Journal of Network and Computer Applications 81 (2017) 12–23

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Towards repeatability & verifiability in networking experiments: A stochastic framework

Swati Sharma[a,*], Alefiya Hussain[b], Huzur Saran[a]

[a] Department of Computer Science & Engineering, Indian Institute of Technology, Delhi (IITD), India
[b] Computer Scientist, Information Sciences Institute, University of Southern California (USC), Los Angeles, USA

ARTICLE INFO

Keywords:
Experiment characterization
Verifiability
Repeatability
Event-based Markov chain characterization
Time-series characterization

ABSTRACT

Growing complexity of networking experiments has rendered the accuracy, repetition and reproduction of experimental results indispensable. This research work aims at developing a tool to facilitate repeatability and verifiability of experiments on diverse experimentation platforms. The process involves studying an experiment with respect to its constituent events and creating time-series based and event based characterizations from logs and measurement data. The framework thus developed was applied to 6000 experiment trials comprising of 1500 DETER testbed DNS-cache poisoning attack, 3200 ns2 web-traffic generation, 1000 real-time web-traffic generation and 300 ns3 wireless mobility experiment trials. We observed that these characterizations were sensitive to experiment configuration. We also investigated the impact of variations in topology, traffic, hardware and mobility on them (characterizations). Differences in characterizations were quantified by standard distance measures: KL Divergence, Total Variation Distance and Euclidean Distance. Our results were statistically verified by hypothesis testing and analysis through ANOVA, chi-square and correlation tests. Thus, our framework provides a direct and precise method to compare two executions of a stochastic networking experiment for simulations, emulation-based testbeds and real-time network experiments. These results form the groundwork for generating a validity management framework and will help to achieve verifiability in experiment executions.

1. Introduction

The experimental nature of networking research involves novel experiments for creation, optimization, design and execution of networking protocols. These experiments are complex and stochastic in nature. As a result, they are repeated multiple times (not necessarily consecutively) on different network platforms (wired/wireless/emulation-testbeds/simulators), network topologies and hardware apparatuses for obtaining optimal results. Other researchers also re-execute these experiments to account for the human element of error. A user repeating his own experiment has a decent idea about its (experiment's) execution and may identify if the experiment is not executed accurately. But another researcher may not possess that level of understanding into other researcher's work. Thus, the ability to accurately repeat an experiment and verify its accuracy is critical to its execution.

An experiment consists of three main components, namely, (a) deterministic components (simple programming code targeted to run a certain way), (b) non-deterministic components (rare error cases in programming code or dynamic network behavior during experiment execution), and (c) opportunistic components (cyber-attack models following multiple code paths for successful execution). Diverse sources of variation in an experiment render attaining repeatability and verifiability very difficult. For instance, an experiment's (1) physical apparatus, (2) topology, (3) software code or binaries, (4) input parameters, (5) hardware and software configuration of the nodes, (6) procedure, (7) measurement and analysis process, (8) output, (9) cross-traffic involved, and (10) other network constraints originating from the dynamic network behavior.

1.1. Denotations

Reproducibility is the ability to accurately recreate experiments (same experiment output) performed by other researchers or on other networks/apparatus. *Repeatability* is the ability to accurately replicate experiments (same experiment output) performed by the same researcher at different times. *Verifiability* is the ability to compare outcomes (output and measurement logs) of two trials. Rapidly

* Corresponding author.
E-mail address: swati.sharma.rs@gmail.com (S. Sharma).

http://dx.doi.org/10.1016/j.jnca.2016.07.001
Received 24 July 2015; Received in revised form 7 April 2016; Accepted 1 July 2016
Available online 03 January 2017
1084-8045/

evolving research on future networking applications, protocols and monitoring demands valuable experimental research along with its validation.

On a network, any network phenomena that we wish to observe can be termed as an experiment. Each execution of this experiment is called a trial and the smallest entity of interest in this trial is called an event. In this paper, when we refer to repeatability, we refer to reproducibility as well.

1.2. Motivation

The motivation of this work comes from some of the major challenges currently encountered by a typical testbed user. These include – (i) inability to directly compare two repeated experiment executions (experiments and traffic models are inherently stochastic). (ii) Lack of user understanding of errors and faults in underlying testbed infrastructure and resulting impact on the experiment results (heterogeneous testbed infrastructure can impact hardware-dependent experiment measurements in unanticipated ways). (iii) Characterization of event logs and experiment measurements is still a manual ad-hoc process. There is a lack of metrics that would archive experimentation scripts as well as provide *measures of agreement* between two experiment trials. (iv) With increasing experiment scale, an experiment's complexity increases exponentially (numerous networked entities need setup and coordination). Thereby, making the task in step iii) even more challenging. (v) Inability to verify an experiment execution (replicating an experiment does not guarantee its correctness).

