PCR - A New Flow Metric

Producer Consumer Ratio

Carter Bullard
QoSient, LLC

carter@qosient.com

John Gerth
Stanford University

gerth@graphics.stanford.edu

FloCon 2014
January 13-16, 2014
Charleston, South Carolina

Problem Statement

- Data Exfiltration is a serious problem
  - Cyber Espionage is primary issue in US Cyber warfare
  - APT1 identifies nation state orchestration and exploitation
  - Represents for some, THE worst case scenario
- Detection can be challenging
  - Novel transport strategies can really make it hard
  - Behavioral threshold based systems easily subverted
  - Some involve internal data consolidation, with physical extraction
- For many, the problem is not Data Loss
  - Unknowingly, many infrastructure are used as stepping stones facilitating the transport of data.
  - Implicating them as co-conspirators.
  - Liability and reputation impacts are very damaging

Exfiltration Methods

- Data Exfiltration can be a complex and protracted event
  - "Night Dragon" APT active ~4 years.
  - APT1 identifies nation state orchestration and exploitation
- Majority of the nodes involved are transport nodes, not sources
- Use of overt and covert channels is common
  - IP Multicasting for extra work group exfiltration (overt)
  - Non-IP LAN based data exchange for consolidation and exfiltration
  - Browser based covert channels - DNS prefetching , Java scripting
  - DNS tunneling, HTTP tunneling, XXX tunneling
  - Piggyback transporting - NTPv3, VoIP, ICMP, SIP
- Insider exfiltration generally uses physical media
  - Normally involves data consolidation to a set of extraction nodes, to facilitate physical removal of the data.

Exfiltration Detection

- Data Loss Protection strategies fall short
  - Exfiltration is not exclusively a data loss problem
    - Many impacted by exfiltration exploits are just stepping stones
    - For these sites, content based detection fails, as its not their data.
  - Distributed exploitation frameworks, such as "mesh - in - mesh - out" fabrics, make load based identification very difficult.
- Bell-La Padula type formal methods may provide some help
  - Detect transformation from normal node to an exfiltration node
  - Need new metrics
- Propose that exfiltration is a shift in producer / consumer roles
  - Better methods to describe producer / consumers will really help
  - Early detection involves identifying leading indicators

Producer Consumer Roles

- The purpose of a communications network is to facilitate producer / consumer functions
    - The nature of the produce / consumer role can be formalized, identified, analyzed, tracked and controlled.
- All network nodes are producers and consumers of data
    - All nodes consume network control services
        - ARP and DNS
    - All nodes provide network services, some more than others
        - Switches, routers, file systems, web servers, name servers, whatever
    - Use of the network is a consumer / producer relationship
        - Nodes using applications generally are producers or consumers
        - Nodes supporting applications, generate application exchange
- Exfiltration is a modification of the highly granular and aggregated consumer / producer relationships of an organization of systems.

Producer Consumer Ratio

Novel Flow Metric

- Basic Computer Science Semantics
- Fundamental Flow Dynamic
- Basis for Behavioral Classification
- Simple Arithmetic / Statistical Operations
- Support All Flow Data Operations
    - Aggregation, Inverse, Filtering, Selection, Search, Bining, Metadata Enhancement

Producer Consumer Ratio

Definition

Intuition

A normalized value indicating directionality of application information transfer, independent of data load or rate.

$$PCR = \frac{SrcApplicationBytes - DstApplicationBytes}{SrcApplicationBytes + DstApplicationBytes}$$

$$\text{Application Bytes} = (\text{Total Bytes} - \text{Sum}(L_{[2, 3, 4]} \text{ Headers })) - \text{Retrans Bytes}$$

Producer Consumer Ratio

Properties

Consumer Producer

Range: -1.0 <= PCR <= 1.0

Proportions: Source's Fraction = (1 + PCR) / 2

Destination's Fraction = (1 – PCR) / 2

Analytics: Aggregation, Selection (filtering), Sorting, Cluster Analysis, Frequency Analysis, Classification

Sample Values:

