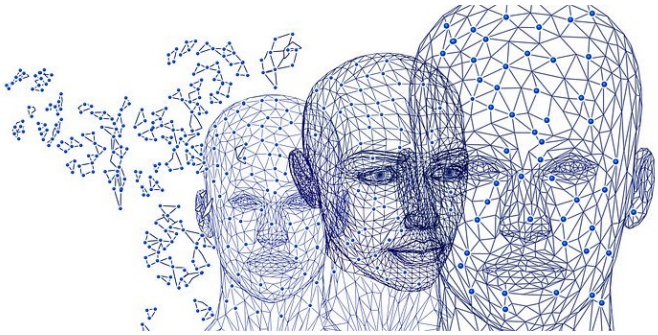
February 2022, Issue Code NL-2022-6

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

Editorial

This issue of the Digital Data Deception (DDD) Technology Watch Newsletter focuses on generation and detection of graph-based data deception techniques. The content of the newsletter resulted from keyword-based searches into the English scientific database Scopus and the China National Knowledge Infrastructure (CNKI) database. We have excluded the following search results: academic papers published prior to 2020, publications in languages other than English or Chinese, and publications which did not include original research or were not review papers. Our searches resulted in 465 papers from Scopus and 34 papers from CNKI. After several rounds of screening and labelling, 31 papers were selected for inclusion in the newsletter.

Three main themes emerged from the searches. They are:

- *False Data Injection* in terms of attack and defence (i.e., detection and resilience to such attacks); this theme is covered in Section 1.
- *Fake News, Fake Reviews and Fake Accounts* in terms of detection and mitigation; this theme is covered in Section 2.
- *Network Topology* and its use in generating fake topology, and topology obfuscation; this theme is covered in Section 3.

We hope you enjoy reading this issue. Feedback is always welcome and should be directed to ddd-newsletter@kent.ac.uk.

List of Acronyms

- AC: Alternating Current
- ARMA: Auto-Regressive Moving Average
- BDD: Bad Data Detection
- C-FATH: Community-based Framework with ATtention mechanism for large-scale Heterogeneous graphs
- CCNN: Center-Cluster-Neural-Network
- CCPT: Coordinated Cyber–Physical Topology
- CNN: Convolutional Neural Network
- DAGA-NN: Domain-Adversarial and Graph-Attention Neural Network
- D-FACTS: Distributed Flexible AC Transmission System
- DC: Direct Current
- DDoS: Distributed Denial of Service
- DNN: Deep Neural Network
- DL: Deep Learning
- EET: Enhanced Evidence Theory
- ET: Evidence Theory
- FDIA: Fake Data Injection Attack
- GAN: Generative Adversarial Network
- GF: Graph Filter
- GNN: Graph Neural Network
- HDGCN: Heterogeneous Deep Convolutional Network
- HMTD: Hidden Moving Target Defence
- IMIA-HCRF: Identify Malicious Injection Attacks using Higher Order Conditional Random Fields
- k-NN: k-Nearest Neighbour
- KB: Knowledge Base
- LDA: Latent Dirichlet Allocation
- LFA: Link Flooding Attack
- LUN: Labelled Unreliable News
- MTD: Moving Target Defence
- NDG: News Detection Graph
- NN: Neural Networks
- ProTo: Proactive Topology Obfuscation
- PSSE: Power System State Estimation
- RF: Random Forest
- RL: Reinforcement Learning
- RNN: Recurrent Neural Network
- SDN: Software Defined Networking
- SE: State Estimation
- SLN: Satirical and Legitimate News
- SOMPS-Net: SOcial graph with Multihead attention and Publisher information and news Statistics Network
- TTL: Time To Live
- UAV: Unmanned Aerial Vehicle
- UFA: Unveiling Fake Accounts

1. False Data Injection

False Data Injection Attacks (FDIA) were originally envisioned in the smart grid domain. The idea is that an attacker stealthily compromises sensor readings by introducing undetected errors into calculations of state variables. Other applications, such as smart healthcare, e-banking, defence and governance, have also been found to be subject to similar methods. In complex adaptive systems for critical applications, FDIA has become one of the top-priority challenges that need to be addressed to provide safety and gain or maintain user trust. It is necessary to raise awareness about these attacks and adopt better mechanisms to counter them. In this section we present examples of FDIAs and countermeasures to defend against them.

1.1. Attack

In this subsection, we summarise four recently published articles which proposed FDIA in different application domains such as recommender systems, energy systems and smart grids.

Wu et al. [41] proposed a graph convolution-based *generative shilling attack* for recommender systems which exploits the robustness of those systems. Shilling attacks happen when attackers inject fake user profiles into recommendation systems to influence the recommendation results. For instance, in e-commerce, attackers can inject fake purchase records and reviews to affect users' choices and judgement on a specific item. The authors adopted a primitive shilling attack paradigm, a sub-class of FDIA, that assigns items for fake users based on user-item interactions. This was further used to construct co-rated items for fake user profiles in the attack model via sampling, and to generate fake ratings using a Deep Learning (DL) based model. The authors deployed a generative adversarial network (GAN) that learns the real rating distribution to generate fake ratings. Moreover, the authors employed a tailored graph convolution structure that leveraged from the correlations between co-rated items to smooth the fake ratings and enhance their feigned authenticity. An overview of the proposed method is shown in Figure 1. It consists of four parts. First, items are sampled from an item-item graph to decide which items should be rated in a fake user profile. Then, an adversarial architecture powered by graph convolution is trained to generate fake ratings for these sampled items. Next, the fake user profile is assembled by combining the sampled items and the target items (i.e., items to be artificially promoted). The method repeats the above process to finally select a number of fake user profiles and inject them into real ratings to deceive the customer or users.

Editorial Comments

The effectiveness of shilling attacks [41] is evaluated using *hit ratio* i.e., the ratio of target items (graph nodes) that appear in real users' top ten recommendation lists. Evaluations based on two public datasets (i.e., Douban and Ciao) demonstrated technical feasibility for building a more powerful and intelligent attack model with a much reduced cost. The authors just named the datasets without providing specific links to them, therefore, it remained unclear whether they used a dataset variant.

Anwar et al. [1] proposed a methodology to determine a power grid topology based on a data-driven approach using measurement signals and a sparse FDIA construction. Utilising the properties of power grid, including its positive semi-definite nature with the null space property, the authors modelled the topology estimation problem as a constraint optimisation problem with the sparsity regularisation. They also demonstrated how to reveal the power grid topology only taking the measurement signals into account. Comparative evaluation using graph theory measures (i.e., *degree centrality and eigenvector centrality, closeness centrality, average degree of neighbouring nodes, and graph energy*) indicated that the proposed solution reveals the topology with high accuracy for each graph theory measure, where the accuracy is evaluated using adjacency matrices of the actual topology and the estimated topology.

Wang et al. [39] proposed a framework with two main components: (1) a Reinforcement Learning (RL) based Coordinated Cyber–Physical Topology (CCPT) attack strategy, and (2) a deep RL based approach to determine and to identify the minimal attack resources. The CCPT attack strategy is firstly used to physically trip (i.e., interrupt power supply) a transmission line. Then the outage signal of the tripped line is masked and a new fake line outage

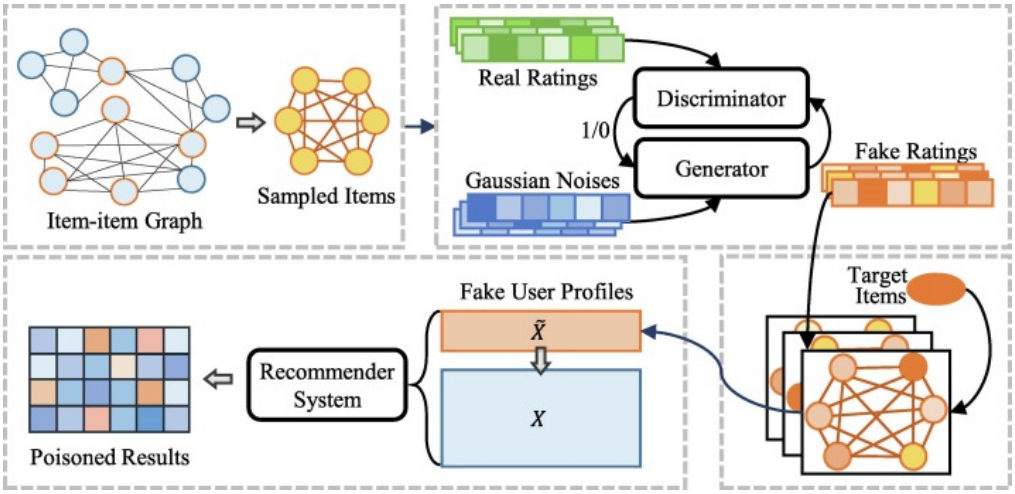


Figure 1: Architectural overview and signal flow graph of the proposed graph convolution-based generative shilling attack design by Wu et al. [41]. The first block in the pipeline are items sampled from an item-item graph to decide which items should be rated in a fake user profile. The second block is an adversarial architecture powered by graph convolution which is trained to generate ratings for these sampled items. This is followed by the assembling of fake user profile by combining the sampled items and assigned high ratings for promotion.

for another transmission line is created by using the CCPT attack. Secondly, in order to block the outage signal of the tripped line and create the fake outage signal (i.e., the FDIA attack) with limited attack resources, a deep RL-based approach is used to determine the minimal attack resources. Since from the attacker's perspective, it is necessary to conduct feasible attacks with limited resources, the proposed deep RL-based method enables identification of the minimal attack resources. The feasibility of such attacks indicates a vulnerability of critical transmission lines and is useful to use in quality assurance design for ensuring protection methods against such attacks.