1.3. Illustration with file transfer experiment

Consider a very simple experiment - a FTP file transfer. The output of this experiment would be obtaining the complete file at the destination node through FTP. Repeated file transfers would form independent trials of this experiment. The conditions of this experiment can be monitored by carrying out the analysis of performance statistics (inserted traffic throughput, cross-traffic interference, experienced packet loss, the time taken for file transfer, etc). Ensuring repeatability and reproducibility in this experiment would mean achieving similar performance statistics from trials executed by one's own self or other researchers, respectively. Ensuring verifiability would mean similarity of measurement logs in the event of packet losses incurred on the network in the two experiment trials (measurement logs may differ due to performance statistics mentioned above).

1.4. Contributions

First and foremost, we provide syntax and semantics of a framework to directly characterize an experiment execution. Its innovative features are the following. (i) This framework extracts information from the event logs and measurement data obtained from the experiment's execution. (ii) It creates characterizations based on Time-series and Event-based Markov Chains. (iii) We demonstrate that these characterizations are experiment specific and therefore distinct. The contribution of this work is not an innovative way to create this Markov Chain distribution, but rather to use this distribution to provide a direct approach to characterize an experiment's execution.

Secondly, we demonstrate that identical experiment trials can be examined to investigate if they are similar or not. Thus, by extension, this research presents an approach to compare two experiment trials (by using standard distance measures like KL distance, Total Variation Distance and Euclidean Distance), thereby verifying an experiment's execution.

Thirdly, we demonstrate that our framework is sound, statistically rigorous as well as sensitive to changes in the experiment's configuration (topology, traffic, hardware and mobility). These changes affect the experiment's characterization. We extensively test our framework for a few classes of experiments (opportunistic and heavy-tailed) on simulators, emulated testbeds and real networks. Our set of experiments includes 1500 DETER testbed DNS-cache poisoning attack experiment trials, 3200 ns2 web traffic generation experiment trials, 1000 real-time web traffic generation experiment trials and 300 ns3 wireless mobility experiment trials. We built different implementations for different experimentation platforms (for instance C based libpcap implementation for DNS cache poisoning attack experiment and scripts for ns2 & ns3 simulations). We also establish that these characterizations are feasible under transformations of scale and complexity.

2. Key idea

A typical networking experiment, $\mathcal{E} = (\mathcal{A}, \mathcal{P}, \mathcal{L}, \mathcal{M})$, is composed of an apparatus description $\mathcal{A}$ and a procedure description $\mathcal{P}$. *Apparatus description* defines the topological structure and device configurations for the experiment. It is annotated with properties like bandwidth, delay, loss rates, and queuing mechanisms. On the other hand, *procedure description* defines experiment work-flow to specify implementations on hosts, events in the experiment, and sensors that record experiment state. For instance, tcpdump, CPU utilization monitors and memory utilization monitors.

Henceforth, in this paper, we refer to each experiment procedure execution as an *experiment trial*. Following are the artifacts that can be analyzed upon a trial's completion. (i) Testbed Allocation Logs and Procedure Orchestrator Event Logs (together denoted as $\mathcal{L}$). (ii) Measurement Data from instrumentation system (denoted as $\mathcal{M}$). In this paper, we use $\mathcal{P}$ and $\mathcal{M}$ (and not $\mathcal{A}$ and $\mathcal{L}$) for the experiment characterization process.

2.1. Brief summary of approach

If the foundation of building an experiment's characterization is laid upon its execution details, this information could also be worthwhile in verifying repeated executions of the said experiment. At the most fundamental level, such a characterization could be based on experiment information like the number & type of events occurring in the experiment's execution and the sequence of occurrence of these events. Just as a network administrator maintains a fingerprint database for device identification in a network, similarly, a repository could be maintained for archiving characterizations corresponding to the different network experiments. The process of verifying the correct execution of an experiment would then simply be reduced to the trivial task of comparing (within an accepted error range) characterizations from fresh experiment executions against those stored in the repository. Fig. 1 depicts this process with sample trials of a typical stochastic networking experiment.

The sequence of execution of a set of events is, hence, critical to creating the experiment's characterization. This sequence also includes the inter-dependencies between the constituent events. We observed that the distribution of these event sequences varies with a change in the context of the experiment. Every experiment has a skeleton set of events (or backbone) that must take place to define the experiment. Additional background traffic and network constraints may add onto the experiment's event skeleton. But the basic event distribution remains the same. Changes that affect an experiment's context (or nature) also influence this event skeleton of the experiment. These changes can be easily spotted in the event based experiment characterization. Our framework uses these event distributions forming the experiment skeleton and further, processes them to create a Markov Chain based characterization. The contribution of this work is not an innovative way to create this Markov Chain distribution, but rather to use these event distributions to provide a direct approach to characterize an experiment's execution.

Similar characterizations do not necessarily imply that the experiment is repeatable. If two characterizations are similar, then it implies that we can compare the two experiment executions (trials) that gave rise to the two characterizations. Thus, this framework functions as a tool to determine if an experiment execution is verifiable by comparing its characterization with the known archived experiment characteriza-