Harry Trinh

🙋🏻‍♂️ Brief introduction

> Hi! I'm a product-focused senior machine learning engineer, currently working at #themedium.io and #mergerity.com.

I was a data scientist at POPS WORLDWIDE from 2018 to 2019 in Vietnam and now I am working as a senior ML engineer at TheMedium.io and Mergerity.com (subsidiary company) based in Seoul, South Korea. Recently, I received a Master of Engineering from Soongsil University and published 2 Q1 international journals and 4 international conference papers as the first author in the field of AI and network security. I was also a co-author of different papers in the same field.

Please refer to this **link** to see my publications.

Or please visit my website: neuralfast.com

👔 Work experience

Senior ML Engineer

Themedium.io (mergerity.com - subsidiary company), *Gangnam, Seoul - Mar 202 - Now.*

Mergerity is a social network company where augmented reality (AR) and artificial intelligence play a key role.

My roles:

- Building a variety of real-time AI applications for AR technology.

- Building an end-to-end data pipeline using AWS services (EKS, ECR).

- Building a landmark retrieval AI model using metric learning.

Pytorch, Tensorflow, Redis, EKS, ECR, Uvicorn, FastAPI, Docker are the main tools that I use them on a daily basis.

Data Scientist

POPS Worldwide, *HCMC, Vietnam, Jul 2018 - Mar 2019*

POPS App is an all-inclusive digital entertainment platform used and beloved by millions of fans every day. Available on all App stores and leading digital platforms, POPS is the leading digital entertainment company in Southeast Asia, providing over 400 million subscribers with access to the very best entertainment the region has to offer. Over the past 13 years, POPS has mastered the art and science of connecting brands with viewers and building up communities through meaningful digital content.

- Led POPS's data team, including one business analyst, one data analyst, and one system administrator.

- Designed and built a complete data pipeline for a finance department with ≈ 35M records per month (2TB in total) on Google Cloud Platform (GCP), including ETL, data warehouses, analytics, and BI.

- Designed and built various reporting dashboards using Power BI, retrieving data from a data warehouse on GCP.

- Created a tool that automatically downloads monthly reports through YouTube Reporting API using Python (saved $2800 each month).

- Migrated on-premises data to GCP and built internal tools: web crawlers, parsing tools using Python.

Complex SQL, Google BigQuery, PowerBI, and Tensorflow were the tools that I mainly used.

Junior ML Engineer

CBD Robotics, *HCMC, Vietnam, Jul 2018 - Mar 2019*

Founded by a group of Vietnamese American and American professionals in the field of data management, data science, Computer Science and Robotics, CBD Robotics is able to provide the best products and services to help agencies, companies and industries address the most sensitive and complex challenges of managing large data, apply artificial intelligence to data to predict events, especially in Marketing and Medical Diagnostics. The combination of sophisticated algorithms and high quality software design makes our software products and services bringing valuable experience to our customers, and help to gain the trust of millions of users, as well as creating the predicting data.

Built various ML-based products with a team:

- Built a deep-learning-based lung cancer detection solution using X-ray images collaborated with Cho Ray hospital.

- Built a CBD diagnosis website integrated with a trained detection model using Flask.

