HIMALAYAS

PIs: SRI: Shalini Ghosh, Vinod Yegneswaran; **UMN:** Arindam Banerjee; **TAMU:** Guofei Gu

Challenge

- Analyze large-scale data of different types (e.g., onion sites from Darkweb, web servers, DNS sequences) to detect hidden malicious activities (e.g., malicious domains / servers).

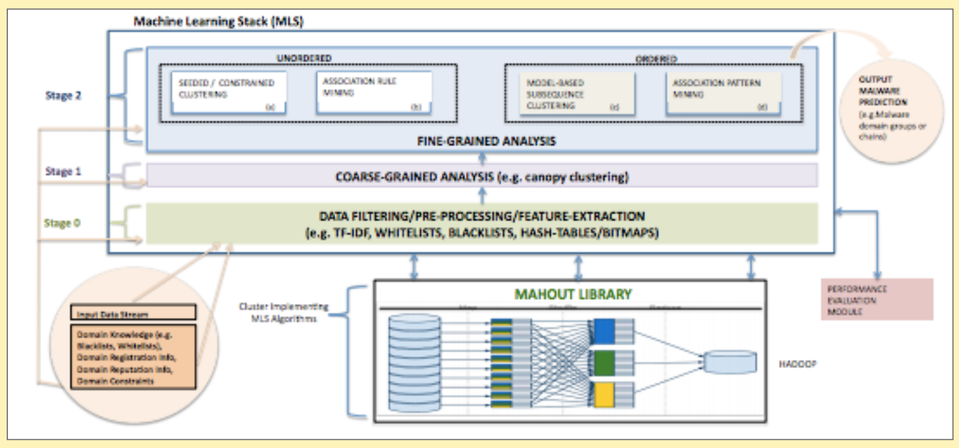

Fig. 1

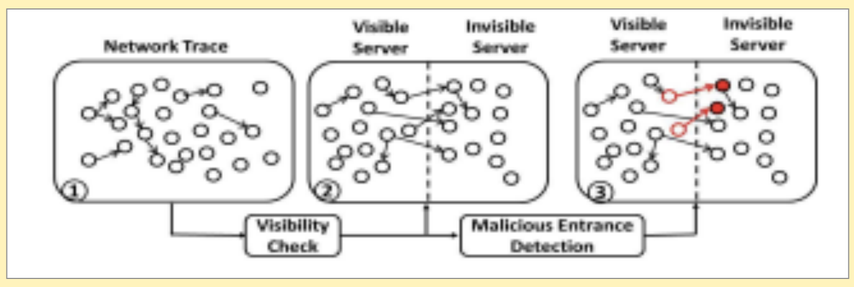

Fig. 2

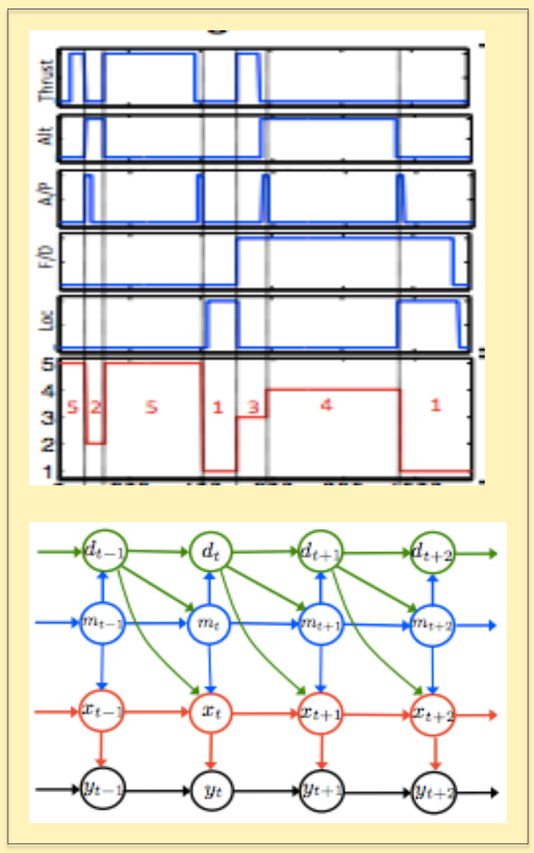

Fig. 3

Approaches

- Develop machine learning models that can:
  - incorporate prior knowledge,
  - operate with minimal supervision,
  - give interpretable results,
  - handle large-scale data.

TAMU contribution [Figure 2]

Conducted a large-scale measurement study of malicious web infrastructure and developed new tools, which can provide ground truth for training relevant ML models.

"Hunting for Invisibility: Characterizing and Detecting Malicious Web Infrastructures through Server Visibility Analysis."Jialong Zhang, Xin Hu, Jiyong Jang, Ting Wang, Guofei Gu, Marc Stoecklin. In INFOCOM 2016.

SRI contribution [Figure 1]

Developed a multi-stage machine learning stack for thematic labeling of onions, which can detect onions associated with different malicious activities on the DarkWeb.

“ATOL: A Framework for Automated Analysis and Categorization of the Darkweb Ecosystem”, Shalini Ghosh, Phillip Porras, Vinod Yegneswaran, Ken Nitz, Ariyam Das. In AICS Workshop, AAAI 2017.

UMN contribution [Figure 3]

Developed novel ML (vector auto-regressive) models for time-series analysis – they can efficiently model temporal and spatial dependence in DNS sequences.

“Estimating Structured Vector Autoregressive Model”, I. Melnyk and A. Banerjee. In ICML 2016.

Impact

Scientific Impact

- Output of our ML analysis will be streamed as a data channel in the Secure Info Exchange, to be used by other INFOSEC researchers.
- Software tools developed will be released on github, facilitating acceleration of research in this area.

Broader Impact

- Improve the security of computing infrastructures by accelerating the identification and take downs of malicious actors.
- The tools built as part of this project can be applied to other relevant domains, e.g., financial analysis.

Interested in meeting the PIs? Attach post-it note below!