NOracle: Who is communicating with whom in my network?

Patrick Kalmbach  
Technical University of Munich  
Munich, Germany

David Hock  
Infosim GmbH & Co. KG  
Würzburg, Germany

Fabian Lipp  
Infosim GmbH & Co. KG  
Würzburg, Germany

Wolfgang Kellerer  
Technical University of Munich  
Munich, Germany

Andreas Blank  
Technical University of Munich  
Munich, Germany

ABSTRACT

This demo presents NOracle: a system using Stochastic Block Models (SBMs) to infer structural roles of hosts and communication patterns of services in networks. NOracle can be used with existing monitoring systems to analyze and visualize networks in an online manner or be used to analyze stored traces. Network operators can use SBMs to monitor and verify network operation, detect possible security issues and change-points. To showcase this, NOracle combines the production-grade network management solution StableNet with an SBM based anomaly detection and network visualization module. StableNet provides network flow statistics in real-time from actual devices. The SBM extracts roles and communication patterns live from the data provided by StableNet. The result can help to reason about communication behaviors, detect anomalous hosts and indicate changes in the large scale-structure of network communication.

KEYWORDS

Anomaly Detection, Stochastic Block Model, Network Monitoring

1 INTRODUCTION

Answering the questions of “who is communicating with whom in my network” can help operating today’s and future self-driving networks in many directions. For instance, knowing the communication pattern of applications in data centers helps improving resource management systems, e.g. speeding up the completion times of distributed data processing applications. In particular, data driven resource management systems for placement and embedding tasks like [1, 2, 15] can use communication patterns as basis for their predictions, and thus help networks to run themselves.

Furthermore, inferring communication patterns can help detecting security holes, e.g., infected hosts being part of a botnet [3, 6, 14]. A better understanding of the communication behaviors of users and services is crucial to make networks self-driving and thus inevitable for future communication paradigms such as application-aware networking needed for low-latency networks, such as 5G.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGCOMM ’19, August 19–23, 2019, Beijing, China  
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-6886-5/19/08...

$15.00  
https://doi.org/10.1145/3342280.3342303

2 NOracle: A DATA-DRIVEN APPROACH

NOracle uses Stochastic Block Models (SBMs) [9] to separate hosts into meaningful groups. A SBM is a PGM that represents a parametric probability distribution over graphs [9]. The model encodes

1 In a TDG nodes correspond to IP addresses and edges to flows/communication between these addresses.
We represent TDGs as unweighted graphs, i.e., $A$ is provided by the user or can be estimated from data using the k-nearest neighbors ($k$NN) approach and node metadata [12] to boost model accuracy.

While the model can work completely unsupervised, human knowledge can still improve the overall system performance drastically. For instance, by roughly knowing the services inside a network, a system administrator can help to faster bootstrap the system or choose a more suitable value for $k$. We will showcase both examples in our demo.

3 DEMO

The demo presents how NOracle can (1) infer the communication structure of applications and (2) based on this information detect anomalous hosts, i.e., hosts infected with malware or generally with a suspicious communication pattern.

Scenario. The demo considers three scenarios: (1) synthetic graphs with known structure, (2) a campus network with more than 5000 hosts and (3) an enterprise network with more than 100 hosts. For all scenarios, network traffic, i.e., the packet level traces or netflow data are fed into NOracle. The data can be evaluated for different parameter settings. For instance, the demo shows how network hosts are grouped for different $k$, exposing structural roles of nodes (e.g., client vs. server). Moreover, the demo shows how live grouping of hosts can help to detect hosts with abnormal behavior, e.g., hosts which suddenly change their communication pattern when infected with malware.

Network data. For (1) we use synthetic data with planted groups. This data is generated using a SBM with pre-set parameters. The demonstration shows how known structural roles can be identified in an completely unsupervised fashion.

For (2), the demo uses the publicly available data set "CTU13 Corpus 9". The data set contains the trace of a campus network with known infected hosts. Those hosts are manually infected with the Neris malware by the authors [4]. The demonstration shows that NOracle can detect the malicious bots shortly after the malware becomes active.

Data for (3) is taken live with the network management system StableNet from a remote enterprise network testbed located in Würzburg, Germany. The enterprise network provides a testbed for trying out network management operations — it consists of more than 100 devices. StableNet is the core part "glueing" all together, i.e., it fetches networking data from all devices and makes it available. Here, the demo shows how a network operator can inspect the communication behavior of the users and services live at run-time. For example, it is possible to select the number of communication groups. Using NOracle’s GUI illustrated in Fig. 2, a network operator/administrator can investigate the evolution of the network over time, or investigate details of the communication structures within or between groups illustrated in Fig. 3. Clients that should be blocked from the outside world should not show any communication with "external" groups. Again, human knowledge is useful or even required to finally infer the semantic meaning of the communication groups.

ACKNOWLEDGEMENT

This work is part of a project that has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement No 647158 - FlexNets).
REFERENCES