- Did various data science problems: NLP classification, clustering, and sentiment analysis at CBD bootcamp.
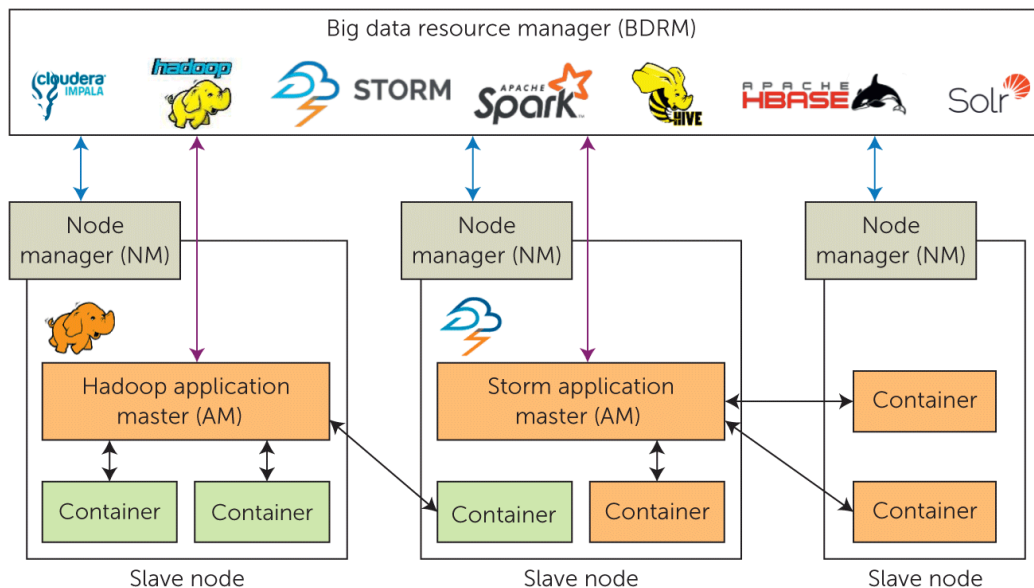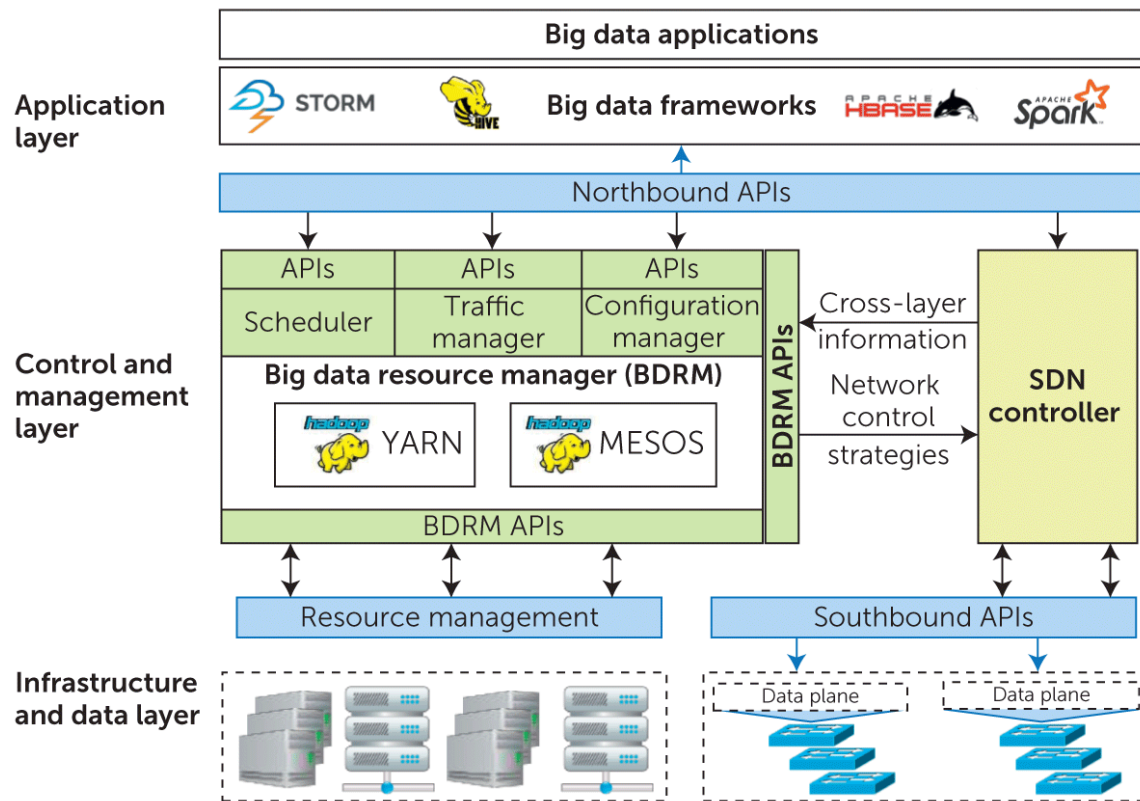
👨‍💻 Side projects

Some typical projects that I have been working on over the years are listed at http://neuralfast.com/#projects**.**

I. Masters Research Projects

### 1. (4-2020) Big Data integrated in SDN architecture

Development of A Framework for DDoS attack in SDN-based cloud.