1.0 – pure push - FTP upload, multicast, beaconing

0.4 – 70:30 export - Sending Email

0.0 – Balanced Exchange - NTP, ARP probe

-0.5 – 3:1 import - HTTP Browsing

-1.0 – pure pull - HTTP Download

PCR and Observation Domains

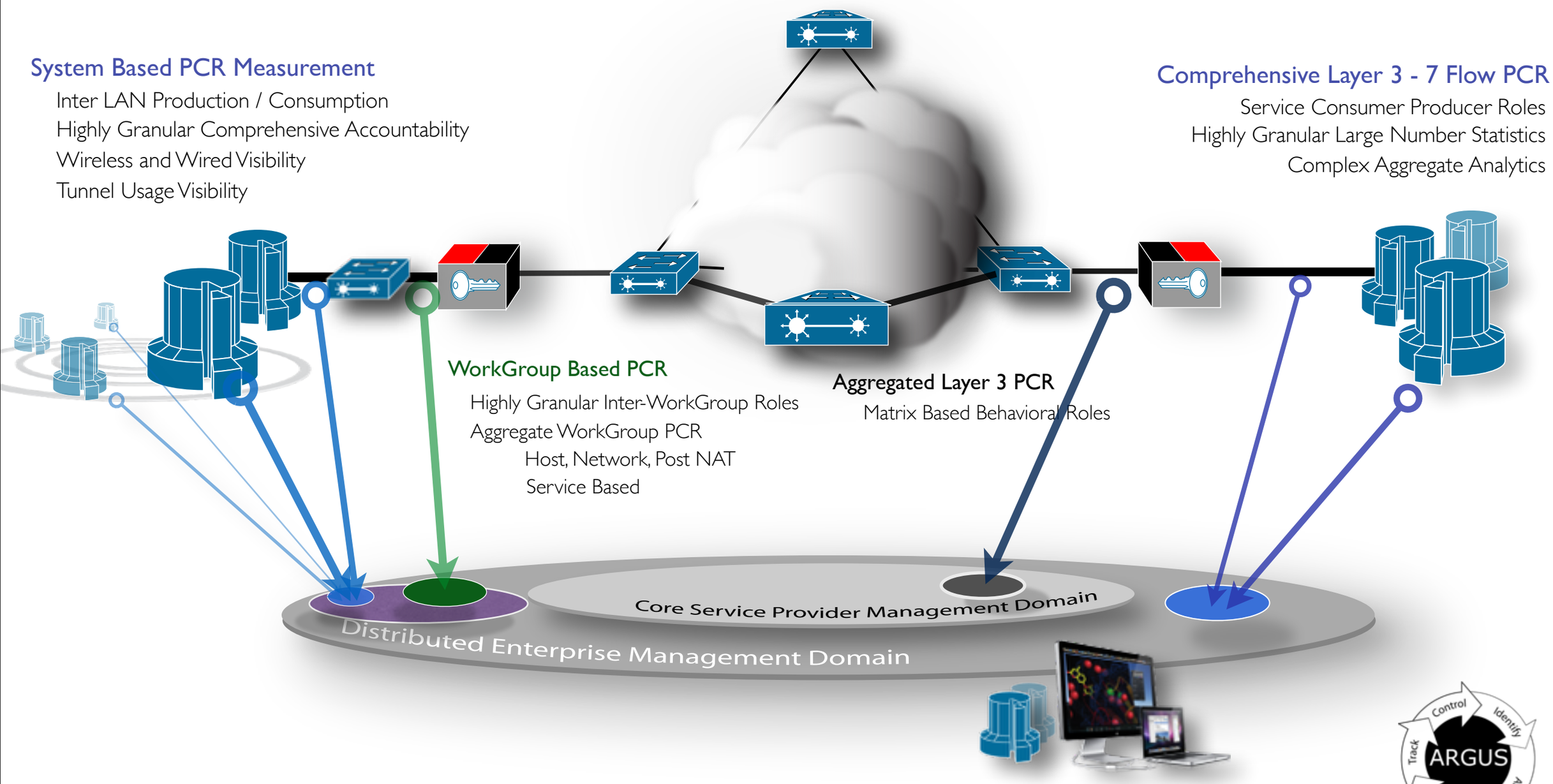
PCR Situational Applicability

System Based PCR Measurement

- Inter LAN Production / Consumption
- Highly Granular Comprehensive Accountability
- Wireless and Wired Visibility
- Tunnel Usage Visibility

Comprehensive Layer 3 - 7 Flow PCR

- Service Consumer Producer Roles
- Highly Granular Large Number Statistics
- Complex Aggregate Analytics

