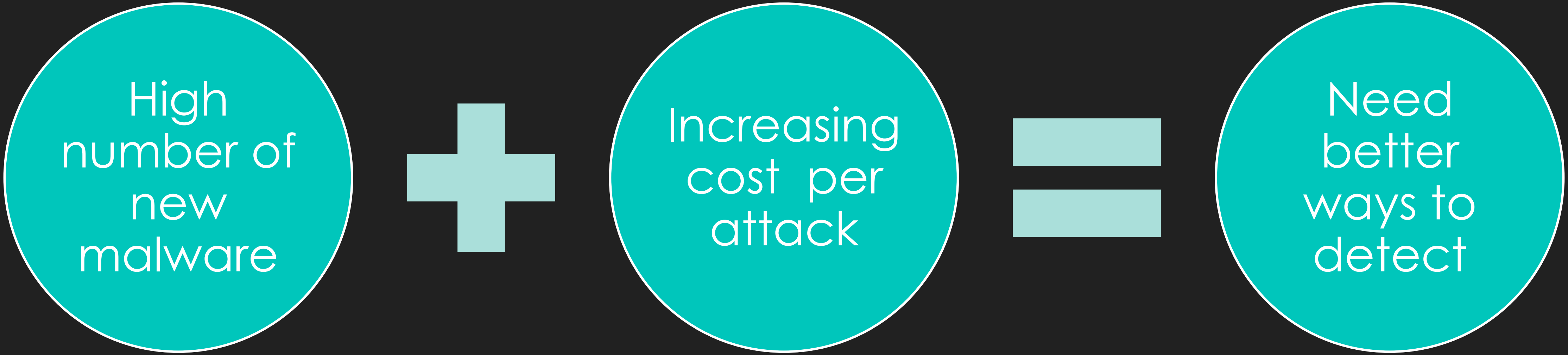
REA605 Presentation

Group SSL:

Gaston Carvallo

Loyd Rafols

Malware detection through machine learning



High number of new malware

Increasing cost per attack

Need better ways to detect

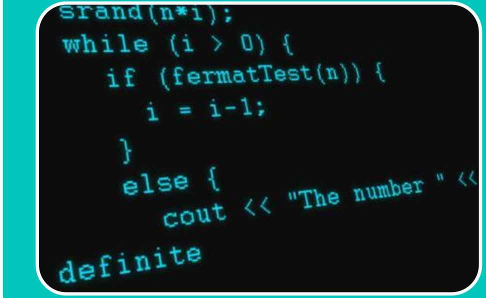
Agenda

Findings

Limitations

Proposal

Malware Detection



```
srand(n*i);
while (i > 0) {
   if (fermatTest(n)) {
      i = i-1;
   }
   else {
      cout << "The number " <<
definite
```

