DARTMOUTH

Institute for Security
Technology and Society
at Dartmouth

Fake Document Generation against IP Theft

In today's digitized world, intellectual property theft has become an increasingly prominent issue. Organizations and businesses must face the reality that hackers will eventually break into their systems and attempt to retrieve information. In order to address this problem, Dartmouth has led a number of projects aiming to improve the viability of generating fake documents to impose additional costs on malicious actors after exfiltration or increasing risk of discovery during exfiltration.

FORGE

The Fake Online Repository Generation Engine (FORGE) is a project which generates fake technical documents by replacing central textual concepts with related concepts based on an ontology. Specifically, FORGE constructs a multi-layer graph from key concepts/terminology in the document to map out relationships between terms in the document and the domain to which the document belongs to (such as Biology, Chemistry). The "meta-centrality" of the nodes is then computed, which is a novel idea that expands single layer graph centrality measures to generalize similarities between multiple layers. From this multi-layer graph, a fake document is produced by replacing concepts using an ontology to find suitable replacement concepts. The fake documents generated were then tested alongside the original document by human subjects. The results show that the fake documents generated are highly effective in deceiving human subjects while maintaining believability of the fake documents.

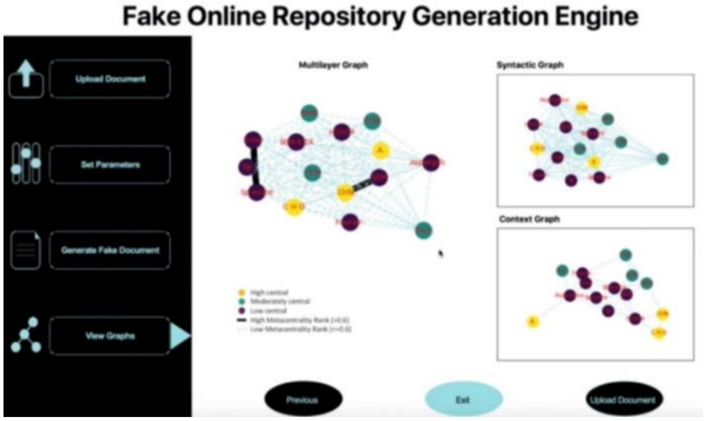
The credibility of FORGE is 98.20%… The high deception factor achieved by FORGE is not only due to the number of fake documents generated by FORGE but also the quality of the generated documents which appear believable to the attackers.

Probabilistic Logic Graphs

FORGE only replaces content in text, which might provide malicious actors with a way to identify documents based on unchanged text in diagrams and tables. This problem is addressed in another project using Probabilistic Logic Graphs (PLG). PLGs are directed graphs with semantic information and probabilistic annotations, which are used in this project to represent text in graphs and diagrams. After the fake PLGs have been generated from the original document, they are then mapped onto fake documents. After testing these generated documents against human subjects, the results show the fake documents generated by the PLGs successfully deceive human subjects a majority of the time.

FakeTables/ITS-GAN

While PLGs allow for replacing text within diagrams and tables, FakeTables were developed to create synthetic tables from data while preserving functional dependencies within the table. FakeTables use a novel framework called ITS-GAN, which adapts traditional generative adversarial networks to generate synthetic tables. ITS-GAN also trains autoencoders to model functional dependencies for both the generator and the discriminator. ITS-GAN was tested using U.S. Census and Air Carrier Statistics data: the results showed that ITS-GAN greatly improves current methods of tabular data augmentation. Though ITS-GAN was not originally developed for fake document generation, it is clear that it is easily able to create deceptive tables from small subsets of a full table that should be hidden from malicious actors.



Demo FORGE GUI showing graphs associated with a document

https://www.cs.dartmouth.edu/~dsail/projects/forge.html

FORGE Details

FORGE begins by finding key terminology, formulas, and other concepts often found in technical documents using neuro-linguistic programming. Using these concepts, FORGE constructs a multi-layer graph to map out relationships between concepts inside the document as well as across the domain related to the document. Thus, the multi-layer graph has a "document" and "domain" layer connected with each other.

The centrality of the nodes within this multi-layer graph is computed through the "meta-centrality." The "meta-centrality" is a novel idea developed for this project that builds upon classical centrality measures for single-layer graphs (such as Degree Centrality or Closeness Centrality) to compute generalized similarities between concepts within a document and domains.

After "meta-centrality" ranks are assigned to the concepts, the process of fake document generation begins. This requires one to answer three questions:

(i) which candidate concepts can replace a given concept?
(ii) which concepts should be replaced?
(iii) how many concepts should we replace?

The first question is answered by using a domain-specific ontology to find suitable replacement concepts, such as "lithium" replacing "sodium."

The latter questions are answered by solving three constrained optimization problems for concepts based on their "meta-centrality" score (highly-ranked, medium-ranked, and low-ranked).

The problem of fake document generation is then defined as: generating n fake documents from d by replacing a set of C concepts in such a way that the sum of the meta-centrality ranking of the selected concept will be minimum and the substitution cost will remain bounded within a certain budget B_n per fake document.

This problem is found to be intractable, so the fake document generation problem is mapped to a knapsack problem, where the value of a concept corresponds to the "meta-centrality" ranking and the weight of a concept corresponds to the budget.

Finally, after fake document generation, FORGE incorporates feedback from security officers, who accept or reject replacement concepts and their corresponding alternatives. If a security officer rejects all candidate replacement concepts or suggests an alternative concept, then FORGE is rerun using new user input.

FORGE was tested using 250 documents from the domains of agricultural chemistry and computer science against Engineering Masters and Computer Science Masters students. In total, 50 sets of documents were created for each domain, with each set including one original document and a number of generated fakes. The participants were asked to rank the top 3 documents that they felt confident to be the original from each set.

The results from this project show that FORGE is highly effective in deceiving human subjects attempting to identify an original document among fakes while maintaining believability of the fake documents.

Additional Information

References

1. T. Chakraborty, S. Jajodia, J. Katz, A. Picariello, G. Sperli, and V.S. Subrahmanian. FORGE: Fake Online Repository Generation Engine, accepted for publication, *IEEE Transactions on Dependable & Secure Computing* (accepted Jan 2019).
2. H. Chen, S. Jajodia, J. Liu, N. Park, V. Sokolov, V.S. Subrahmanian. FakeTables: Using GANs to Generate Functional Dependency Preserving Tables with Bounded Real Data, *Proc. 2019 Intl. Joint Conference on Artificial Intelligence (IJCAI 2019)*, Aug 2019, Macao.

Video

https://www.cs.dartmouth.edu/~dsail/projects/forge/forge-video.mp4

Presentation

https://www.cs.dartmouth.edu/~dsail/projects/forge/forge-slides.pdf

PARTICIPANTS

Lead: V.S. Subrahmanian

Tanmoy Chakraborty, Haipeng Chen, Qian Han, Sushil Jajodia, Jonathan Katz, Jing Liu, Cristian Molinaro, Noseong Park, Antonio Picariello, Vadim Sokolov, Giancarlo Sperlì, Yanhai Xiong.

Dartmouth
Security and Artificial
Intelligence Laboratory