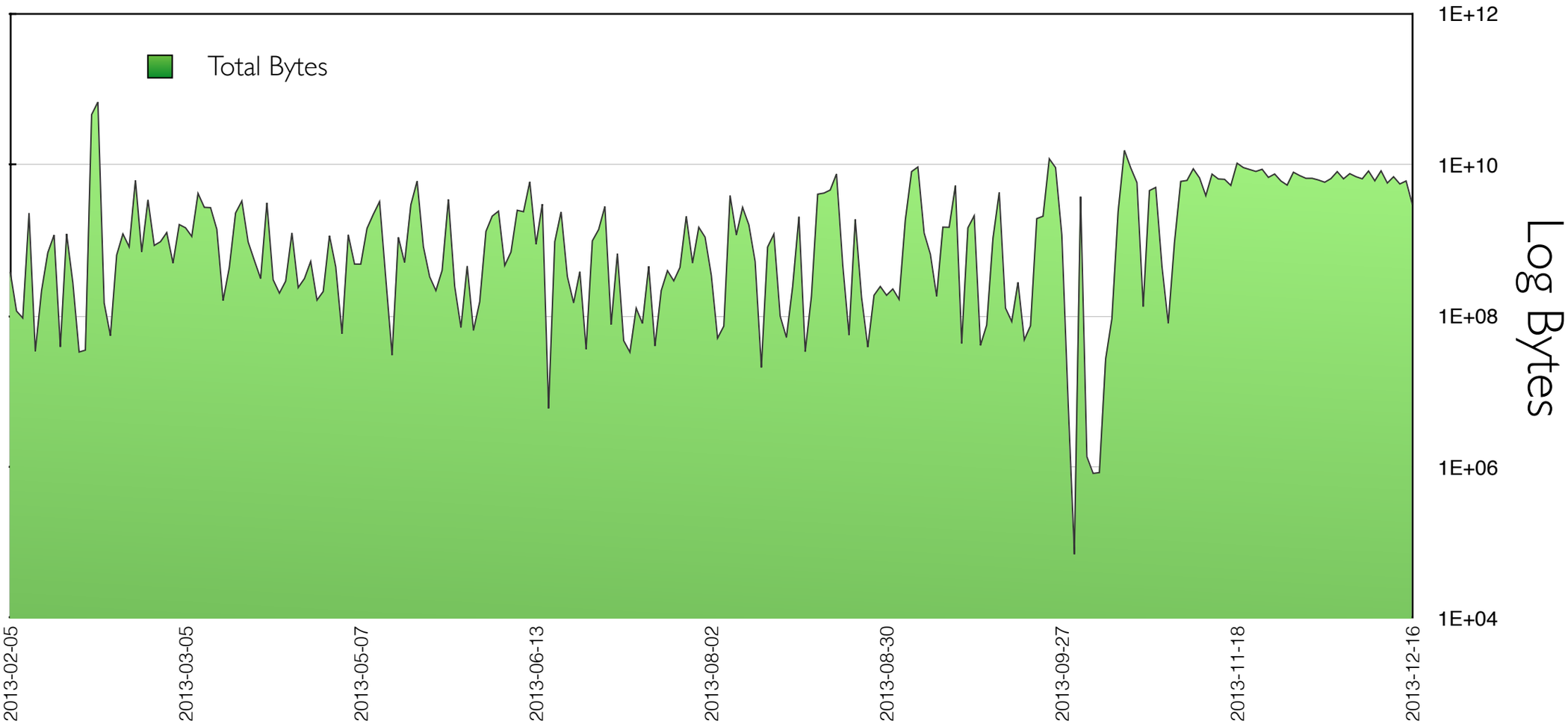


Results

PCR Application Characterization

Enterprise Aggregate PCR

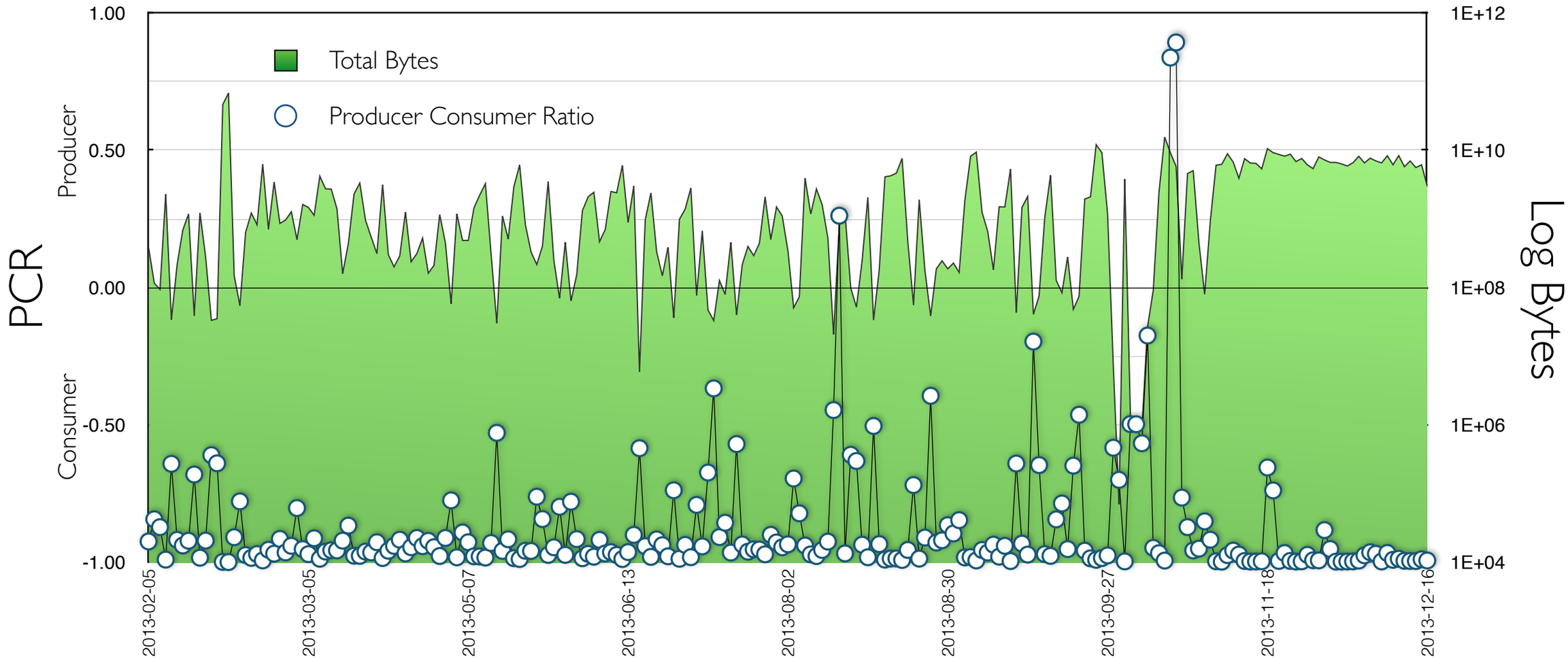
QoSient WHQ Aggregate Daily Producer Consumer Ratio