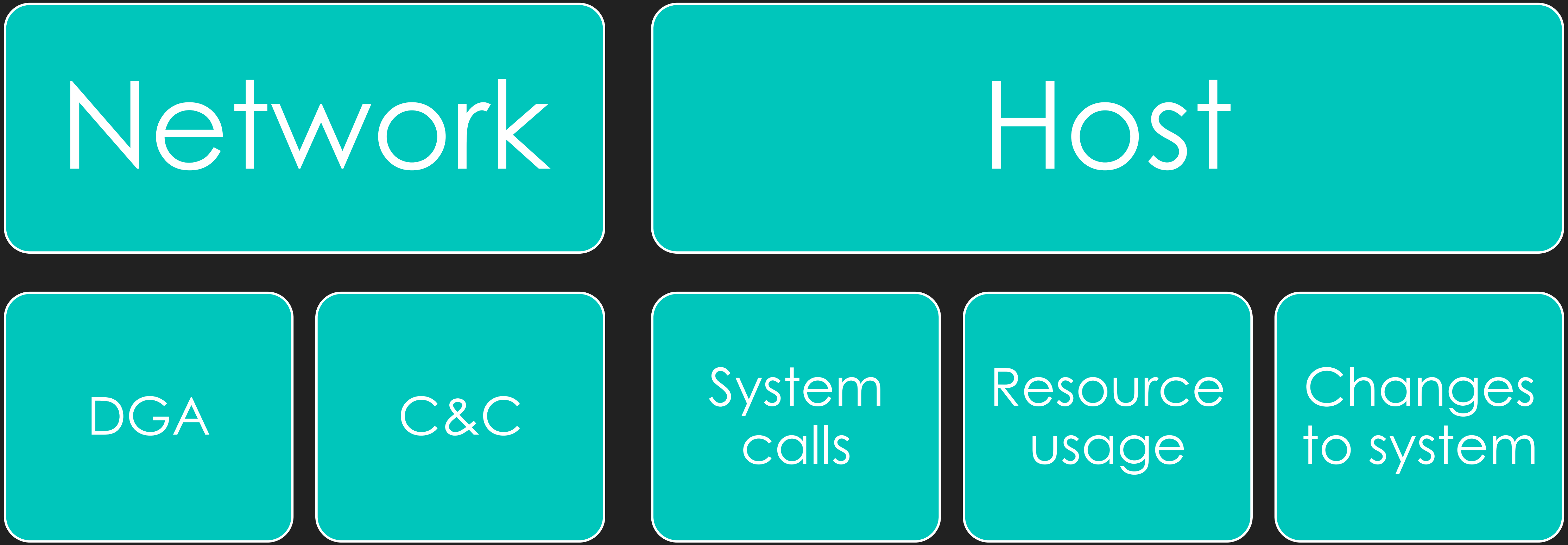
Static Analysis

- Fast and efficient for known malware
- Easier to evade

Dynamic Analysis

- Higher overhead
- Harder to evade

Dynamic Analysis



Network

Host

DGA

C&C

System calls

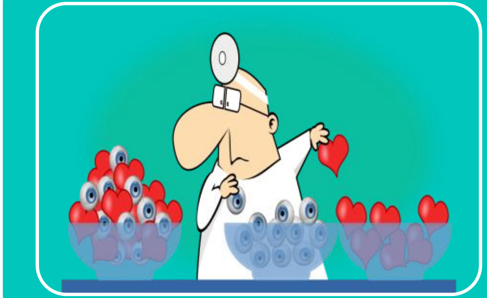
Resource usage

Changes to system

Machine Learning

Feature Selection

- Expertise
- Heuristics



Classifiers

- Supervised
- Semi-Supervised

Gaps and Limitations

- Availability of research datasets
- Lack of operational solutions
- Arms race dynamics
- Narrow detectors

Research Objective

- Most research focuses on limited scope
  - Bae focuses only on file activity [4]
- Significance
  - Combined/hybrid model
  - Increase malware detection rate
- Expected outcomes
  - PoC model to detect malware using host-based and network-based features
  - Low false-positive, false-negative rate

Proposed Solution

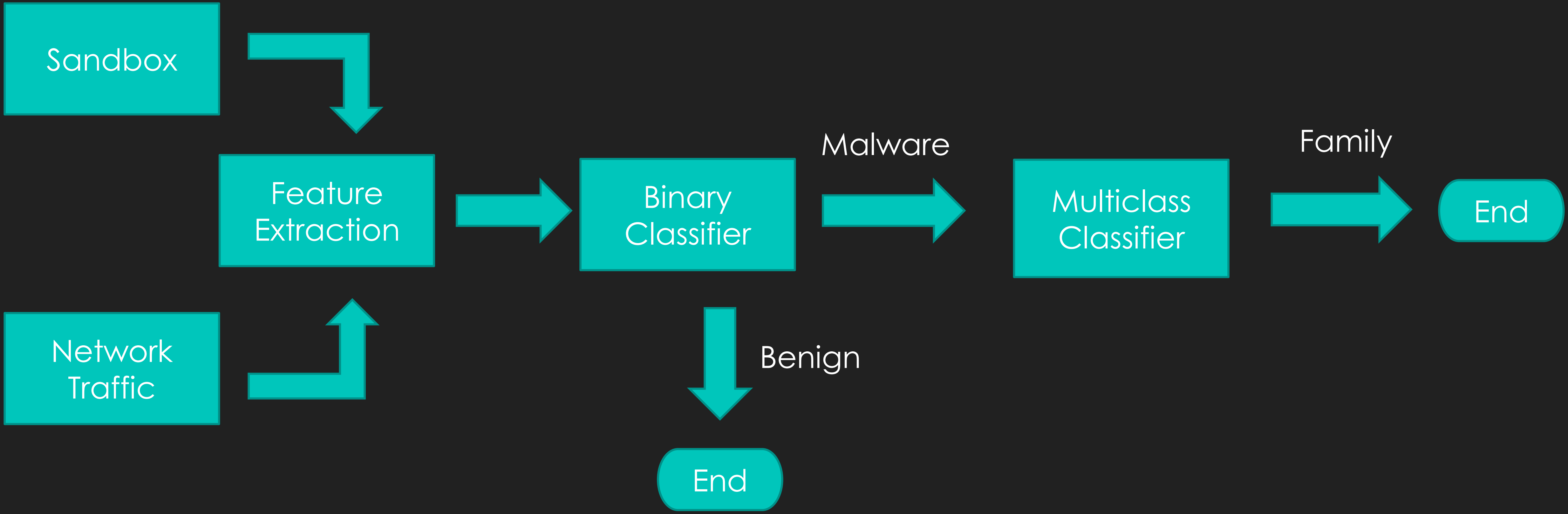
Working Prototype: Hybrid malware detection scheme using machine learning

Multiple detectors

- Host based
- Network based

Binary (benign/malware) and Multiclass classifiers (malware family)

Proposed Model



Unknowns

- Integrating Network – Host detectors
- Feature selection
- Research datasets

Thank You

- Group SSL
  - Gaston Carvallo
  - Loyd Rafols
- Website: https://rea.000109.xyz
- Course: REA605 Winter 2021, Mark Shtern

References

- [1] Symantec. 2020, “Internet Security Threat Report Volume 24, February 2019”. ISTR, Symantec Corporation, Last modified 2019. https://docs.broadcom.com/doc/istr-24-2019-en
- [2] Brunau, Chris, 2018. “Ransomware News: WannaCry Attack Costs NHS Over $100 Million”. Datto, last modified October 18, 2018. https://www.datto.com/uk/blog/ransomware-news-wannacry-attack-costs-nhs-over-100-million
- [3] IBM, 2021. Cost Of A Data Breach Report 2020. Ebook.
- [4] Bae, Seong Il, Gyu Bin Lee, and Eul Gyu Im. 2018. "Ransomware detection using machine learning algorithms." *Concurrency and Computation: Practice and Experience* 32, no. 18 (2020): e5422. doi: 10.1002/cpe.5422