CERIAS

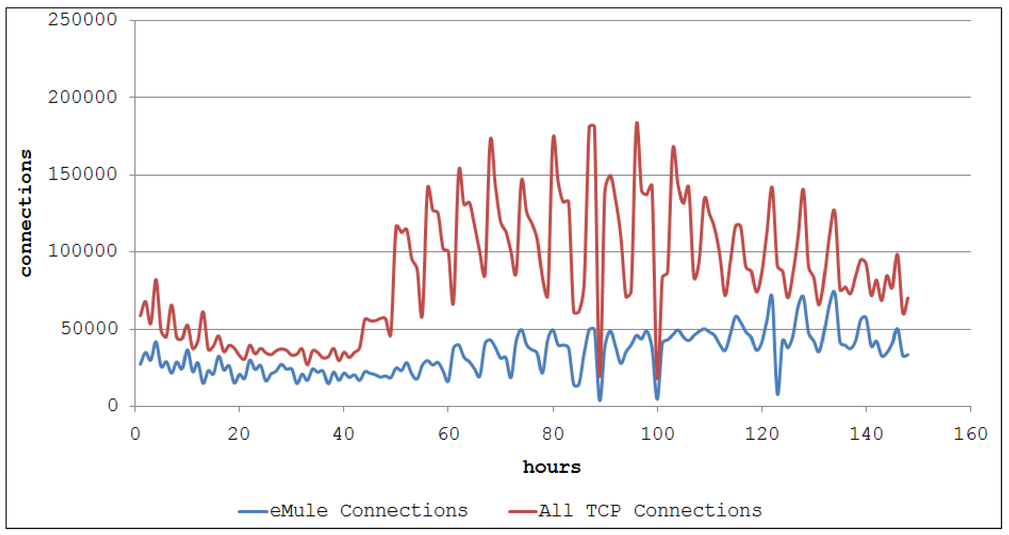
the center for education and research in information assurance and security

Behavior-Based Characterization of Peer-to-Peer (P2P) Traffic

Ruben Torres, Mohammad Hajjat, and Sanjay Rao

P2P traffic is **dominant**

Number of eMule connections passing through an ISP in 150 hours



But, do we really **understand** it?

P2P traffic has similar behavior to worms

- Contacting many nodes
- High failure ratio
- *Content prevalence*: common substrings among many packets

P2P classification methods provide **no insight** into **intrinsic characteristics** of P2P traffic

- Why P2P systems have high *failure ratio*?
- Why are some metrics good to classify P2P traffic and others are not?

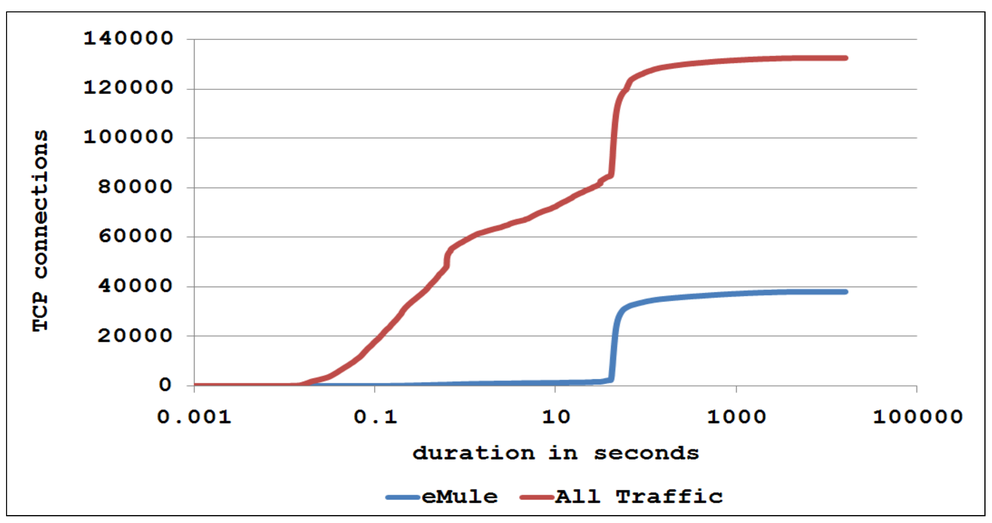
Goal: Provide an accurate P2P **traffic characterization** based on **intrinsic** understanding of P2P clients behavior

I. Selection of intuitive metrics to characterize P2P clients → Failure Ratio, Average connections per destination IP, Average connections duration, number of distinct destinations, etcetera.

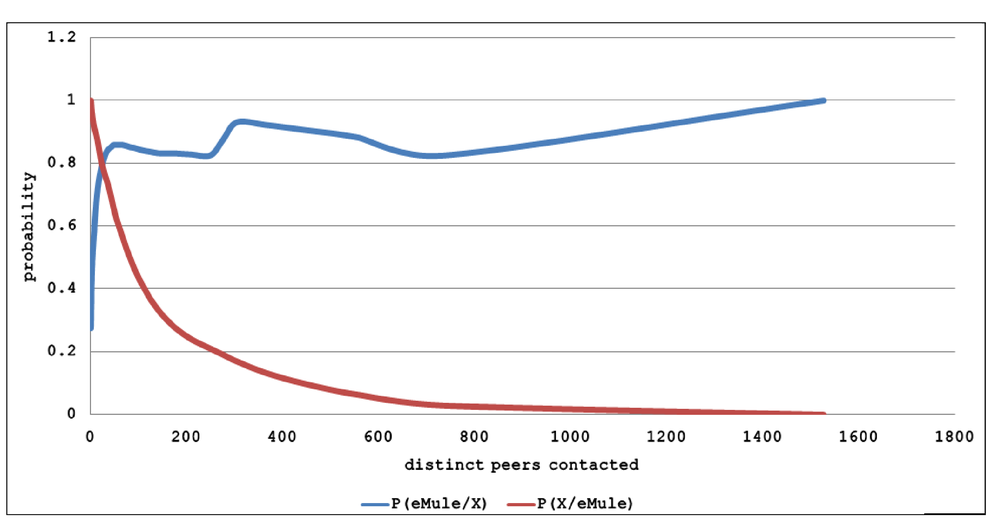
II. Understanding the distribution of metrics

III. Using simple probabilities to characterize P2P nodes behavior

CDF of duration of TCP connections: eMule connections last longer than other applications



Pr(Host runs eMule | number of distinct destinations) is 0.95 if number of distinct destinations is more than 400 in 5 min



PURDUE UNIVERSITY

CERIAS

Discovery Park e-Enterprise Center