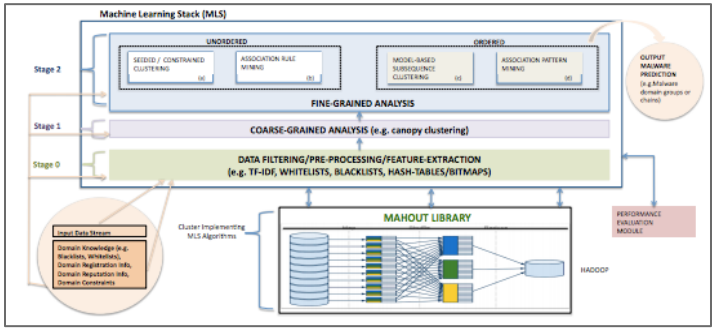
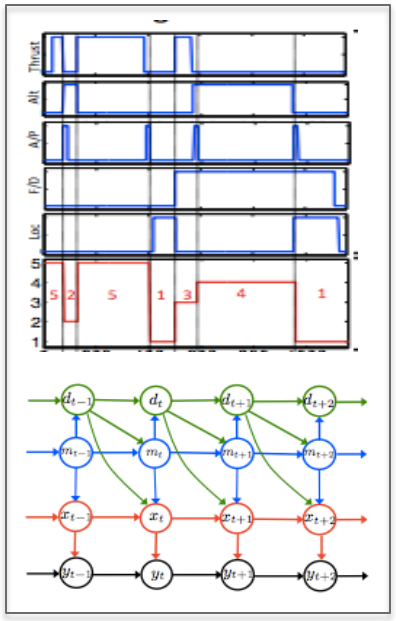
HIMALAYAS

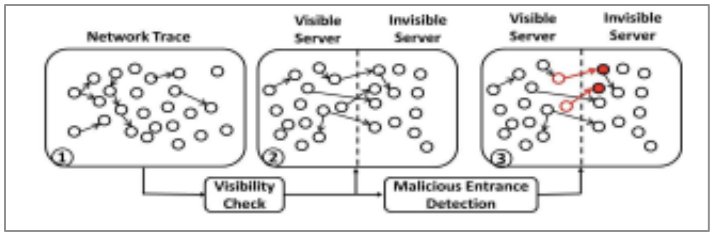
SRI



UMN



TAMU



Challenges:

- 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 or servers).
- Develop models that can incorporate prior knowledge, operate with minimal supervision, and give interpretable results.

Milestones:

- **SRI:** Developed a multi-stage machine learning stack for thematic labeling of onions.
- **TAMU:** Conducted a large-scale measurement study of malicious web infrastructure and developed new detection tools.
- **UMN:** Developed novel ML models for time-series analysis, applicable to DNS sequences.

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.
- Most of the software tools developed will be released on github, facilitating the 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 potentially applied to other domains, e.g., financial analysis.

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