# Measuring and understanding errors in the Internet

Susmit Shannigrahi, Craig Partridge, Anton Betten

The Internet incorporates a wide variety of applications, protocols, and hardware. Each is capable of introducing errors in data handling. In recent years, the emergence of white box hardware and high-speed links (both wired and wireless) has introduced new sources of errors into the Internet. The concurrent advent of Big Data applications has sharply increased the chance an application (if it uses Big Data) will encounter a previously rare error. Unfortunately, a comprehensive census of the errors that occur on today's Internet has not been taken. Enumerating and understanding the errors that happen in today's Internet would be useful to not only network operators and researchers but also the application developers who seek to build robust and high performance applications. The research questions we seek to answer are then (a) what errors are happening in today's Internet? (b) what is causing them? (c ) what do we need to do to fix them?

**Research Question 1: What errors are happening in today's Internet?**

***Data needed:*** Packets/frames with errors from a set of diverse network operators (e.g., Universities, ISPs, Research testbeds, Cloud providers). If possible, corresponding good packets if they were retransmitted. ***Infrastructure needed:*** Packet capture software/containers across different networks. The ability to capture packets with errors without slowing down forwarding. The ability to capture retransmission of packets/frames with errors. ***How NSF can support:*** **(a)** New infrastructure to continuously capture and store Internet errors over a long period of time (months to years). This will provide us with the ability to understand how errors evolve. **(b)** New analytic techniques to better understand what causes these errors **(c )** New techniques to minimize the errors while maintaining the ability to observe. For example, bypassing a faulty switch using an SDN controller is straightforward but it hides the errors.

**Research Question 2: What is causing these errors?**

**Data needed:** Packets/frames with errors. Operational details (make and model of hardware, software versions, operating environment) that are often very difficult to obtain. **Infrastructure needed:** Packet capture software that works with new high-speed links, ground truth (which conditions cause what errors), and state of the network/application at the time of the error (e.g., congested WiFi) **How NSF can support: (a)** Capture and storage of errors over a long time. **(b)** Analytic techniques that can map errors to their root causes, and ML techniques that can automatically identify root causes and predict future errors.

**Research Question 3:  How do we fix these errors?**

Based on the findings of #1 and #2, we might need to look at potentially new error detection/correction mechanisms. Moreover, given an error, finding the most efficient solutions can be tricky. **How NSF can support:** supporting fundamental research in error detection and correction methods, supporting techniques that can map potential solutions to different types of errors, supporting deployment and testing in real world scenarios (e.g., the FABRIC testbed).