light-Weight CNN Detector. *Security and Communication Networks* 2021, Article 8848473 (2021), 17 pages. https://doi.org/10.1155/2021/8848473

[188] Jean-Paul A. Yaacoub, Ola Salman, Hassan N. Noura, Nesrine Kaaniche, Ali Chehab, and Mohamad Malli. 2020. Cyber-physical systems security: Limitations, issues and future trends. *Microprocessors and Microsystems* 77 (2020), 103201. https://doi.org/10.1016/j.micpro.2020.103201

[189] Chen Yan, Guoming Zhang, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2021. The Feasibility of Injecting Inaudible Voice Commands to Voice Assistants. *IEEE Transactions on Dependable and Secure Computing* 18, 3 (2021), 1108–1124. https://doi.org/10.1109/TDSC.2019.2906165

[190] Hao Yang, Kun Du, Yubao Zhang, Shuai Hao, Haining Wang, Jia Zhang, and Haixin Duan. 2021. Mingling of Clear and Muddy Water: Understanding and Detecting Semantic Confusion in Blackhat SEO. In *Computer Security – ESORICS 2021: 26th European Symposium on Research in Computer Security, Darmstadt, Germany, October 4–8, 2021, Proceedings, Part I*. Springer, 263–284. https://doi.org/10.1007/978-3-030-88418-5_13

[191] Kaichen Yang, Tzungyu Tsai, Honggang Yu, Tsung-Yi Ho, and Yier Jin. 2020. Beyond Digital Domain: Fooling Deep Learning based Recognition System in Physical World. In *Proceedings of the 2020 AAAI Conference on Artificial Intelligence*, Vol. 34. AAAI, 1088–1095. https://doi.org/10.1609/aaai.v34i01.5459

[192] Michael Yip, Nigel Shadbolt, and Craig Webber. 2013. Why Forums?: An Empirical Analysis into the Facilitating Factors of Carding Forums. In *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, 453–462. https://doi.org/10.1145/2464464.2464524

[193] Michael Yip, Craig Webber, and Nigel Shadbolt. 2013. Trust among cybercriminals? Carding forums, uncertainty and implications for policing. *Policing and Society* 23, 4 (2013), 516–539. https://doi.org/10.1080/10439463.2013.780227

[194] Jianzhou You, Shichao Lv, Yue Sun, Hui Wen, and Limin Sun. [n.d.]. In *Proceedings of the 2021 IEEE International Conference on Communications, title=HoneyVP: A Cost-Effective Hybrid Honeypot Architecture for Industrial Control Systems, year=2021, numpages=6, doi=10.1109/ICC42927.2021.9500567, publisher=IEEE,*.

[195] Kara Zaffarano, Joshua Taylor, and Samuel Hamilton. 2015. A Quantitative Framework for Moving Target Defense Effectiveness Evaluation. In *Proceedings of the 2nd ACM Workshop on Moving Target Defense*. ACM, 3–10. https://doi.org/10.1145/2808475.2808476

[196] Vahid Zangeneh and Mehdi Shajari. 2018. A cost-sensitive move selection strategy for moving target defense. *Computers & Security* 75 (2018), 72–91. https://doi.org/10.1016/j.cose.2017.12.013

[197] Kimberly Zeitz, Michael Cantrell, Randy Marchany, and Joseph Tront. 2017. Designing a Micro-Moving Target IPv6 Defense for the Internet of Things. In *Proceeding of the 2017 IEEE/ACM 2nd International Conference on Internet-of-Things Design and Implementation*. IEEE, 179–184. https://ieeexplore.ieee.org/document/7946873

[198] Kimberly Zeitz, Michael Cantrell, Randy Marchany, and Joseph Tront. 2018. Changing the Game: A Micro Moving Target IPv6 Defense for the Internet of Things. *IEEE Wireless Communications Letters* 7, 4 (2018), 578–581. https://doi.org/10.1109/LWC.2018.2797916

[199] Guoming Zhang, Xiaoyu Ji, Xinfeng Li, Gang Qu, and Wenyuan Xu. 2021. EarArray: Defending against DolphinAttack via Acoustic Attenuation. In *Proceedings of the 2021 Conference on Network and Distributed Systems Security Symposium*. Internet Society, 14 pages. https://doi.org/10.14722/ndss.2021.24551

