Chair of Network Architectures and Services
School of Computation, Information, and Technology
Technical University of Munich

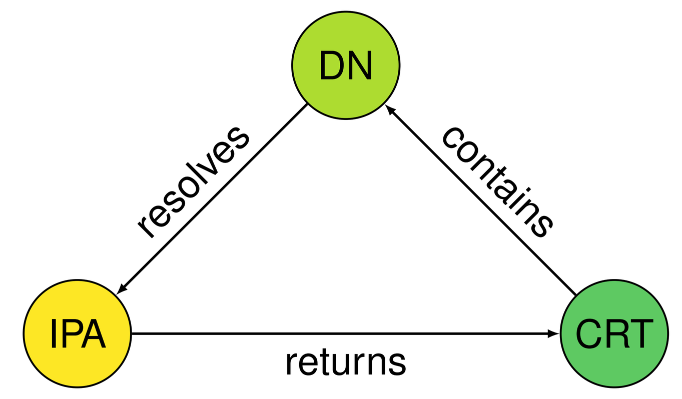
Graph-based Modeling and Analysis of the TLS Ecosystem

Motivation

- Blocklists often have a limited view of which resources are malicious, depending on how they amass their information
- Other researchers have proposed multiple methods to use the blocklists and Internet measurements to find more malicious actors that are not listed on the blocklists
- This work evaluates different approaches and compares their results and how well they perform on a larger dataset

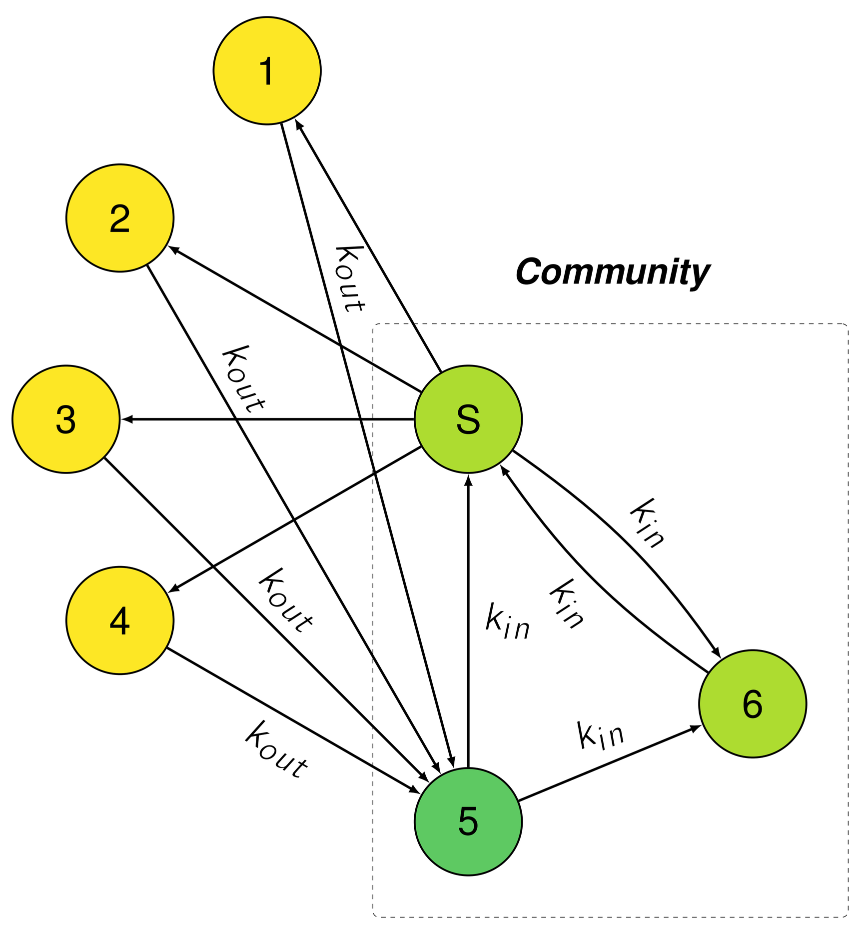
Database

- Created from DNS and TLS queries executed over a period of a month
- Blocklists designate malicious actors and suspicious ressources
- The blocklists are split into a training and evaluation set.