Editorial Comments

According to Wang et al. [39], the topology of the power system can be represented by a bus–branch model $G=(V,L)$. A physical disconnection of a transmission line Pl by an attacker alters the topology of the power system to $\mathbf{G_1} = (\mathbf{V}, \mathbf{L}\backslash\{\mathbf{Pl}\})$, where V is set of buses and L is set of transmission lines.

Wang (王胜锋) et al. [40] proposed an FDIA scheme to impact the electricity market based on topology tampering. Modern grids acquire grid data through numerous sensors such as Remote Terminal Units or Phasor Measurement Units. Grid operators can estimate generation costs and assign appropriate market prices using this data. Attackers profit from the market by compromising a certain number of sensors to attack the system and sending fake measurements to the power monitoring centre. Since the monitoring systems usually allow a certain rate of false positives, attackers take advantage of this feature to remain undetected. The authors also validated the proposed scheme and showed that it is more stealthy and more profitable than other similar FDIAs.

Editorial Comments

As we have seen, both graph based [1, 41] and DL based [39] models can be used to design FDIA attacks. However, the performance of both is subject to the target architecture and the nature of the attack. Nevertheless, in both scenarios FDIA attacks represent a risk in applications such as smart grids and recommender systems. Such attacks should be taken into consideration before deploying any architecture in real-life.

1.2. Defence (Detection)

This subsection reviews recently proposed defence mechanisms against FDIAs in applications

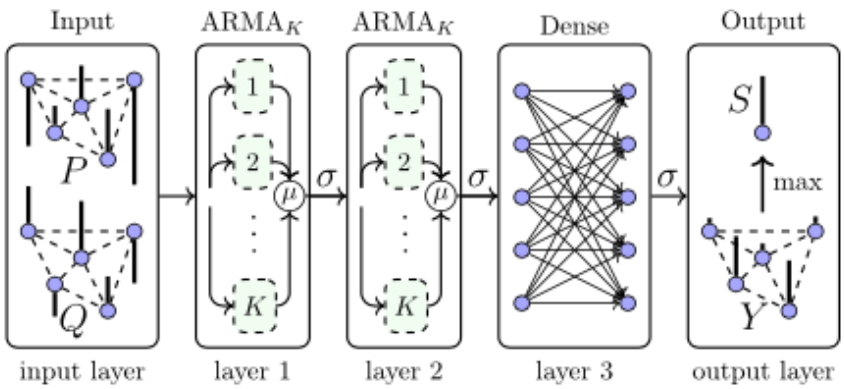


Figure 2: Architectural overview of the proposed design by [4], where each $ARMA_K$ layer consists of K parallel layers. Each one of the K dashed blocks in an $ARMA_K$ layer corresponds to an $ARMA_1$ block depicted with a dashed block. While complex power injections P, Q and predicted attack probabilities Y, S at the node and graph level are visualised with thick bars at each node; activation and mean value functions are represented with σ and μ, respectively. Moreover, the nodes and edges of the graphs represent buses and transmission lines, respectively.

such as power systems and smart grids.

Jorjani et al. [17] proposed an algorithm to detect FDIAs for alternating current (AC) state estimation (SE). The algorithm applied outlier detection techniques on the SE results and then detected a possible attack using a graph theory based approach. The main idea of the proposed method is the fact that, due to power flow equations, the attacker needs to change the measurements obtained from the target bus and some of its adjacent buses of branches to ensure a successful attack. As proposed by the authors, first, the probability distribution of the estimated state variables and measurement variations were used to detect the outliers in the current SE results. Then, the method looked for adjacencies among the identified outliers and constructed a graph with them. The attack is detected if there is enough evidence that the identified outliers are adjacent and related to each other. This procedure of analysing the detected outliers based on graph theory concepts makes the proposed method capable of detecting AC FDIAs. Furthermore, the proposed approach used historical data to calculate the amount of difference between two consecutive SE results. Using these derived variations, the probability distribution of each of the system SE variables and measurements were calculated. After the injection of false data into the system, the amount of variation in system variables would be higher than in normal conditions. This was used to identify the individual outliers in SE results for a given time step. The performance of detecting normal and attack samples (outliers) was evaluated by considering different combinations of nodes and predefined threshold values. For each of these combinations, the balanced accuracy of the detection algorithm was calculated.

Boyaci et al. [4] proposed a GNN-based model to address FDIA in power systems by integrating the underlying graph topology of the grid and spatial correlations of its measurement data. The purpose is to jointly detect and localise FDIAs, whereas the full AC power flow equations are employed to address the physical architecture and power flow of the network. The model leveraged from the auto-regressive moving average (ARMA) graph filters (GFs). Such filters can better adapt to sharp changes in the spectral domain due to their rational filter composition compared to polynomial GFs (e.g., Chebyshev). Filter weights were learned automatically during training by an end-to-end data-driven approach. The architecture of their proposed ARMA GNN based detector and localiser with three hidden layers can be seen in Figure 2.

Boyaci et al. [5] presented a generic, localised and stealthy attack generation methodology to address FDIA attacks in power grids. They proposed a GNN based scalable and real-time FDIA detection method, which combines model-driven and data-driven approaches. The method incorporated the inherent physical connections of modern AC power

grids and explored the spatial correlations of measurements (e.g., power flow and meter readings). An overview of their proposed architecture is shown in Figure 3. The blue boxes represent smart grids and their operations run by an operator, while red and green boxes denote functional blocks of the attacker and the defender, respectively. Note that the operator gets attack measurements z_a instead of the original z_o due to the FDIA. The defender, on the contrary, tries to detect possible attacks by using z_a. Here, a represents the attack vector, $x^{\wedge}$ and $x^{\vee}$ denote the estimated (original) state vector and the false data injected state vector.

Editorial Comments

Overall, the studies conducted in applications such as power grid [4, 5] suggest that architectural differences in neural networks (NN) play an important role in the performance of FDIA detection. Multi-Layer Perceptron based detectors have the problem of over fitting due to the dense connections, hence they have poor generalisation. This can be overcome by using dropout layers. RNNs underperform because they work on sequenced data and the node values cannot be transformed into sequenced values. CNNs are able to model temporal and spatial relations of the input in Euclidean space but the graph structure of power grids cannot be modelled in Euclidean space, therefore, CNNs lag behind GNN-based models. GNNs are more suitable for graph data because they can show adjacency in node data.

Yang et al. [43] presented a unified detection approach called *identify malicious injection attacks using higher order conditional random fields* (IMIA-HCRF) to address FDIAs in recommender systems. The authors explored attributes of both nodes and edges of the behaviour association graph. They proposed to incorporate unary potential and pairwise higher order conditional random fields for informative representations of rating and co-visitation behaviours. This was further used to develop a unified detection approach to identify two types of FDIA attacks, namely *profile injection attacks* and *co-visitation injection attacks*. In particular, the authors applied a divide-and-conquer strategy to detect the FDIA attacks for online recommender systems. Experimental results on both synthetic data and real-world data (ML-100K, Amazon, LibraryThing, and TripAdvisor) showed that the elimination of disturbed data, determination of dense behaviours, and potential segmentation exhibit considerable stability and discriminability among nodes for detecting malicious injection behaviours.

1.3. Defence (Resilience)

Sometimes only identifying attacks, such as FDIAs in a peer-to-peer network, is not enough. Further information about those attacks is crucial, such as their provenance, development, ownership, location, changes to system components and personnel and/or processes involved. This information enables system administrators to trace attacks back to the root cause and take action to address them, e.g., by revoking access, changing policies and collecting evidence for legal action.