[200] Jiliang Zhang and Chen Li. 2020. Adversarial Examples: Opportunities and Challenges. *IEEE Transactions on Neural Networks and Learning Systems* 31, 7 (2020), 2578–2593. https://doi.org/10.1109/TNNLS.2019.2933524

[201] Shaohu Zhang and Anupam Das. 2021. HandLock: Enabling 2-FA for Smart Home Voice Assistants Using Inaudible Acoustic Signal. In *Proceedings of the 24th International Symposium on Research in Attacks, Intrusions and Defenses*. ACM, 251–265. https://doi.org/10.1145/3471621.3471866

[202] Yue Zhao, Hong Zhu, Ruigang Liang, Qintao Shen, Shengzhi Zhang, and Kai Chen. 2019. Seeing isn't Believing: Towards More Robust Adversarial Attack Against Real World Object Detectors. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1989–2004. https://doi.org/10.1145/3319535.3354259

[203] Jianjun Zheng and Akbar Siami Namin. 2019. Enforcing Optimal Moving Target Defense Policies. In *Proceeding of the 2019 IEEE 43rd Annual Computer Software and Applications Conference*, Vol. 1. IEEE, 753–759. https://doi.org/10.1109/COMPSAC.2019.00112

[204] Man Zhou, Zhan Qin, Xiu Lin, Shengshan Hu, Qian Wang, and Kui Ren. 2019. Hidden Voice Commands: Attacks and Defenses on the VCS of Autonomous Driving Cars. *IEEE Wireless Communications* 26, 5 (2019), 128–133. https://doi.org/10.1109/MWC.2019.1800477

[205] Caixia Zhou (周彩霞). 2008. An Analysis of Spyware Industry Chain / 对 "流氓软件" 产业链的理论分析. *Journal of Guangdong University of Finance & Economics* /《广东商学院学报》 98, 3 (1 2008), 26–31. https://d.wanfangdata.com.cn/periodical/gdsxyxb200803005

[206] Minghui Zhu, Zhisheng Hu, and Peng Liu. 2014. Reinforcement Learning Algorithms for Adaptive Cyber Defense against Heartbleed. In *Proceedings of the 1st ACM Workshop on Moving Target Defense*. ACM, 51–58. https://doi.org/10.1145/2663474.2663481

[207] Quanyan Zhu and Tamer Başar. 2013. Game-Theoretic Approach to Feedback-Driven Multi-Stage Moving Target Defense. In *Proceeding of the 2013 4th International Conference on Decision and Game Theory for Security*. Springer, 246–263. https://doi.org/10.1007/978-3-319-02786-9_15

[208] Rui Zhuang, Su Zhang, Alex Bardas, Scott A. DeLoach, Xinming Ou, and Anoop Singhal. 2013. Investigating the Application of Moving Target Defenses to Network Security. In *Proceeding of the 2013 6th International Symposium on Resilient Control Systems*. IEEE, 162–169. https://doi.org/10.1109/ISRCS.2013.6623770

[209] Rui Zhuang, Su Zhang, Scott A. DeLoach, Xinming Ou, Anoop Singhal, et al. 2012. Simulation-based Approaches to Studying Effectiveness of Moving-Target Network Defense. In *Proceeding of the 2012 National Symposium on Moving Target Research*.

12 pages. https://csrc.nist.gov/CSRC/media/Projects/Measuring-Security-Risk-in-Enterprise-Networks/documents/mtd_paper_final.pdf

[210] Lukáš Zobal, Dušan Kolář, and Radek Fujdiak. 2019. Current State of Honeypots and Deception Strategies in Cybersecurity. In *Proceedings of the 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*. IEEE, 9 pages. https://doi.org/10.1109/ICUMT48472.2019.8970921

[211] Vít Šembera, Masarah Paquet-Clouston, Sebastian Garcia, and Maria Jose Erquiaga. 2021. Cybercrime Specialization: An Exposé of a Malicious Android Obfuscation-as-a-Service. In *Proceedings of the 2021 IEEE European Symposium on Security and Privacy Workshops*. IEEE, 213–226. https://doi.org/10.1109/EuroSPW54576.2021.00029