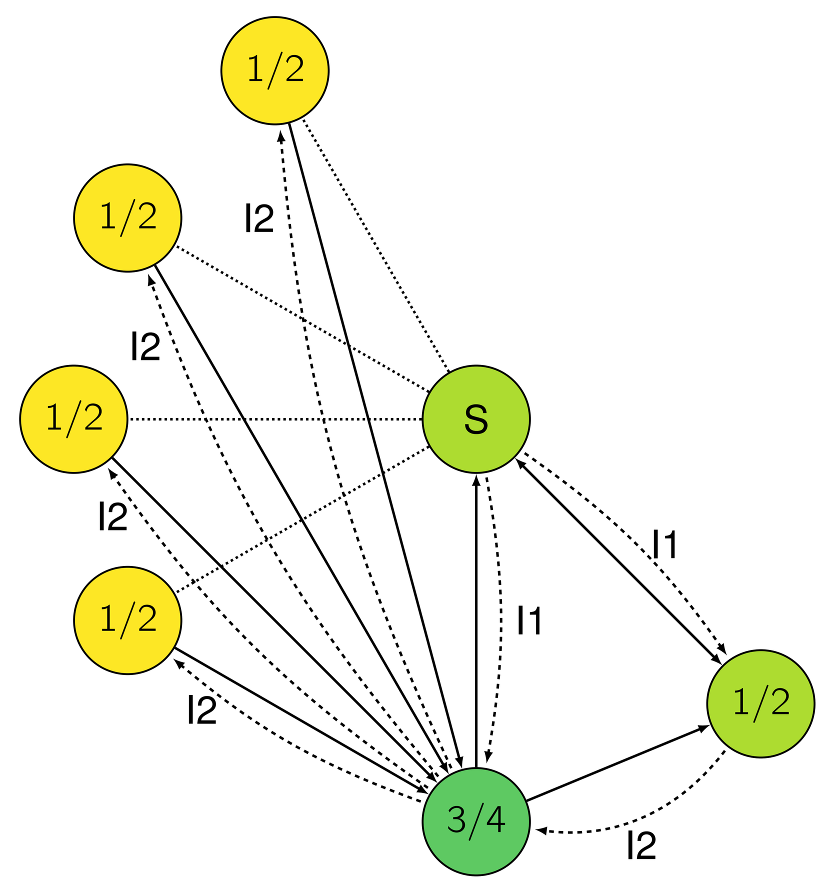
Local Community Detection

- Community Detection based on edges
- Different fitness functions [5, 4, 3]
- $f_M(C) = \frac{1}{|E|}(k_{in}^C - \frac{(2k_{in}^C + k_{out}^C)^2}{4|E|})$
- $f_\alpha(C) = \frac{2k_{in}^C + 1}{(2k_{in}^C + k_{out}^C)^\alpha}$



Probabilistic Threat Propagation

- Proposed by Carter et al. [1, 2]
- Propagate maliciousness score
- All blocked nodes have a fixed score of 1
- Each iteration, all nodes with a score push their score to their neighbors
- Algorithm stops, when scores converge or the change falls below a certain threshold



Results

Total Number

Blocklist	LCD Modularity	LCD $\alpha_{0.5}$	LCD α_1	PTP
Feodo	2,955	931	788	3,055
SSLBL	865	778	755	$2.57 \cdot 10^5$
StrongIPs	7,869	1,815	1,413	8,596
OpenPhish	$4.22 \cdot 10^5$	$1.55 \cdot 10^5$	$1.31 \cdot 10^5$	$4.49 \cdot 10^6$

Recall

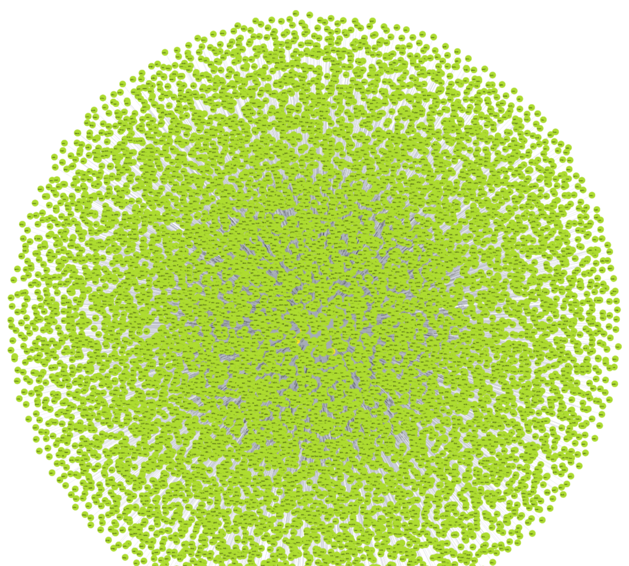
Blocklist	LCD Modularity	LCD $\alpha_{0.5}$	LCD α_1	PTP
Feodo	7.5%	7.5%	7.5%	7.5%
SSLBL	/	/	/	/
StrongIPs	14%	12%	11%	6.8%
OpenPhish	14%	5.7%	5.2%	19%