2. (11-2019) Economic Denial of Substainability Attack in SDN architecture on cloud (in collaboration with KAIST University)
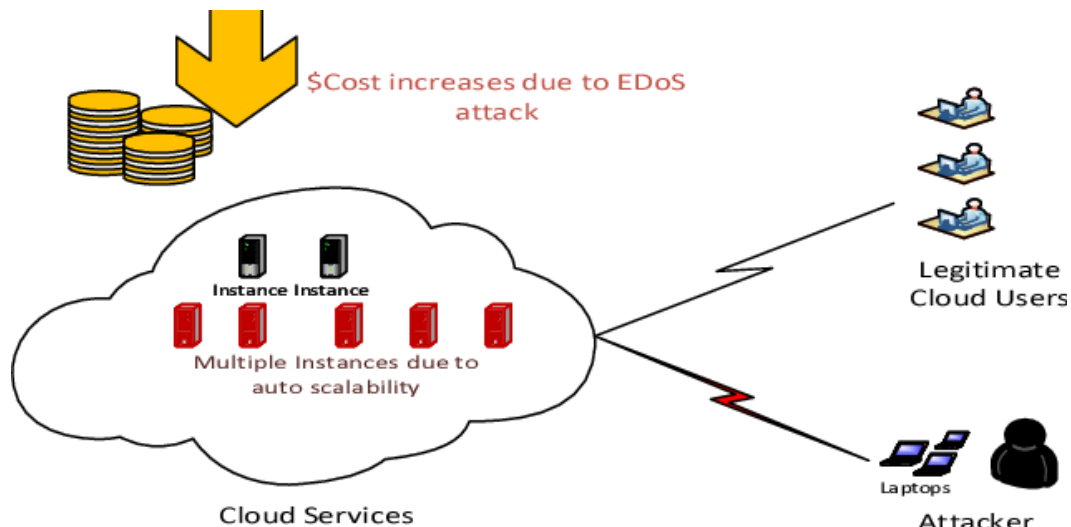
Development of Machine-learning based Network Attack Detection and Dynamic Defense Security System using SDN-based NFV To Detect EDoS Attacks

3. (8-2019) - Compromised Switch Attack Detection in SDN architecture (in collaboration with KAIST University)
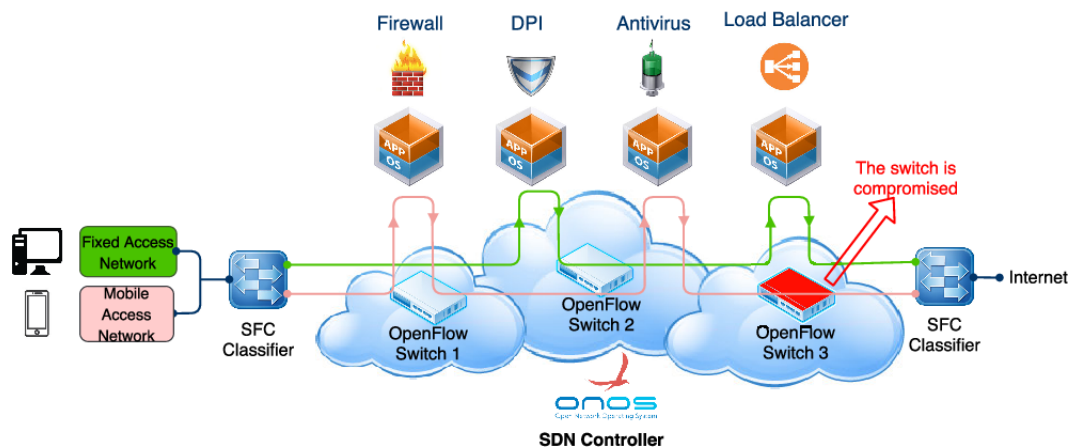
FIGURE 1. SDN-based cloud environment prototype including a compromised switch.

Development of Machine-learning based Network Attack Detection and Dynamic Defense Security System using SDN-based NFV To Detect Compromised Switches.

Link to the project: http://neuralfast.com/project/researchProject/

All achievements from 3 mentioned-above projects are listed in here

II. Global Covid19 Prediction Project

Covid-19 Research project

Dr. Truong Van Nguyen (Jimmy)
Lecturer in Operation & Information System,
Brunel University London, UK
TruongVan.Nguyen@brunel.ac.uk

Prof. Li Zhou
Professor in System Management & Strategy,
University of Greenwich, UK
Li.zhou@greenwich.ac.uk

Phuc Trinh (Harry)
Data scientist, SNS lab, Seoul, South Korea
phuctrinh@soongsil.ac.kr

Toan Viet Tran
Ho Chi Minh City University of Technology, VN
1613611@hcmut.edu.vn

Minh Cong Tran
Data scientist, MB shinsei Financial, VN
minhtc.uet@gmail.com

1. **Research background**

In January 2020 the COVID-19 virus started to spread from Wuhan to other countries and the numbers of those affected and those sadly dying began to rise at 20-30% per day in Iran, Italy, South Korea and then in other European countries which became the main focus of spread in March, and then into USA, India, and Africa.