Ge et al. [12] presented a provenance-aware distributed trust model called UAVN-pro to address such challenges for Unmanned Aerial Vehicle (UAV) networks. Their aim was to achieve accurate peer-to-peer trust assessment and maximise the delivery of correct messages received by destination nodes while minimising the message delay and communication cost under such resource-constrained network environments. Their proposed method leveraged the history of ownership, referred to as the provenance of the message transmitted on the network. The behaviour of message creators and operators can be effectively evaluated based on message integrity and can be used to generate the observational evidence. The evidence considered was of two types: (1) positive evidence, where the observer observed the positive behaviour of the node and made a positive recommendation, which does not rule out false positives, and (2) negative evidence, where the observed values of the node reflected the node's negative behaviour, which does not rule out false negatives. The authors observed evidence based on their distributed trust evaluation model and then identified malicious nodes to revoke or isolate them from the network. In order to reduce the computational cost in the UAV network where malicious nodes exist, they adopted a data-driven method to ensure secure communications. Their experiment showed that UAVN-pro is compatible with the existing UAV network routing protocols and can effectively identify attacks, such as black hole, grey hole, message modification,

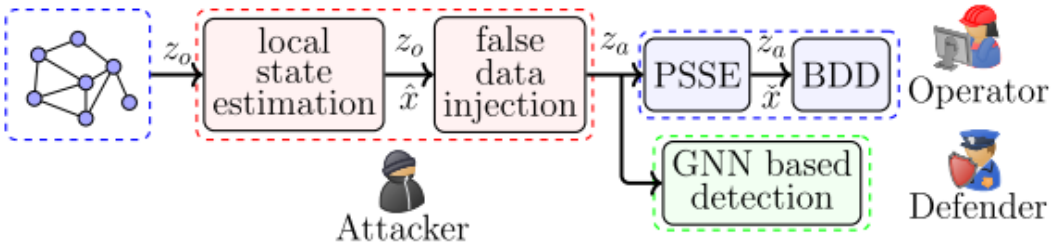


Figure 3: Architectural overview and signal flow graph of the proposed design in [5]. The blue boxes represent smart grids and their operations run by operator, while red and green boxes depict the functional blocks of the attacker and defender, respectively. It should be noted that the operator gets attacked measurements z_a instead of the original ones z_o due to the FDIA. The defender, on the other hand, tries to detect the FDIA attacks by using z_a. The bad data detection (BDD) module takes z_a as input and, along with Power system state estimation (PSSE), is used to detect the "bad" measurement data.

fake recommendation, and fake identity in UAV networks. The experimental results showed that their provenance-aware trust model outperforms the existing UAV network routing protocols, and can effectively identify a range of FDIA attacks.

2. Fake News, Fake Reviews and Fake Accounts

Social media has paved the way for easier and faster information propagation, including false and misleading information. In addition, the increasing use of e-commerce platforms led fraudsters to target the users of these platforms through fake reviews. Recent advances in graph-based machine learning approaches (e.g., graph attention neural networks) enabled graph structures to be used in analysing social networks to detect such anomalies more effectively. This section provides some recent studies utilising graphs for detecting fake news, fake reviews, and fake accounts.

2.1. Fake News Detection

Yu and Long [44] reviewed recent studies on graph-based fake news detection. The authors focused on modelling different types of information (e.g., content, context, and temporal propagation) in social media as a graph, and identifying measurements or properties of the graph or subgraphs (e.g., dense subgraph mining, graph statistics, and temporal graph structure) to characterise target anomalies. They covered recent fake news detection methods for different fake news detection scenarios, including unimodal, dynamic (dual-modal), and multi-modal scenarios, depending on the type(s) of information utilised in detection.

Yuan et al. [45] proposed a framework, named *Domain-Adversarial and Graph-Attention Neural Network* (DAGA-NN), to detect fake news across events and domains (e.g., politics). The architecture of DAGA-NN is shown in Figure 4. The feature extractor identifies and combines textual and image features from the corresponding text and image data in the news. The combined features are used by the domain discriminator to separate the domain-specific information through a min-max game, and graph-attention-based fake news classifier to determine if the news is fake. The graph used in the classifier is an undirected graph where each node represents a news document, and each edge represents the cosine similarity between the textual contents of the two connected news nodes. The authors also introduced a learning strategy for optimising graph attention networks to improve the performance of the proposed framework. DAGA-NN was evaluated on two multimedia datasets from Twitter and Weibo. The results indicated that the proposed model achieved 88-90% F1-score for the Twitter dataset [3], and 95% F1-score for the Weibo dataset.

Editorial Comments

The framework proposed by Yuan et al. [45] is capable of detecting fake news containing multiple events/domains since it uses both text and image parts of the news for detection. This makes the proposed solution more suitable for real-world cases when compared to traditional methods that suffer from limited efficiency in detecting fake news across domains due to insufficient representativeness [11], considering the nature of fake news online.

Dhanasekaran et al. [8] presented a graph-based framework, called *SOcial graph with Multi-head attention and Publisher information and news Statistics Network* (SOMPS-Net), for early detection of fake health news. SOMPS-Net leveraged social engagements of associated users along with the news publisher details and social media statistics of the news article. The authors evaluated the framework with the FakeHealth data repository [7], which includes two health news datasets: HealthStory and HealthRelease. Their results indicated that SOMPS-Net achieved 79.6% F1-score and 72.7% accuracy. In addition, the authors observed that the proposed model reached 75% F1-score within 4 hours from the broadcast of fake news articles, and its maximum F1-score level, ~80%, after 8 hours. This indicated that the model can detect fake news at its early stages.

Kang et al. [18] leveraged the connections between multiple news items (e.g., their relevance in time, content, topic and source) for fake news detection. The authors constructed a heterogeneous graph (i.e., containing different types of nodes and edges), called *News Detection Graph* (NDG), to integrate multivarious information as different node types, including news content, domain, reviews, and source. They also proposed a Heterogeneous Deep Convolutional Network (HDGCN) to learn deep representation of news nodes. To evaluate the proposed framework, the authors used Fakeddit [30] and Weibo [27] datasets. The experiment results indicated that the proposed framework achieved 83.2-88.5% F1-score on the Fakeddit dataset, and 96% F1-score on the

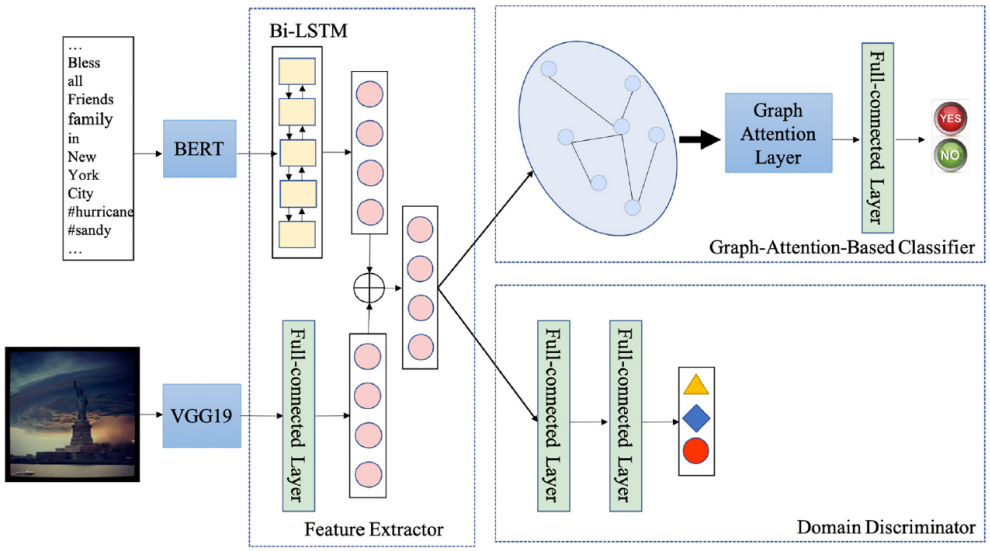


Figure 4: Architecture of the DAGA-NN framework, proposed by Yuan et al. [45]. Textual and image features are identified and fused by the feature extractor. Then, domain-specific information is separated via a min-max game between the feature extractor and the domain discriminator. Finally, a graph-attention-based fake news classifier determines if the news is fake.

Weibo dataset.

Editorial Comments

The framework proposed by Kang et al. [18] considered the connection between multiple news items for fake news detection. This enabled identification of previously-detected fake news through similarity measures.

Hu et al. [16] introduced an end-to-end graph neural model, *CompareNet*, that compared the news to the Wikipedia knowledge base (KB) through contextual entities to identify fake news. The authors first constructed a directed heterogeneous document graph incorporating topics and entities, as shown in Figure 5. The graph contained nodes for topics, sentences, and contextual entities. The topics were extracted from the sentences via unsupervised Latent Dirichlet Allocation (LDA) while the entities were identified and mapped to Wikipedia through the entity linking tool TAGME. Then, they developed heterogeneous graph attention networks to learn topic-enriched news representations and contextual entity representations, based on the document graph. For the next step, the authors leveraged an entity comparison network to compare the contextual entity representations with the corresponding KB-based entity representations. Finally, they combined the obtained entity comparison features with the topic-enriched news document representation for detecting fake news. The authors used the Satirical and Legitimate News (SLN) [34], and the Labelled Unreliable News (LUN) [32] datasets for evaluating their model. The experiment results showed that CompareNet achieved 89% F1-score for the SLN dataset, and 69% F1-score for the LUN dataset.

Editorial Comments

The source code of the model proposed by Hu et al. [16] is publicly available at https://github.com/ytc272098215/FakeNewsDetection.