PCR Application Characterization

Enterprise Aggregate PCR

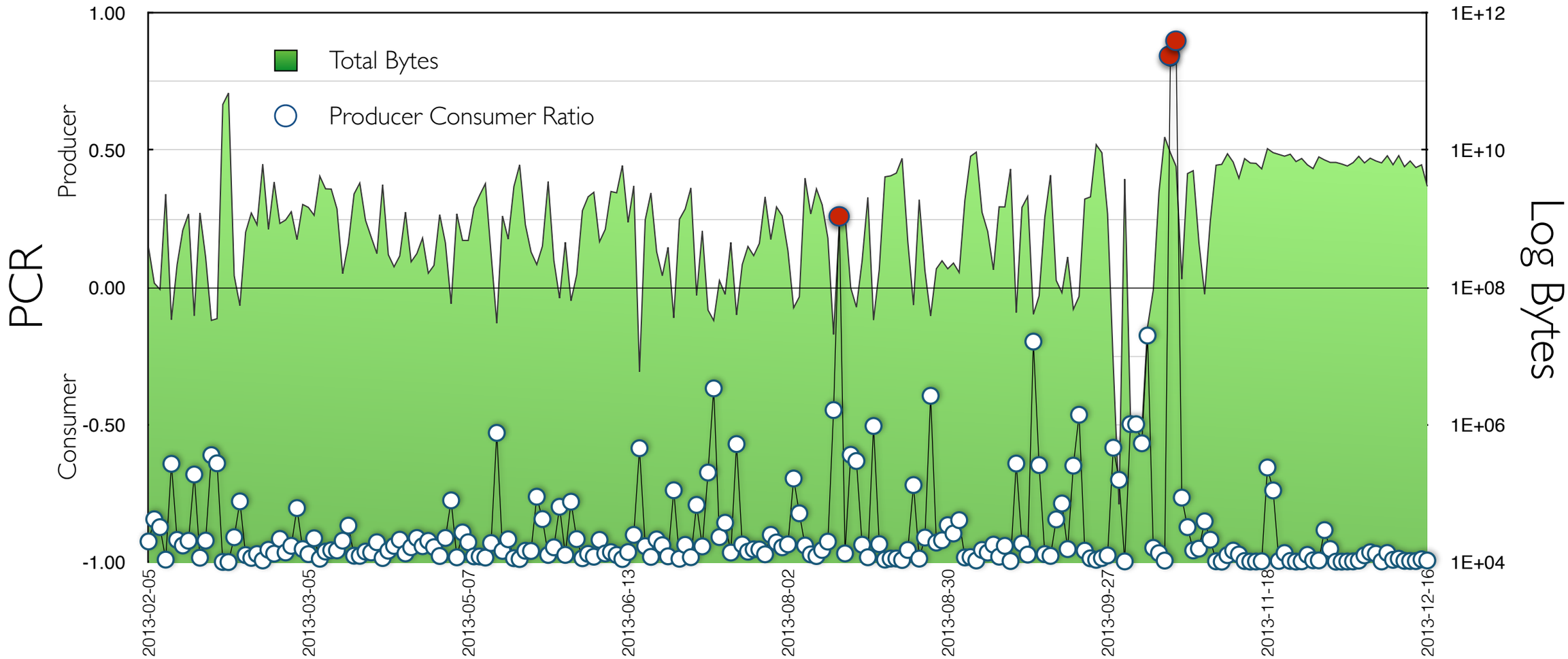
QoSient WHQ Aggregate Daily Producer Consumer Ratio