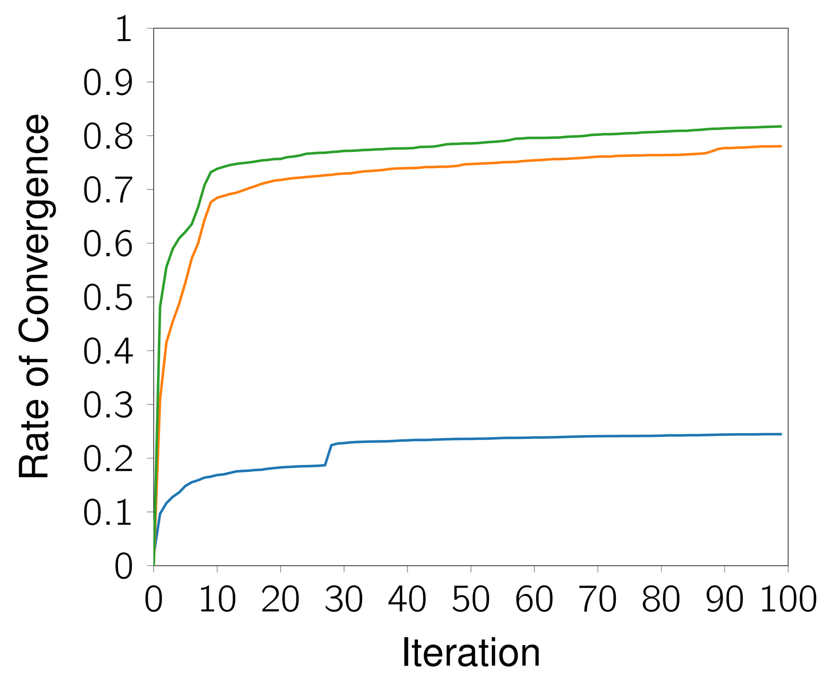
Precision

Blocklist	LCD Modularity	LCD $\alpha_{0.5}$	LCD α_1	PTP
Feodo	0.43%	0.63%	0.79%	0.43%
SSLBL	/	/	/	/
StrongIPs	0.44%	0.69%	0.86%	0.06%
OpenPhish	0.31%	0.33%	0.37%	0.04%

- Changes in the blocklists are used to evaluate the results
- High number of potential candidates
- Good recall rate
- Precision is low, because the blocklists are in comparison much smaller

Summary





- Performance is a important characteristic for the algorithms
- Local Community Detection has better recall and precision rates
- Probabilistic Threat Propagation has a much better performance
- Graph and blocklists characteristics have a large impact on the analysis
  - Highly connected graphs lead to a low convergence rate for LCD
  - Large structures lead to a large number of nodes with a score for PTP
- Outlook
  - Performance improvements
  - Building a blocklists accumulator with additional information is possible

[1] K. M. Carter, N. Idika, and W. W. Streilein. Probabilistic threat propagation for network security. *IEEE Transactions on Information Forensics and Security*, 9(9):1394–1405, Sept.
[2] K. M. Carter, N. Idika, and W. W. Streilein. Probabilistic threat propagation for malicious activity detection. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, May 2013.
[3] F. Havemann, M. Heinz, A. Struck, and J. Gläser. Identification of overlapping communities and their hierarchy by locally calculating community-changing resolution levels. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(01):P01023, Jan. 2011.
[4] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2), Feb. 2004.
[5] A. Zakrzewska and D. A. Bader. A dynamic algorithm for local community detection in graphs. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ASONAM '15. ACM, Aug. 2015.

Tim Betzer
betzer@net.in.tum.de