He (何韩森) and Sun (孙国梓) [14] proposed a fake news detection model based on feature aggregation, called *Center-Cluster-Neural-Network* (CCNN), which combines the advantages of CNN and RNN. The authors divided the overall process of the CCNN model into four steps: (1) *Data acquisition and annotation*: they used the Politifact [38], Fake News detection from Kaggle and Buzzfeed [35] datasets. (2) *Text pre-processing*: they segmented words and turned them into structured data. (3) *Feature extraction*: feature extraction of the input text was performed by CNN. At the same time, the RNN was used to collect the global temporal fea-

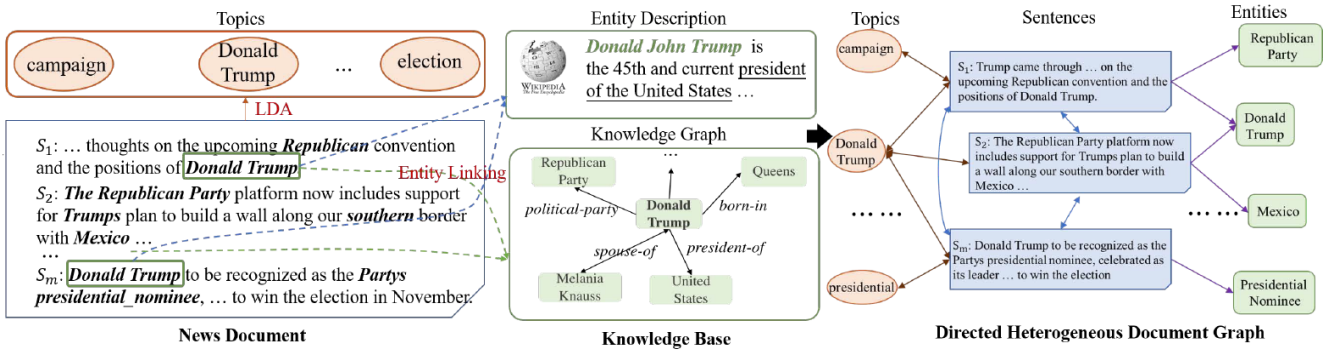


Figure 5: An example of directed heterogeneous document graph incorporating topics and entities from the study by Hu et al. [16]. From each sentence in the news document, topics are extracted with unsupervised LDA. Then, the entities are identified and mapped to Wikipedia via an entity linking tool.

tures of the input text. Then the features were combined to form the fully connected layer. (4) *Classification*: the model classifier was trained with the fully connected layer and the improved uniform loss function to distinguish between fake and genuine news. The results indicated that the proposed model achieved an 80.5% F1-score for the three aforementioned datasets.

2.2. Fake Review Detection

Manaskasemsak et al. [28] proposed a semi-supervised graph partitioning approach, called *BeGP*, and its extension *BeGPX*, for fake reviewer detection. The main idea of BeGP was to first construct a behavioural graph where nodes were behavioural feature vectors, i.e., combination of statistical features derived from reviewers, and edges were weighted by calculating the cosine similarity of both feature vectors. Then, the algorithm performed graph partitioning by utilising a Greedy Algorithm approach to identify suspicious reviewers, starting from a set of known fake reviewers. BeGPX, on the other hand, enhanced BeGP by employing additional analysis of semantic content and emotions expressed in reviews with the help of a Deep Neural Network (DNN). As shown in Figure 6, it benefited from review and reviewer characteristics obtained from the review metadata, as well as word embedding and emotion representation generated from the review contents. The authors evaluated the proposed methods on two datasets collected from Yelp.com, i.e., YelpNYC and YelpZip. The results showed that both proposed methods outperformed several baseline methods in terms of precision metric.

Editorial Comments

BeGP and BeGPX, introduced by Manaskasemsak et al. [28], required a small set of known fake reviewers. In addition, their main goal was to detect fake reviewers, rather than fake reviews, with the assumption that all reviews posted by fake reviewers are fake.

Wang et al. [37] proposed a *Community-based Framework with ATtention mechanism for large-scale Heterogeneous graphs* (C-FATH) to detect fake reviews in e-commerce. The authors aimed to solve two challenges in fake review detection (or more generally, fraud detection): *mixture of structure-inconsistency* because of the extremely unbalanced positive and negative samples, and *mixture of content-inconsistency* due to the difference between various item categories. While community-based filtering was applied to solve the former challenge, the latter was alleviated with similarity-based sampling. C-FATH was evaluated on two public datasets, YelpZip and YelpNYC, as well as two large-scale datasets collected from JD.com (one of the largest e-commerce platforms in China) for fake shopping orders and spam reviews. The datasets collected from JD.com contained over 5.8M reviews on ~870k products from ~4.3M users. The experiment results showed that the proposed framework achieved 68-87% F1-score, depending on the dataset.

Rathore et al. [33] introduced a framework for detection of fake reviewer groups on the Google Play Store, based on a modified semi-supervised clustering method, PCKMeans, and a DeepWalk approach for the topological structure of the re-

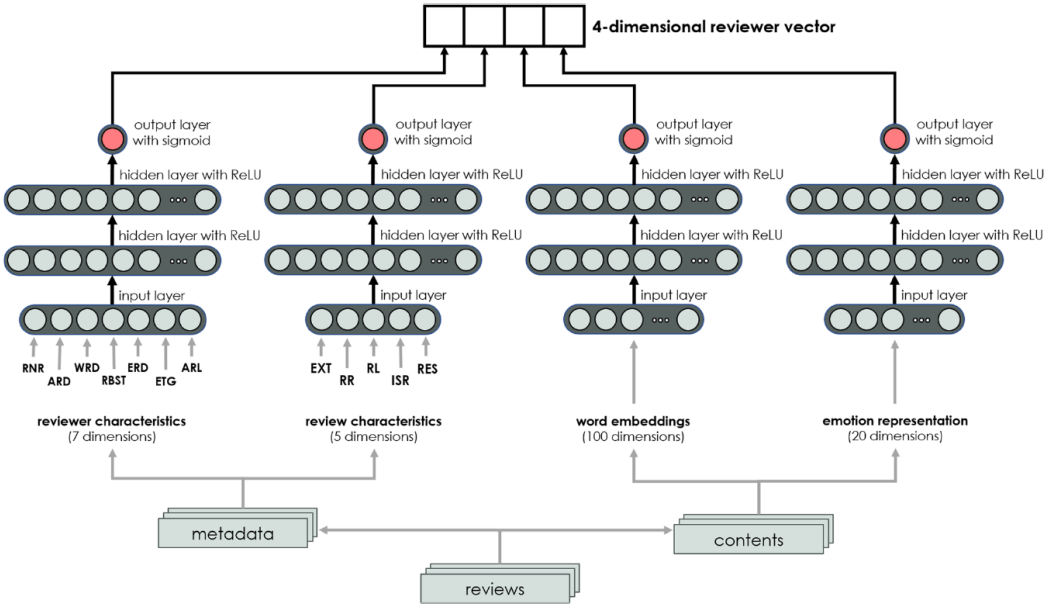


Figure 6: Reviewer node representation in BeGPX from the study by Manaskasemsak et al. [28]. Unlike BeGP, which only used statistical features, BeGPX learned review and reviewer characteristics from the review metadata, and extracted word embedding and emotion representation from the reviews' contents.

viewer's graph. The authors considered reviewers as nodes, and edges representing the number of products commonly reviewed by both a and b. To evaluate the framework, the authors recruited 23 fraud freelancers from Fiverr – a marketplace for freelance services – and collected review samples of the participants through a custom application installed on their local computers. The obtained review samples were used as the ground-truth for the experiments. In addition, they collected review data from 640 Google Play apps, reviewed by over 38k unique reviewer-IDs. The results showed that the proposed framework detected fake reviewer groups both from the ground-truth data and the entire (both labelled and unlabelled) data with ~67% accuracy.

2.3. Fake Accounts Detection

Li et al. [21] presented a model for shilling attack detection, called *SpDetector*, by fusing hypergraph spectral features. As illustrated in Figure 7, the authors used three main features to train their DNN-based model: the *user spectral features* extracted from the user hypergraph, *item similarity offset* between the user's highest-rated and lowest-rated items extracted from the item hypergraph, and the mean square error between real ratings and predicted ratings, i.e., *rating prediction error*. While the user hypergraph contains n user nodes and m item hyperedges, the item hypergraph involves m item nodes and n user hyperedges. SpDetector was evaluated on MovieLens and Amazon [42] datasets, and the results indicated that it achieved over 95% F1-score. Furthermore, the authors showed that the proposed model was robust under different attack sizes.

