Capturing the Internet presence of Organizations with the help of a TLS Graph

The Internet undergoes an increasing adoption of the TLS to encrypt and authenticate traffic. Essential for the authentication of servers is the interaction of the DNS and the x509 certificate infrastructure. With active scans it is possible to cryptographically verify whether a server on a specific IP address has the proper certificate for a requested domain name. This creates a relation among TLS servers that can be used to better understand server deployments. Previous work has established a pipeline that scans and parses the data into the TLS Graph (see the schema in the figure) which combines information from the DNS, TLS, and BGP.

Manual inspection and first analyses show a lot of potential to reveal the Internet presence of organizations. Such an organization could be a company or a single administrator. Security systems can benefit from such data to improve their assessment of security risks (e.g., is the newly observed update server of an IOT device similar to the previously observed ones). However, in the Internet strange things happen and these relations are not as clearly defined as one would hope. For example, what happens if an IP address seems to be shared between multiple organizations? This can be caused by cloud providers or CDNs.

Community detection algorithms can be used to automatically find all nodes related to such organizations and cross-validated with indicators in certificates, domain names, and public domain-portfolio lists. Results could be the a domain-portfolio of different companies.

- Research on public available ground-truth to validate clusters
- Research on suitable graph algorithms
- Implement and evaluate different community detection approaches
- Python 3, networkx, scikit-learn
- Community detection algorithms from shallow learning (e.g., LPA)

Markus Sosnowski   sosnowski@net.in.tum.de
Patrick Sattler         sattler@net.in.tum.de