PCR Application Characterization

Enterprise Aggregate PCR

QoSient WHQ Aggregate Daily Producer Consumer Ratio

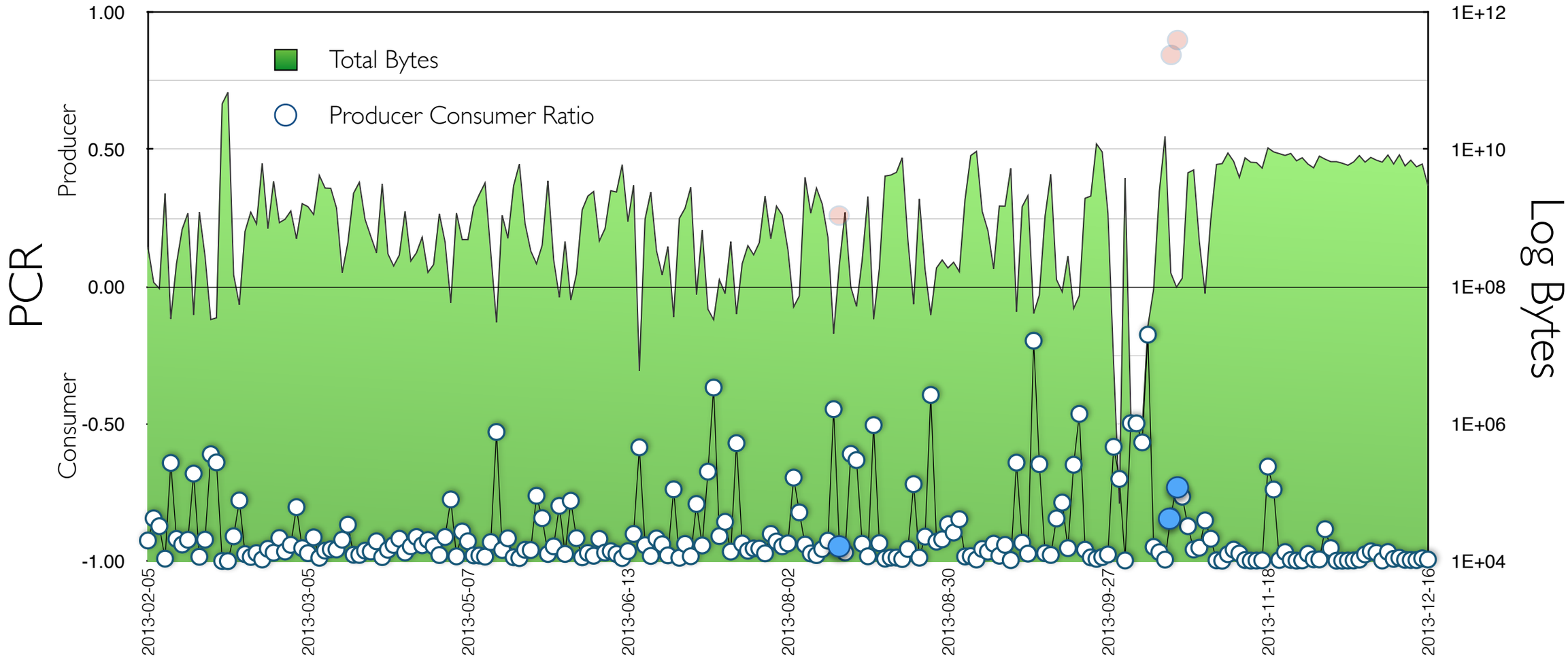


Positive Aggregate PCR caused solely by iCould awacsd - apple wide area connectivity service daemon