Liang et al. [22] proposed the *Unveiling Fake Accounts* (UFA) method to detect fake accounts immediately after they were registered. The authors observed that fake accounts were likely to cluster on outlier registration patterns, e.g., when using the same IP address or phone number, and being active at midnight, through a measurement study. Thus, they extracted features which revealed outlier registration patterns for their unsupervised learning algorithm. A registration graph was constructed to capture the correlation between registration accounts. Since fake accounts were likely to be densely connected in the registration graph, the authors utilised a community detection algorithm to cluster the registration accounts into communities, and considered all accounts in a community as fake accounts if the community size exceeded a threshold. Experiments on seven WeChat registration datasets showed that UFA achieved 86.62% F1-score. In addition, UFA was deployed by the WeChat application for more than one year, and the authors reported that it detected 500k fake accounts per day with a precision

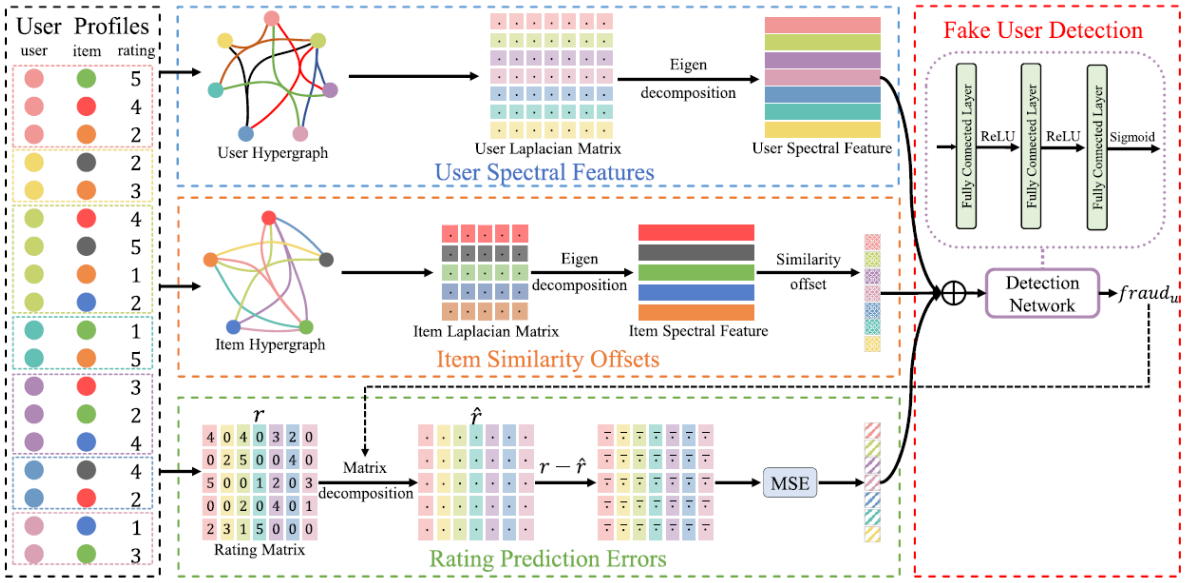


Figure 7: Illustration of the SpDetector model proposed by Li et al. [21] consisting of three crucial features extraction components: user spectral features component, item similarity offsets component, and rating prediction errors component. The DNN takes the features extracted by the three components for detection.

of ~93% on average.

Editorial Comments

The system proposed by Liang et al. [22] was deployed by WeChat for more than one year to detect fake accounts. The authors shared the details regarding the deployment process in the paper, which could provide a good example of utilising a fake account detector in a real-world application.

Bebensee et al. [2] proposed an approach for identifying fake accounts and bots by exploiting differences between ego networks (i.e., networks where a central node exists and all other nodes are directly connected to this central node) of benign users and fake accounts as well as weak links between communities of real users and communities of fake accounts. For this purpose, the authors extracted *egograph features* and *aggregated neighbourhood features* by analysing a dataset expanded from the Cresci-2018 dataset [6], containing 4.6M Twitter user profiles. The generated features included median outdegree of predecessors, median favourites of predecessors, median status count of predecessors, median account age of predecessors, median favourites of successors, and the egograph density and reciprocity. The authors evaluated their method on Cresci-2018 dataset [6], and the experiment results indicated that random forest (RF) and neural network-based bot detection models improved when the introduced features were applied.

Editorial Comments

Although the features introduced by Bebensee et al. [2] improved the performance of existing fake account detection models, the improvement in F1-scores is quite limited (from ~87% to ~88%).

Poupko et al. [31] suggested two approaches from graph theory, i.e., graph conductance and vertex expansion, for a social network to grow without admitting too many sybils (i.e., fake identities). While graph conductance assumes that real users tend to distrust fake identities and aims to measure the connectivity of a graph by quantifying the minimal edge cut, the latter relies on the assumption that there are not too many fake identities in the community and aims to measure the connectivity of a graph by quantifying the minimal vertex cut. The authors aimed to keep the fraction of sybils below a certain threshold through the construction of a new social network where the users can authenticate each other,

rather than discovering sybils in existing social networks. Based on this, the authors modelled a digital community via a trust graph where vertices represented identities, and edges represented trust relations between the owners of the identities. Finally, the authors applied the proposed methods to two real network samples taken from Facebook, containing over 4k connected nodes, and DBLP (DBLP: Digital Bibliography & Library Project), containing ~317k nodes and over 1M edges, respectively. They reported that the proposed methods achieved the binding of the population of fake identities, although the proposed methods were not efficient enough for larger communities.

Editorial Comments

The study conducted by Poupko et al. [31] assumed the construction of a new, more trustworthy social network by trying to prevent fake identities from engaging in the first place, rather than focusing on fake account detection on existing social networks. In addition, the authors reported that more efficient methods are needed, especially when the network contains bigger communities.

3. Network Topology

Network topology shows how computers are connected in a network. Attackers typically conduct extensive network reconnaissance to discover exploitable vulnerabilities on target networks before a formal attack. Distributed denial of service (DDoS) attacks target nodes sending a large number of data packets by controlling bots. Similarly, a link flooding attack (LFA) targets the critical links in network topology and impacts normal operation. These attacks can lead to Internet service interruption, power outages, unreliable IoT devices, and more. Moreover, it is often challenging to balance network security (i.e. defence) and network quality (i.e. reliability). In this section, we present some papers that addressed such attacks.

3.1. Fake Topology Generation

Evidence theory (ET) based attacks are commonly used to target Base Stations (BS) in wireless sensor networks. The adversary may locate the BS by applying traffic analysis, i.e., they intercept radio transmissions and correlate them to identify the presence of a BS. The ET attack model only uses spatial aspects of intercepted transmissions in order to deduce knowledge about data routes. Generally, ET includes observation of packet transmissions to form the links between a source and a destination, and is used widely in the literature as a traffic analysis attack model.

Ebrahimi and Younis [9] presented an Enhanced Evidence Theory (EET) that correlates the intercepted transmissions both spatially and temporally. The basic idea behind EET is that it factors the temporal relations between two network nodes by elevating the corresponding spatially inferred evidence through the addition of a bonus value. The added time-based correlation feature increases the accuracy in converging to the location of the BS. The effectiveness of EET is dependent on the selection of the bonus value and the temporal correlation window within which two transmissions are deemed to be related. The authors have provided guidelines on how to tune these two parameters for better effectiveness. The current countermeasures for ET-based attacks are not resilient against the temporal correlation of EET. Ebrahimi and Younis [9] proposed to counter the EET-based attack model through a novel countermeasure known as *Assisted Deception* (AD). It aims to be a node-aware, distributed, and EET-resilient scheme that coordinates transmissions among neighbouring nodes to inject deceptive packets in a timely manner. AD not only prevents the time correlation of data transmissions, but it also disturbs the EET analysis and tricks the adversary not to observe the location of the BS, by providing a satisfactory camouflage to the BS.

Editorial Comments

(1) The new anonymity metrics Success Rate, BS Rank, and Safe Distance were introduced by Ebrahimi and Younis [9] to better gauge the BS anonymity and the impact of countermeasures.
(2) Experimental results [9] showed that current countermeasures could not sustain the anonymity of the BS against an adversary that employs EET. Yet, the countermeasure introduced using fake data injection (i.e., Assisted Deception) has potential to safeguard against the EET.

3.2. Network Topology Obfuscation

Network topology obfuscation techniques can hide the topology of the target network to prevent attackers from reconnaissance of the network. Based on the different network topology inference techniques used by attackers, two obfuscation approaches can help protect the real network topology. The first approach is to modify and/or reroute data packets sent by attackers against traceroute. The second approach is to induce fake metrics in the topology of target networks (e.g., link metrics, end-to-end path metrics) – a process known as *Network Tomography*. Liu et al. [24] proposed a network topology obfuscation mechanism called *AntiTomo* which meets the following four design criteria. (1) **Deception**: the obfuscated network topology should have been deceptive for attackers but followed the fundamental characteristic of a real-world network. (2) **Security**: compared to the real network topology, the obfuscated network topology should differ in key nodes and links. i.e., attackers will not threaten the real network based on the obfuscated network topology. (3) **Low cost**: the obfuscated network will have minimal impact on real

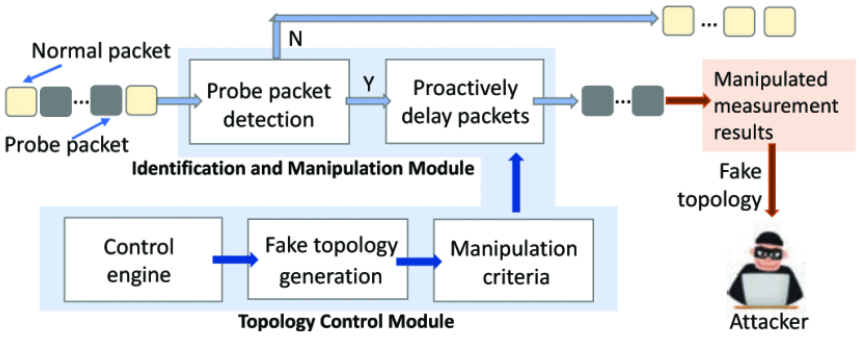


Figure 8: The system architecture of *ProTO* by Hou et al. [15]. The system consists of two main components: an identification and manipulation module and a topology control module. The former is used to detect the probing packet. The latter provides a control interface for network administrators, which can be used to create packet delay related criteria for the fake topology.

users. (4) **High efficiency**: in order to achieve the best defence effect, the obfuscated topology needs to be replaced regularly. This requires that the generation speed of the obfuscated topology should be faster than the replacement cycle. The AntiTomo randomly generates some candidate trees leading to candidate forests used for the obfuscated network topology. A linear programming model is then used to calculate the most suitable candidate tree and achieve the four above-mentioned design criteria through optimisation. The performance evaluation results show that AntiTomo can efficiently defend against the tomography-based network topology reconnaissance.

