Paper Title: Detecting Malicious URL

Paper Link: 10.1109/ICCIT-144147971.2020.9213792

1. Summary

1.1. Motivation:

- The increasing ubiquity of web-based services
- Cybersecurity threat posed by malicious URLs
- Hypothesis: Machine learning can effectively detect malicious URLs

1.2. Contribution:

- Application of four classification algorithms (Random Forest, KNN, J48, BayesNet)
- Experimentation on a public dataset with 20 features and 1781 records
- Identification of Random Forest as the top-performing algorithm (96% accuracy)

1.3. Methodology:

- Two test phases: all features vs. selected features
- Feature selection using CfsSubsetEval
- Evaluation metrics: accuracy, recall, precision
- Cross-validation with 10 folds

1.4. Conclusion:

- Random Forest outperforms other algorithms
- Feature selection improves performance and reduces execution time
- Importance of detecting malicious URLs for web security

2. Limitations

2.1 First Limitation

- Size of the dataset (1781 records) may limit generalizability
- Consideration of additional features for a more comprehensive model

2.2 Second Limitation

- Focus primarily on malicious URLs, not addressing other web application attacks
- Future work could include broader threat detection mechanisms

3. Synthesis

3.1 Potential Applications:

- Enhancing cybersecurity measures for web-based services
- Integrating machine learning into web security systems

3.2 Future Scopes:

- Expanding feature sets for more robust models
- Investigating broader web application attack detection
- Exploring diverse datasets for generalizable results