PCR Application Characterization

Enterprise Aggregate PCR

QoSient WHQ Aggregate Daily Producer Consumer Ratio

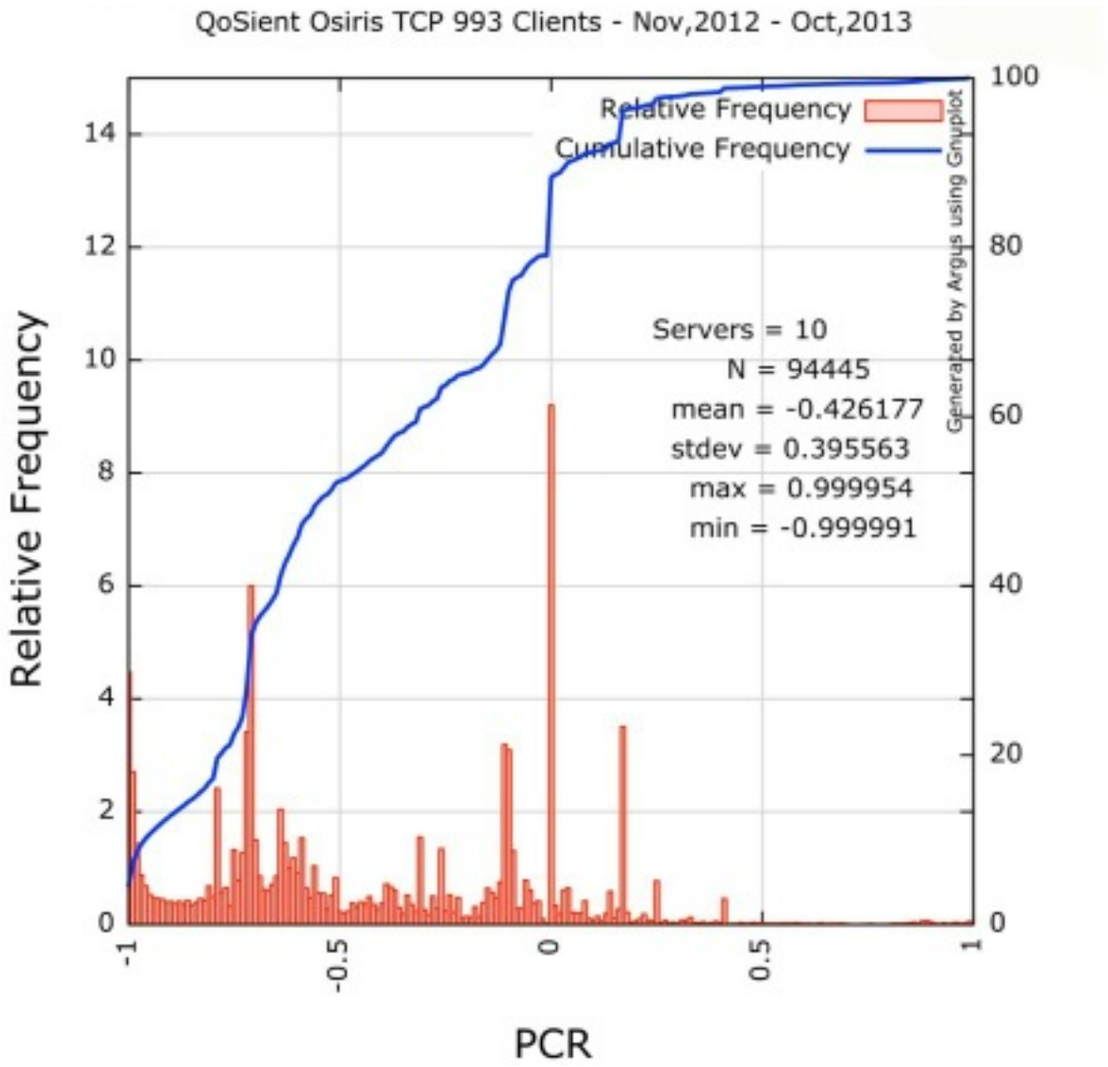