Hou et al. [15] proposed the *Proactive Topology Obfuscation* (ProTo) system, which involves a mechanism to detect reconnaissance activities and a topology obfuscation approach to prevent and/or slow down an attacker's ability to obtain the real network topology. Figure 8 shows the system architecture of ProTo. In order to detect the probing behaviour, a machine learning-based classification framework is proposed. The framework uses the lightweight k-Nearest Neighbour (k-NN) approach to detect the probe packets. Compared with traditional k-NN, light-k-NN is more efficient for real-time network devices because it can dynamically adjust the weights adaptively. ProTo can also implement a lazy learning update strategy based on voting. For topology obfuscation, in addition to ensuring that attackers only get the fake topology, the probe packets are also delayed impacting the attacker's ability to do reconnaissance. The obfuscated topology should meet the requirement that attackers are not able to reverse engineer the real topology from a mathematical perspective.

Editorial Comments

Experimental results showed that, compared with no protection and ProTo [15], AntiTomo [24] has advantages in terms of security, efficiency and cost. However, this scheme only considers single-source attacks. In a real-world network scenario, attacks could be multi-sourced and simultaneous, and this is more challenging to defend.
Even in the performance evaluation results, AntiTomo has an advantage, but the authors also proposed some useful evaluation metrics. Detection rate and false alarm rate have been introduced as performance metrics. The similarity of the real network and the inferred network topology is used as the effectiveness metric. The performance cost is defined as the ratio of the extra latency to the normal latency of normal packets going through the network.

LFA is a type of DDoS attack which utilises a known network topology to attack target links. Figure 9 shows the attacker's process of choosing target links in a crossfire attack. This is a type of LFA where large amounts of low-rate traffic sent by bots are relayed to the target links, causing the target area to lose connectivity. The premise of this attack is that the attacker knows the topology of the target network. Liu et al. [25] designed a defence system called *NetObfu* by using network topology obfuscation techniques. There are four modes to obfuscate

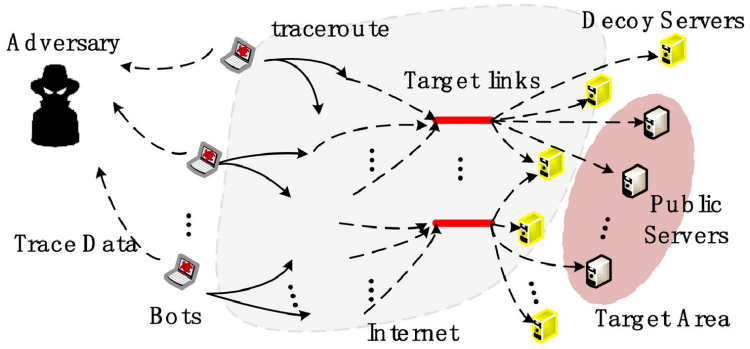


Figure 9: The process of choosing target links in crossfire attack [25]. A large amount of low-rate traffic sent by bots relayed the target links (in red), causing the target area to lose connectivity.

the network topology: (1) hide the node, (2) disguise a node as another node, (3) disguise a node as two or more nodes or (4) keep the node. First, NetObfu generates the virtual obfuscated network topology. Then, according to the obfuscated network topology, the data plane of NetObfu controls the TTL (Time To Live) of the probe packet and modifies its return address in order to interfere with the attacker's ability to perform reconnaissance. In addition, NetObfu supports honey links (i.e., links used to deceive attackers) used to identify bots and to collect logs for further investigation. The authors believe that deploying 20%-30% of nodes in a software-defined network (SDN) environment can protect the majority of the network while keeping low latency for normal users at runtime.

Liu (刘亚群) et al. [26] also proposed a topology obfuscation mechanism named TopoObf, which can effectively defend against attacks such as LFA in the reconnaissance phase based on SDN. The execution process of TopoObfu includes two steps: Node Updates and Route Updates. The former refers to adding virtual links to the real network topology. The latter allows attackers to obtain obfuscated topology by modifying routing rules. The authors designed an algorithm to make the importance of links on all paths, from the entry node to the exit node, as similar as possible, in order to reduce the probability of critical links being identified without impacting normal users. The experiment was carried out under the SDN based network with Topology Zoo datasets [19].

Editorial Comments

Both experiments of NetObfu and TopoObfu were carried out using SDN based networks with topology zoo datasets [19]. For NetObfu, the deployment of this solution is limited by the number of SND nodes in the real world. That is, the number of SDN nodes should not be less than a certain proportion. Furthermore, the trustworthiness of the virtual network topology needs to be considered in order to obfuscate the attacker's information. For TopoObfu, it is not a machine learning based solution. TopoObfu updates a small number of routers in the traditional network to SDN switches according to the node update algorithm, which has lower cost.

Existing IoT devices often have limited hardware resources, therefore, some security solutions cannot be deployed on them. Unlike intrusion detection systems, which react by identifying malicious activity in the system, Ge et al. [13] proposed an integrated approach by leveraging both cyber deception (i.e., a decoy system) and moving target defence (MTD) (i.e., network topology shuffling) for intrusion prevention. This approach can prevent reconnaissance activities before an attack happens. The authors described five aspects of the solution, as follows. (1) *Deployment of decoy nodes*: Decoy nodes are created based on the deployment of real nodes in each virtual LAN and connections are added to some randomly selected real nodes. (2) *When-to-shuffle*: Performing MTD at fixed time intervals is called the fixed shuffle. Correspondingly, the time of the random shuffle is random, but the average time interval is fixed. The execution of adaptive shuffle depends

on the current system security vulnerabilities and state. Hybrid shuffle combines both fixed and adaptive shuffle. (3) *How-to-shuffle*: Genetic Algorithm-based optimisations are used with related metrics (e.g., mean time to a security failure, defence cost) to make the "best" adjustment. Decoy path-based optimisation maximises the number of decoy paths for each IoT node. Random shuffling is a baseline strategy, which is probability-based. (4) *System Measurements*: Related metrics have been proposed for measuring system security, performance, and service availability. (5) *Security Modelling*: The authors designed a graphical security model for security analysis in SDN networks. The workflow has also been proposed for the defence system.

Editorial Comments

Ge et al. [13] used several evaluation metrics, including the number of attracting paths toward decoy targets, mean time to a security failure, defence cost, packet delivery ratio and service availability. The paper concluded that the proposed proactive defence method could be applied to any IoT environment.

Smart grids, as modern transmission networks, also suffer from FDI attacks. The distributed flexible AC transmission system (D-FACTS) can actively perturb to invalidate the grid knowledge gained by an attacker through probing. A hidden MTD (HMTD) is an advanced MTD method, which is good to detect FDI attacks and is more stealthy for attackers. However, in the power system, the optimal planning and operation of MTD can be further improved. Liu and Wu [23] proposed a depth-first-search-based planning algorithm and the direct current (DC)- and alternating current(AC) -HMTD operation models. In practice, the D-FACTS equipment must be installed on a subset of transmission lines during the planning phase in order to implement the HMTD. However, the planning algorithm designed by the authors is more flexible and can deploy HMTD in the case of uncertain D-FACTS setpoints and different conditions in the planning phase. In terms of operation models, those models propose optimisations for the trade-off of saving power generation costs and hiding MTD. They can also be integrated into the existing energy management system.

Editorial Comments

As reported by Liu and Wu [23], both the DC- and AC-HMTD operation models overcome the shortcomings of the exiting HMTD operation and minimise the power generation costs.

The idea of MTD is to make the system dynamic by increasing the randomness and uncertainty of the system. Cyber deception defence mechanisms have developed from the idea of honeypots. The existing network deception method mostly adjusts the honeynet network structure according to the current network environment, which relies heavily on malicious traffic detection technology and is passive. Those solutions are often unable to defend against attackers with anti-honeypot capabilities effectively. Therefore, Gao (高春刚) et al. [10] proposed an MTD enhanced cyber deception defence system based on SDN. Figure 10 shows the system architecture, which consists of three main components, including a virtual network topology module, an IP randomisation module, and a deception server. The virtual network topology module is mainly responsible for generating the virtual network topology and distributing the flow table according to the specifications. The IP randomisation module coordinates the address translation of hosts and decoy nodes in the network. The deception server spoofs malicious scanners by crafting a response based on the specification.