The rapid spread of COVID-19 globally has created a wide range of responses from governments. Common measures include school closings, travel restrictions, bans on public gatherings, emergency investments in healthcare facilities, new forms of social welfare provision, contact tracing and other interventions to contain the spread of the virus, augment health systems, and manage the economic consequences of these actions.

Link to the project: http://neuralfast.com/project/GlobalCovid19/

🧤 Other Activities

- Marshal Team Leader of a team with 55 members in challengevietnam.com

Dates volunteered: Feb 2015 – Apr 2015

Volunteer duration: 3 months

🛠️ Skills and Technologies

- Programming Languages: Python, Scala, Java, R, DAX, C++.

- Databases: Oracle SQL, MySQL, SQL Server, PostgreSQL.

- Big data: AWS Services (EC2, S3, EFS, RDS, DynamoDB, Aurora, CloudWatch, SageMaker, EMR, Kinesis, ElasticSearch, Glue, S3DistCP, Athena, Redshift), BigQuery, Google Cloud
Engine, Apache Kafka, Apache Spark, Apache Hadoop, Apache Parquet, Zookeeper, Cloudera.

- ML framework: Scikit learn, Keras, Tensorflow.

- BI tools: Power BI, Tableau.

- Other: Git, Slack, Confluence, Jira.

📚 Education

[Master of Engineering in Computer Science]

Soongsil University, Mar *2019 - Feb 2021*

GPA: 4.13/4.5

[Bachelor of Engineering in Computer Science]

Telecommunications *University*, *2014 - 2018*

GPA: 3.34/4.0

Publications

R-EDoS: Robust Economic Denial of Sustainability Detection in an SDN-based Cloud through Stochastic Recurrent Neural Network

IEEE ACESS - The Multidisciplinary Open Access Journal [Q1 - SCI, IF=5.16].

Cloud computing is now known as the most cost-effective platform for delivering big data and artificial intelligence services over the Internet to enterprises and

cloud consumers. However, despite many recent security developments, many cloud consumers continue to express great concern about using these platforms because they still have significant vulnerabilities. Typically, Economic Denial of Sustainability (EDoS) attacks exploit the pay-as-you-go billing mechanisms used by cloud service providers, so that a cloud customer is forced to to pay an extra fee for the additional resources triggered by the attack activities...

Trinh Dinh Phuc, Minho Park

PDF Project DOI

BDF-SDN: A Big Data Framework for DDoS Attack Detection in Large-Scale SDN-Based Cloud

The 2021 IEEE Conference on Dependable and Secure Computing - Japan, Jan 2021.

Software-defined networking (SDN) nowadays is extensively being used in a variety of practical settings, provides a new way to manage networks by separating the data plane from its control plane. However, SDN is particularly vulnerable to Distributed Denial of Service (DDoS) attacks because of its centralized control logic. Many studies have been proposed to tackle DDoS attacks in an SDN design using machine-learning-based schemes; however, these feature-based detection schemes are highly resource-intensive and they are unable to perform reliably in such a large-scale SDN network where a massive amount of traffic data is generated from both control and data planes. This can deplete computing resources, degrade network performance, or even shut down the network systems owing to being exhausting resources. To address the above challenges, this paper proposes a big data framework to overcome traditional data processing limitations and to exploit distributed resources effectively for the most compute-intensive tasks such as DDoS attack detection using machine learning techniques, etc. We demonstrate the robustness, scalability, and effectiveness of our framework through practical experiments.

Trinh Dinh Phuc, Minho Park.

PDF Project DOI

Economic Denial of Sustainability (EDoS) Detection using GANs in SDN-based Cloud

2021 IEEE Eighth International Conference on Communications and Electronics - Phu Quoc, Vietnam, Jan 2021.

Cloud computing is now considered to be the most cost-effective platform for offering business and consumer IT services over the Internet. However, it is prone to new vulnerabilities. Specifically, a newly discovered type of attack, called an economic-denial-of-sustainability attack known as EDoS, exploits the pay-per-use model to scale up the resource usage over time to the degree that the cloud user has to pay for the unexpected usage charge. To prevent EDoS attacks, we propose an effective solution in the SDN-based cloud computing environment. We first introduce a machine-learning-based approach adopting a framework called MAD-GAN which applies an unsupervised multivariate anomaly detection technique based on Generative Adversarial Networks (GANs)...

Trinh Dinh Phuc, Minho Park

Project PDF DOI

ECSD: Enhanced Compromised Switch Detection in an SDN-Based Cloud Through Multivariate Time-Series Analysis

IEEE ACESS - The Multidisciplinary Open Access Journal [Q1 - SCIE, IF=5.16].