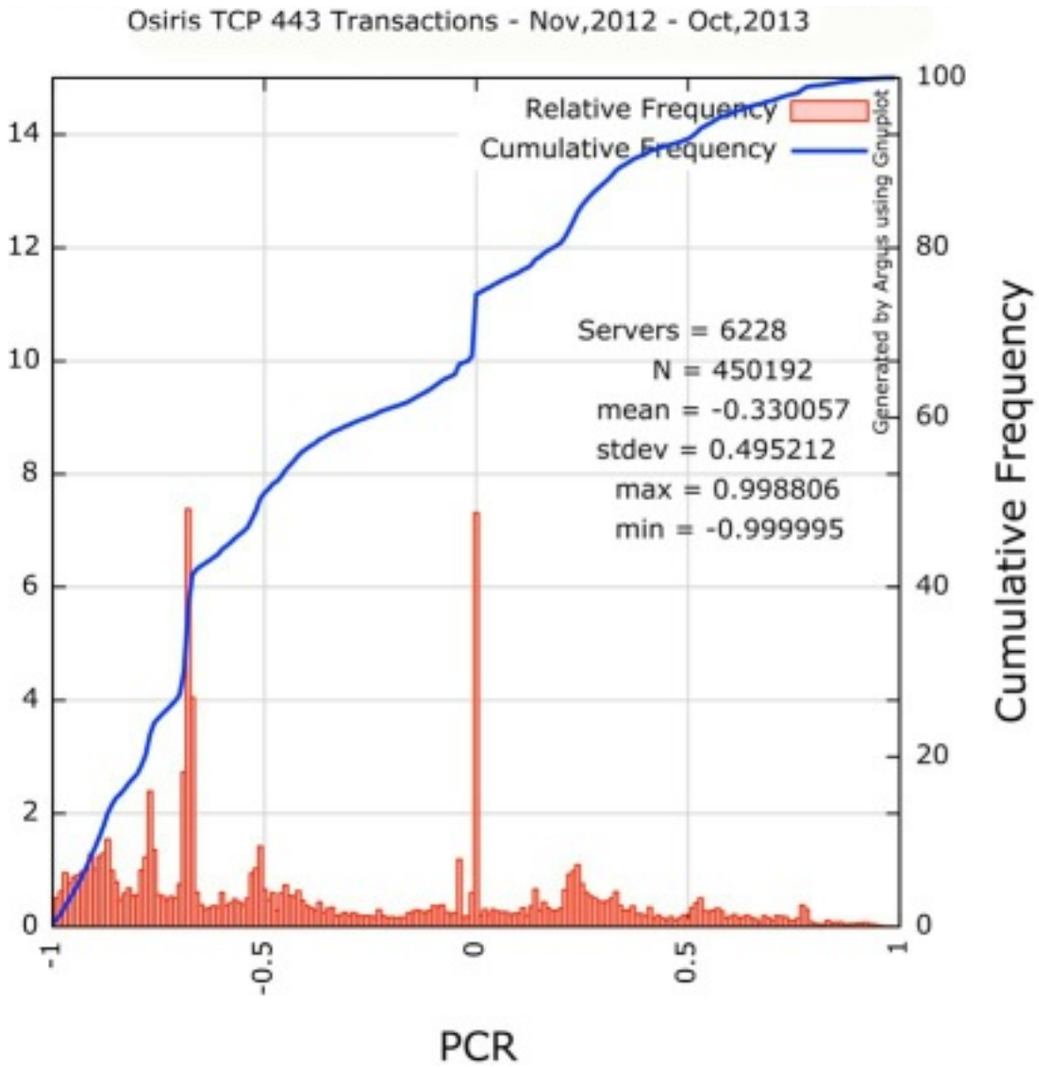
Positive Aggregate PCR caused solely by iCould awacsd - apple wide area connectivity service daemon