Editorial Comments

Gao (高春刚) et al. [10] evaluated the performance of the MTD enhanced cyber deception defence system. The results showed that the minimum time for the attacker to find a real vulnerable host is the same, but the maximum time and the average time are delayed by more than seven times. Combined with IP address randomisation, not only does it further delay the time for an attacker to succeed, but it also reduces the probability of an attacker's success by 83%. However, the limitation is that the network latency increases by 2.2% - 10.4%.

Governments are seeking to profile users based on their online behaviour (interactions). This be-

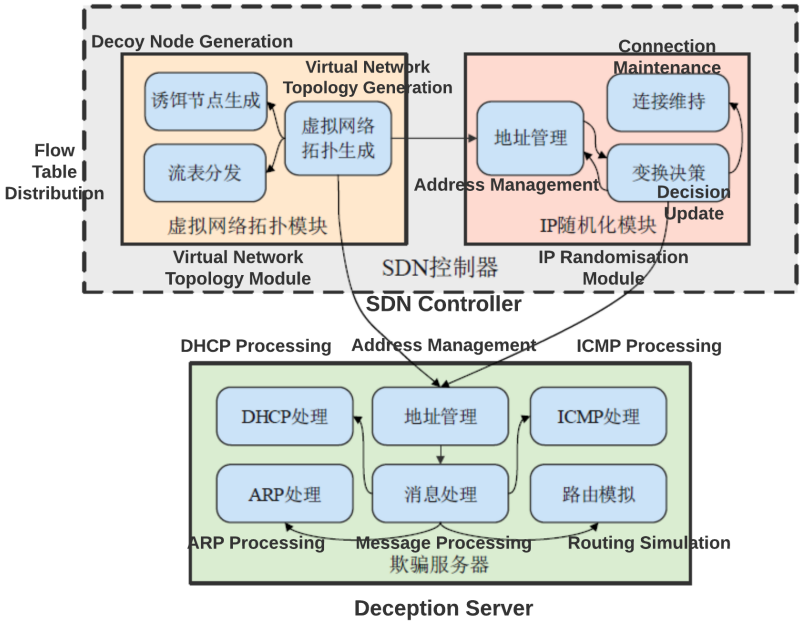


Figure 10: MTD enhanced cyber deception defence model proposed by Gao (高春刚) et al. [10]. This model consists of three main components: a virtual network topology module, an IP randomisation module, and a deception server. The virtual network topology is strictly separated from the real network. By deploying the virtual network topology on the demilitarised zone server, the IP addresses of the intranet hosts detected by the attacker are all virtual.

haviour is often represented using graphs (networks of interactions) that, despite containing sensitive information, are made publicly available for various purposes. However, developing effective methods to anonymise (obfuscate) datasets of user interactions before making them public is of great importance due to privacy and security concerns. Graph anonymisation aims to reduce the ability of an attacker to identify the nodes of a graph by obfuscating its structural information. K-anonymity is a widely used method for anonymisation, it aims at making each node indistinguishable from at least other $k-1$ nodes. Once the identity of a node in a graph is revealed, other potentially sensitive information can be inferred. Therefore, a straightforward way to achieve anonymisation is to hide the real identity label of a graph. However, with enough structural knowledge about the graph, an attacker can still recover the node identities. To address this and to enforce k-anonymity, Minello et al. [29] proposed an algorithm based on the Szemerédi regularity lemma [20]. Given a graph, the proposed method starts by computing a k-regular partition of its nodes. The Szemerédi regularity lemma ensures that such a partition exists and that the edges between the sets of nodes behave quasi-randomly. With such a partitioning in hand, anonymisation is achieved in the graph by randomising the edges within each set, obtaining a graph that is structurally similar to the original one, but the nodes within each set are structurally indistinguishable. With this, the authors are able to create anonymous groups that are resilient to any type of structural attack (including attack to network topology) while minimising the structural information loss.

Editorial Comments

In [29], a scenario where the attacker knows both the original structure of the graph and the node identities has not been considered. However, the attacker may correlate the known identities from the original graph to the obfuscated one, especially in small obfuscated groups where the structural deviation from the original graph is minimal [36]. This scenario has been neglected in the paper.

References

[1] Adnan Anwar, Abdun Naser Mahmood, Zahir Tari, and Akhtar Kalam. 2022. Measurement-Driven Blind Topology Estimation for Sparse Data Injection Attack in Energy System. *Electric Power Systems Research* 202 (2022), 11 pages. https://doi.org/10.1016/j.epsr.2021.107593

[2] Björn Bebensee, Nagmat Nazarov, and Byoung-Tak Zhang. 2021. Leveraging Node Neighborhoods and Egograph Topology for Better Bot Detection in Social Graphs. *Social Network Analysis and Mining* 11, 1 (2021), 1–14. https://doi.org/10.1007/s13278-020-00713-z

[3] Christina Boididou, Katerina Andreadou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, Yiannis Kompatsiaris, et al. 2015. Verifying Multimedia Use at Mediaeval 2015. In *Working Notes Proceedings of the MediaEval 2015 Workshop*. CEUR Workshop Proceedings (CEUR-WS.org), 3 pages. http://ceur-ws.org/Vol-1436/Paper4.pdf

[4] Osman Boyaci, Mohammad Rasoul Narimani, Katherine R. Davis, Muhammad Ismail, Thomas J. Overbye, and Erchin Serpedin. 2022. Joint Detection and Localization of Stealth False Data Injection Attacks in Smart Grids Using Graph Neural Networks. *IEEE Transactions on Smart Grid* 13, 1 (2022), 807–819. https://doi.org/10.1109/TSG.2021.3117977

[5] Osman Boyaci, Amarachi Umunnakwe, Abhijeet Sahu, Mohammad Rasoul Narimani, Muhammad Ismail, Katherine R. Davis, and Erchin Serpedin. 2021. Graph Neural Networks Based Detection of Stealth False Data Injection Attacks in Smart Grids. *IEEE Systems Journal* (2021), 1–12. https://doi.org/10.1109/JSYST.2021.3109082

[6] Stefano Cresci, Fabrizio Lillo, Daniele Regoli, Serena Tardelli, and Maurizio Tesconi. 2018. $FAKE: Evidence of Spam and Bot Activity in Stock Microblogs on Twitter. In *Proceedings of the 12nd International AAAI Conference on Web and Social Media (ICWSM'18)*. AAAI Press, 580–583. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17871/17055

[7] Enyan Dai, Yiwei Sun, and Suhang Wang. 2020. Ginger Cannot Cure Cancer: Battling Fake Health News with a Comprehensive Data Repository. In *Proceedings of the 14th International AAAI Conference on Web and Social Media (ICWSM'20)*, Vol. 14. AAAI Press, 853–862. https://ojs.aaai.org/index.php/ICWSM/article/view/7350

[8] Prasannakumaran Dhanasekaran, Harish Srinivasan, S. Sowmiya Sree, I. Sri Gayathri Devi, Saikrishnan Sankar, and Vineeth Vijayaraghavan. 2021. SOMPS-Net: Attention Based Social Graph Framework for Early Detection of Fake Health News. In *Proceedings of the 19th Australasian Conference on Data Mining (AusDM'21)*. Springer, 165–179. https://doi.org/10.1007/978-981-16-8531-6_12

[9] Yousef Ebrahimi and Mohamed Younis. 2021. Traffic Analysis Through Spatial and Temporal Correlation: Threat and Countermeasure. *IEEE Access* 9 (2021), 54126–54151. https://doi.org/10.1109/ACCESS.2021.3070841

[10] Chungang Gao (高春刚), Yongjie Wang (王永杰), and Xinli Xiong (熊鑫立). 2021. MTD Enhanced Cyber Deception Defense System / MTD 增强的网络欺骗防御系统. *Computer Engineering and Applications / 计算机工程与应用* (2021). https://kns.cnki.net/kcms/detail/11.2127.TP.20210806.0830.004.html

[11] Marianela García Lozano, Joel Brynielsson, Ulrik Franke, Magnus Rosell, Edward Tjörnhammar, Stefan Varga, and Vladimir Vlassov. 2020. Veracity Assessment of Online Data. *Decision Support Systems* 129 (2020), 14 pages. https://doi.org/10.1016/j.dss.2019.113132

[12] Chunpeng Ge, Lu Zhou, Gerhard P. Hancke, and Chunhua Su. 2021. A Provenance-Aware Distributed Trust Model for Resilient Unmanned Aerial Vehicle Networks. *IEEE Internet of Things Journal* 8, 16 (2021), 12481–12489. https://doi.org/10.1109/JIOT.2020.3014947

[13] 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 (TOIT)* 22, 1, Article 24 (2021), 31 pages. https://doi.org/10.1145/3467021

[14] Hansen He (何韩森) and Guozi Sun (孙国梓). 2020. Fake News Content Detection Model based on Feature Aggregation / 基于特征聚合的假新闻内容检测模型. *Journal of Computer Applications / 计算机应用* 40, 8 (2020), 2189–2193. https://doi.org/10.11772/j.issn.1001-9081.2019122114