In our previous work, we proposed an efficient scheme for detecting compromised SDN switches based on chaotic analysis of network traffic using an autoregressive-integrated-moving-average model. This scheme showed good results overall; however, it still showed high false-alarm rates due to a hard-set threshold. In this paper, we propose an enhanced scheme to detect compromised SDN switches effectively and reliably. The scheme consists of two phases (online and offline), leveraging the advantages of a stochastic recurrent neural network variant of multivariate time-series-based anomaly detection. Our main idea is to capture the normal patterns of multivariate time series by learning strong representations with the key techniques, such as planar normalizing flow and stochastic variable connection, then reconstruct input data by the representations, and use the reconstruction probabilities to find anomalies. Evaluation results of our proposed scheme yield outstanding performance in comparison with our previous work and other solutions.

Trinh Dinh Phuc, Minho Park

PDF Project DOI

Dynamic Economic-Denial-of-Sustainability (EDoS) Detection in SDN-based Cloud

The Fifth International Conference on Fog and Mobile Edge Computing (FMEC 2020), 01 July 2020- Paris, France.

To prevent EDoS attacks, we propose an efficient solution in the SDN-based cloud computing environment. In this paper, we first apply an unsupervised learning approach called Long Short-Term Memory (LSTM), which is a multivariate time series anomaly detection, to detect EDoS attacks. Its key idea is to try to predict values of the resource usage of a cloud consumer (CPU load, memory usage and etc). Furthermore, unlike other existing proposals using a predefined threshold to classify the anomalies which generate high rate errors, in this work, we utilize a dynamic error threshold which delivers much better performance. Through practical experiments, the proposed ...

Trinh Dinh Phuc, Minho Park

PDF Project DOI

An Effective Defense Against SYN Flooding Attack in SDN

The 10th International Conference on ICT Convergence, Ramada Plaza Hotel, Jeju Island, Korea, Sep 2019.

SYN Flooding Attack, one of the typical Denial of Service attacks, may not only exhaust the resource of a victim but also paralyze the entire SDN network by a large number of control messages between controllers and SDN switches. Although various approaches have been proposed to defend the SYN flooding attack, they still have some drawbacks such as packet processing overload and delay. Therefore, this paper proposes an efficient SYN flooding defense scheme utilizing the TCP Time Out mechanism and Round-Trip Time (RTT). The experiment results show the proposed scheme can defend the attack with low bandwidth occupation between the controller and SDN switches and little computing resources...

Kim D., **Trinh Dinh Phuc**, Noh Sinchul, Minho Park

PDF Project DOI

Abnormal SDN switches detection based on chaotic analysis of network traffic

2019 25th Asia-Pacific Conference on Communications (APCC), Ho Chi Minh - Vietnam, Sep 2019.

Network flow is susceptible to disruption through a software-defined network caused by malicious switches. The malicious behaviors such as dropping traffic, adding or delaying traffic are diverse. Once a switch is compromised by an attacker, the switch could be malfunctioning or configured incorrectly. In this paper, we propose a real-time method of detecting compromised SDN switches based on chaotic analysis of network traffic. An ARIMA model is used to predict the number of flows in every following three seconds. Then, by calculating the maximum Lyapunov exponent, the chaotic behavior of prediction error time-series is analyzed. Simulation findings indicate that 99.63% of traffic states can be accurately classified by the proposed algorithm.

Trinh Dinh Phuc, TaeHee Lee, Thang Nguyen Canh, Sa Pham Dang, Noh Sinchul, Minho Park

PDF Project DOI

International Talks

1. Feb 01, 2021 13:30 AM Aizuwakamatsu, Fukushima, Japan **[link]**

2. Jan 12, 2020 16:40 PM — 17:00 PM Phu Quoc island, Vietnam **[link]**

3. Jun 30, 2020 11:40 AM — Jul 2, 2020 12:00 PM Paris - France **[link]**

4. 2019 KAIST University and Ho Chi Minh city - Vietnam **[link]**

🗣️ Languages

English 🇺🇸

Full professional working proficiency.

Vietnamese 🇻🇳

Native

Korean 🇰🇷

Beginner

☎️ Contact information

📧 Email: phuccoi996@gmail.com

✍🏻 Blog: neuralfast.com

🔗 LinkedIn: https://www.linkedin.com/in/phuccoi96/

🙏🏻 Thank you!

> Looking forward to hearing from you.
>
> - Harry Trinh