The Center for Cyber Defenders

Expanding Computer Security Knowledge

GPROC

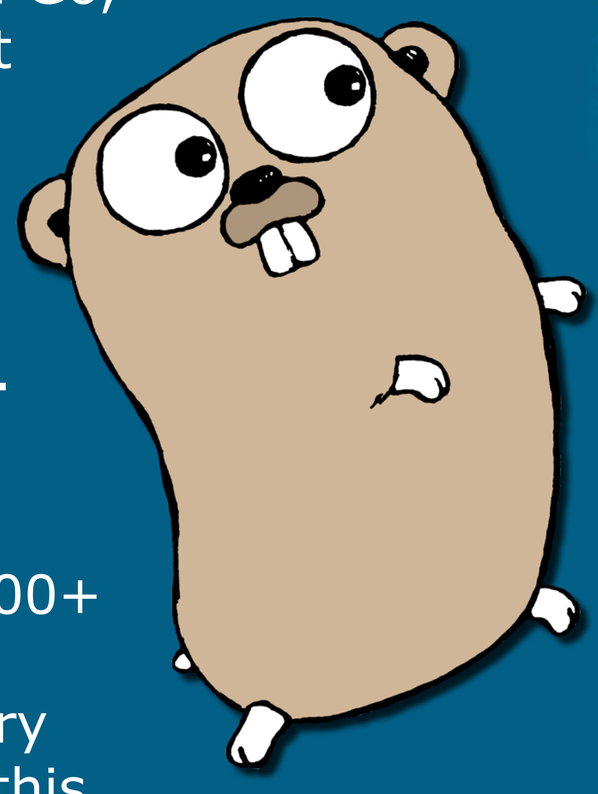
David Fritz, Oklahoma State University;
Kasimir Gabert, New Mexico Institute of Mining and Technology

Project Mentor: Ronald Minnich, 8961

SAND2011-4786P

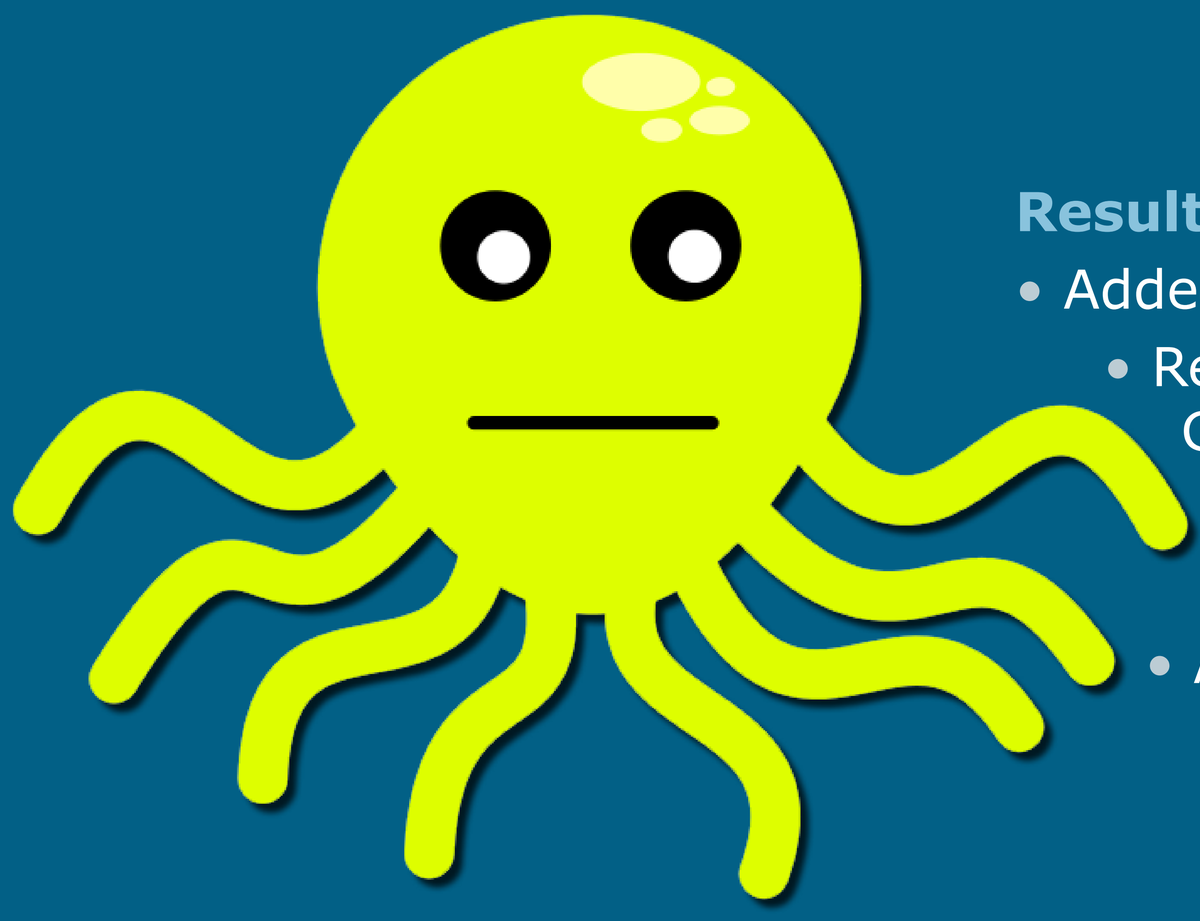
Problem Statement:

- Gproc is a large scale cluster management tool, written in Go, that provides a process startup mechanism for lightweight cluster nodes. A command and a set of nodes upon which to run it are specified, and then the binary, any required dependencies and additional files are packaged and sent to the nodes for execution. Dependencies are determined by Gproc programmatically.



Objectives and Approach:

- Gproc aims to facilitate running large scale clusters (10,000+ nodes with millions of virtual machines). A number of problems arise including synchronization, graceful recovery of lost nodes, and I/O, which require unique solutions at this scale. For example, a minimum size TCP packet with no data on 10 million nodes is about 150MB, so even a simple command requires novel transmission and merging techniques. Gproc accomplishes this by using the Go programming language, which provides easy mechanisms for scalability and concurrency. Gproc nodes form a dynamic tree structure with the ability to add or remove nodes and to partition subtrees for use by the client. To reduce I/O volume back to the client during runs, each internal node merges data in a diff-style format. Gproc also supports batch or interactive sessions.



Results:

- Added web interface to Gproc
- Rewrote network transmission code to fit the Go model
- Added diff-style I/O merging
- Ported Gproc to FreeBSD
- Added IPv6 support

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND2011-4786P.