Mitigating Cyber Attacks Using Domain Adaptation

Latifur Khan, Bhavani Thuraisingham, Zhiqiang Lin

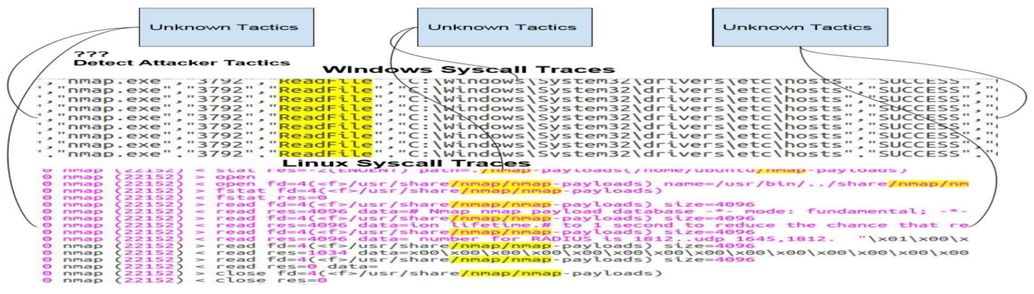
The University of Texas At Dallas

Motivation:

- **Increase in sophisticated attacks by nation states**
- **Limited training data for APT attack detection across multiple domains.**
- **Detect illegal goods sold on the dark web using data from multiple forums.**
- **Create a domain adaptation algorithm, multistream domain adaptation(MSDA) that can leverage data from multiple domains.**

Challenges

- Heterogeneous feature space: the dimension and distribution of data from source and target domains are different;
- Mismatch in feature space such as system calls across two domains e.g Linux vs Windows



Scientific Impact:

- Dataset can be leveraged to evaluate other cyber detection approaches.
- Increase in adoption of machine learning based approach over static rules for APT detection
- Provide better understanding of APT attacks from low level data traces such as system calls and packet data.
- Our MSDA algorithm can be applied to problems in other domains.

Solution:

Find a latent feature space where two domains are adapted and their original structure is preserved.

The objective function is as follows:

$$O = \min_{L_s, L_t} \ell(B_s, L_s) + \ell(B_t, L_t) + \beta \mathbf{D}(L_s, L_t)$$

Where B_s B_t are data instances from original feature space, and L_s L_t are data instances from the projected latent feature space

Multistream classification for cyber threat data with heterogeneous feature spaceYF Li, Y Gao, G Ayoade, H Tao, L Khan, B Thuraisingham

The World Wide Web Conference, 2992-2998

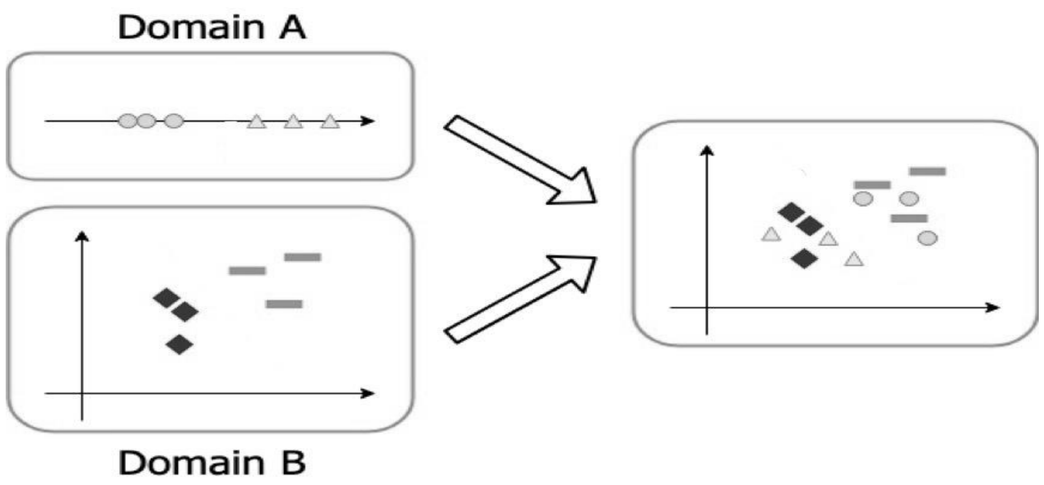


Fig 1: Overall idea of MSDA

Broader Impact on Society:

- Creates a workforce that is able to apply machine learning based principles to cyber defense.
- Continue to build tools that will enable further development in APT detection.

Broader Impact on Education and outreach

- Create a framework that can be leveraged upon by researchers in cyber security.
- Courses from CS, machine learning and cyber security may adopt the content of the project for designing class projects

Broader Impact (quantify potential impact)

- Dataset from this project can be used by other researchers to extend their research.
- Domain adaptation method developed can be leveraged in other domains with similar challenges.