



# Numerical Representations of Internet Hosts and Network Signatures

TECHNOLOGY NUMBER: 7670

## OVERVIEW

Converts raw information about Internet hosts into low-dimensional numerical embeddings

- Can be used to perform scalable analysis of entities on the Internet
- Available open source under MIT license

## BACKGROUND

The development of open-source network scanners provides researchers and security/network analysts with large amounts of raw information on arbitrary hosts on the Internet, including web servers, routers, webcams, and so forth. These measurements can be used to reveal misconfigurations, uncover hosts susceptible to specific software vulnerabilities, or collect statistics and trends on usage of various protocols and technologies.

## INNOVATION

Researchers at the University of Michigan have utilized deep learning methods to develop a methodology and an associated process for converting raw information about Internet hosts to low-dimensional numerical embeddings (representations), and propose techniques for aggregating these embeddings at the network level to obtain numerical fingerprints (signatures) of entities on the Internet (e.g. autonomous systems or organizational networks).

## ADDITIONAL DETAILS

The code can be accessed in GitHub repositories, located at either of the following sites:

<https://github.com/arsarabi/jsonvectorizer>

<https://github.com/arsarabi/vae>

## References

1. Sarabi A, and Liu M. , Characterizing the Internet Host Population Using Deep Learning: A Universal and Lightweight Numerical Embedding. Proceedings of the Internet Measurement Conference 2018. October 2018, pages 133-146.

## Technology ID

7670

## Category

Software & Content

## Inventor

Armin Sarabi

Mingyan Liu

## Further information

Ashwathi Iyer

[ashwathi@umich.edu](mailto:ashwathi@umich.edu)

## Learn more