[15] Tao Hou, Zhe Qu, Tao Wang, Zhuo Lu, and Yao Liu. 2020. ProTO: Proactive Topology Obfuscation Against Adversarial Network Topology Inference. In *IEEE Conference on Computer Communications (IEEE INFOCOM'20)*. IEEE, 1598–1607. https://doi.org/10.1109/INFOCOM41043.2020.9155255

[16] Linmei Hu, Tianchi Yang, Luhao Zhang, Wanjun Zhong, Duyu Tang, Chuan Shi, Nan Duan, and Ming Zhou. 2021. Compare to The Knowledge: Graph Neural Fake News Detection with External Knowledge. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, 754–763. https://doi.org/10.18653/v1/2021.acl-long.62

[17] Mohsen Jorjani, Hossein Seifi, and Ali Yazdian Varjani. 2020. A Graph Theory-Based Approach to Detect False Data Injection Attacks in Power System AC State Estimation. *IEEE Transactions on Industrial Informatics* 17, 4 (2020), 2465–2475. https://doi.org/10.1109/TII.2020.2999571

[18] Zhezhou Kang, Yanan Cao, Yanmin Shang, Tao Liang, Hengzhu Tang, and Lingling Tong. 2021. Fake News Detection with Heterogenous Deep Graph Convolutional Network. In *Proceedings of the 25th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'21)*. Springer, 408–420. https://doi.org/10.1007/978-3-030-75762-5_33

[19] Simon Knight, Hung X. Nguyen, Nickolas Falkner, Rhys Bowden, and Matthew Roughan. 2011. The Internet Topology Zoo. *IEEE Journal on Selected Areas in Communications* 29, 9 (2011), 1765–1775. https://doi.org/10.1109/JSAC.2011.111002

[20] János Komlós, Ali Shokoufandeh, Miklós Simonovits, and Endre Szemerédi. 2000. The Regularity Lemma and Its Applications in Graph Theory. *Summer School on Theoretical Aspects of Computer Science* 2292 (2000), 84–112. https://doi.org/10.1007/3-540-45878-6_3

[21] Hao Li, Min Gao, Fengtao Zhou, Yueyang Wang, Qilin Fan, and Linda Yang. 2021. Fusing Hypergraph Spectral Features for Shilling Attack Detection. *Journal of Information Security and Applications* 63 (2021), 10 pages. https://doi.org/10.1016/j.jisa.2021.103051

[22] Xiao Liang, Zheng Yang, Binghui Wang, Shaofeng Hu, Zijie Yang, Dong Yuan, Neil Zhenqiang Gong, Qi Li, and Fang He. 2021. Unveiling Fake Accounts at the Time of Registration: An Unsupervised Approach. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD'21)*. ACM, 3240–3250. https://doi.org/10.1145/3447548.3467094

[23] Bo Liu and Hongyu Wu. 2021. Optimal Planning and Operation of Hidden Moving Target Defense for Maximal Detection Effectiveness. *IEEE Transactions on Smart Grid* 12, 5 (2021), 4447–4459. https://doi.org/10.1109/TSG.2021.3076824

[24] Yaqun Liu, Changyou Xing, Guomin Zhang, Lihua Song, and Hongxiu Lin. 2022. AntiTomo: Network Topology Obfuscation against Adversarial Tomography-Based Topology Inference. *Computers & Security* 113 (2022), 10 pages. https://doi.org/10.1016/j.cose.2021.102570

[25] Yaqun Liu, Jinlong Zhao, Guomin Zhang, and Changyou Xing. 2021. NetObfu: A Lightweight and Efficient Network Topology Obfuscation Defense Scheme. *Computers & Security* 110 (2021), 14 pages. https://doi.org/10.1016/j.cose.2021.102447

[26] Yaqun Liu (刘亚群), Changyou Xing (邢长友), Yazhuo Gao (高雅卓), and Guomin Zhang (张国敏). 2021. TopoObfu:A Network Topology Obfuscation Mechanism to Defense Network Reconnaissance / TopoObfu: 一种对抗网络侦察的网络拓扑混淆机制. *Computer Science / 计算机科学* 48, 10 (2021), 278–285. https://doi.org/10.11896/jsjkx.210400296

[27] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting Rumors from Microblogs with Recurrent Neural Networks. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI'16)*. AAAI Press, 3818–3824. https://www.ijcai.org/Proceedings/16/Papers/537.pdf

[28] Bundit Manaskasemsak, Jirateep Tantisuwankul, and Arnon Rungsawang. 2021. Fake Review and Reviewer Detection through Behavioral Graph Partitioning Integrating Deep Neural Network. *Neural Computing and Applications* (2021), 1–14. https://doi.org/10.1007/s00521-021-05948-1

[29] Giorgia Minello, Luca Rossi, and Andrea Torsello. 2021. k-Anonymity on Graphs Using the Szemerédi Regularity Lemma. *IEEE Transactions on Network Science and Engineering* 8, 2 (2021), 1283–1292. https://doi.org/10.1109/TNSE.2020.3020329

[30] Kai Nakamura, Sharon Levy, and William Yang Wang. 2020. Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC'20)*. European Language Resources Association, 6149–6157. https://aclanthology.org/2020.lrec-1.755

[31] Ouri Poupko, Gal Shahaf, Ehud Shapiro, and Nimrod Talmon. 2021. Building a Sybil-Resilient Digital Community Utilizing Trust-Graph Connectivity. *IEEE/ACM Transactions on Networking* 29, 5 (2021), 2215–2227. https://doi.org/10.1109/TNET.2021.3084303

[32] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP'17)*. Association for Computational Linguistics, 2931–2937. https://doi.org/10.18653/v1/D17-1317

[33] Punit Rathore, Jayesh Soni, Nagarajan Prabakar, Marimuthu Palaniswami, and Paolo Santi. 2021. Identifying Groups of Fake Reviewers Using a Semisupervised Approach. *IEEE Transactions on Computational Social Systems* 8, 6 (2021), 1369–1378. https://doi.org/10.1109/TCSS.2021.3085406

[34] Victoria L. Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. In *Proceedings of the 2nd Workshop on Computational Approaches to Deception Detection*. Association for Computational Linguistics, 7–17. https://aclanthology.org/W16-0802.pdf

[35] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data* 8, 3 (2020), 171–188. https://doi.org/10.1089/big.2020.0062

[36] Vicenç Torra and Julián Salas. 2019. Graph Perturbation as Noise Graph Addition: A New Perspective for Graph Anonymization. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology*. Springer, 121–137. https://doi.org/10.1007/978-3-030-31500-9_8

[37] Li Wang, Peipei Li, Kai Xiong, Jiashu Zhao, and Rui Lin. 2021. Modeling Heterogeneous Graph Network on Fraud Detection: A Community-Based Framework with Attention Mechanism. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM'21)*. ACM, 1959–1968. https://doi.org/10.1145/3459637.3482277

[38] William Yang Wang. 2017. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, 422–426. https://doi.org/10.18653/v1/P17-2067

[39] Zhenhua Wang, Haibo He, Zhiqiang Wan, and Yan Sun. 2021. Coordinated Topology Attacks in Smart Grid Using Deep Reinforcement Learning. *IEEE Transactions on Industrial Informatics* 17, 2 (2021), 1407–1415. https://doi.org/10.1109/TII.2020.2994977

[40] Shengfeng Wang (王胜锋), Zhou Ding (丁洲), Jingsong Wu (吴劲松), and Aibing Qiu (邱爱兵). 2021. False Data Injection Attack Scheme of Electricity Market based on Topology Tampering / 基于拓扑篡改的电力市场虚假数据注入攻击方案. *Electric Power Automation Equipment / 电力自动化设备* 41, 11 (2021), 147–152. https://doi.org/10.16081/j.epae.202106004

[41] Fan Wu, Min Gao, Junliang Yu, Zongwei Wang, Kecheng Liu, and Xu Wang. 2021. Ready for Emerging Threats to Recommender Systems? A Graph Convolution-Based Generative Shilling Attack. *Information Sciences* 578 (2021), 683–701. https://doi.org/10.1016/j.ins.2021.07.041

[42] Chang Xu, Jie Zhang, Kuiyu Chang, and Chong Long. 2013. Uncovering Collusive Spammers in Chinese Review Websites. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management (CIKM'13)*. ACM, 979–988. https://doi.org/10.1145/2505515.2505700

[43] Zhihai Yang, Qindong Sun, Yaling Zhang, and Wei Wang. 2021. Identification of Malicious Injection Attacks in Dense Rating and Co-Visitation Behaviors. *IEEE Transactions on Information Forensics and Security* 16 (2021), 537–552. https://doi.org/10.1109/TIFS.2020.3016827

[44] Kaiqiang Yu and Cheng Long. 2021. Graph Mining Meets Fake News Detection. In *Data Science for Fake News: Surveys and Perspectives*. Springer, 169–189. https://doi.org/10.1007/978-3-030-62696-9_8

[45] Hua Yuan, Jie Zheng, Qiongwei Ye, Yu Qian, and Yan Zhang. 2021. Improving Fake News Detection with Domain-Adversarial and Graph-Attention Neural Network. *Decision Support Systems* 151 (2021), 11 pages. https://doi.org/10.1016/j.dss.2021.113633