Baseline Enterprise PCR - Aggregate PCR value after removing awacsd flow metrics

PCR Application Characterization

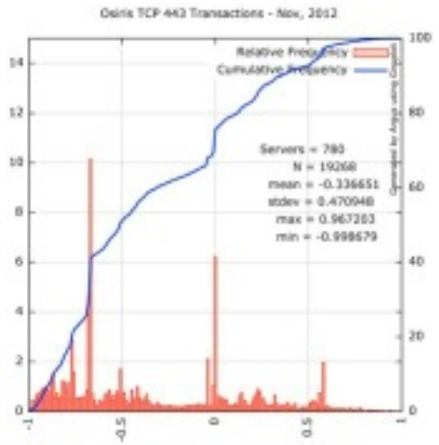
IMAPS, HTTPS

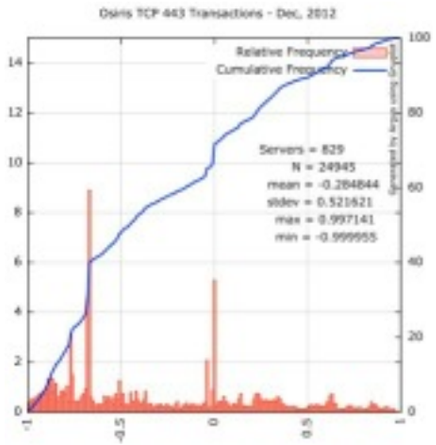


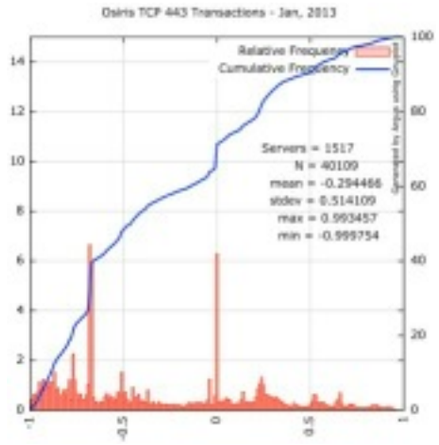


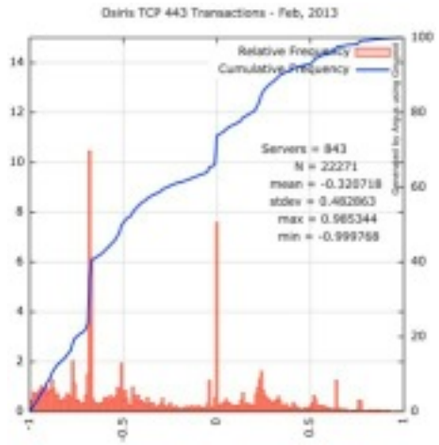
PCR Application Stability

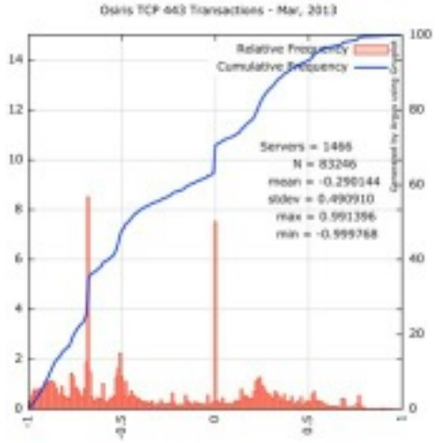
HTTPS - One Year Period

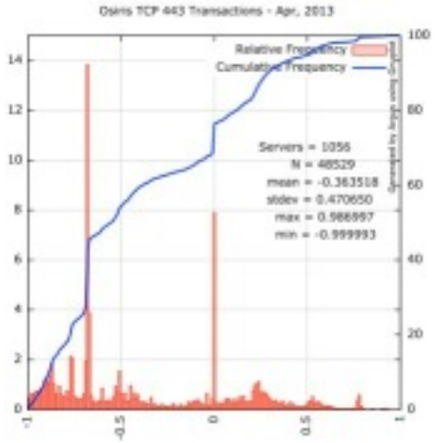


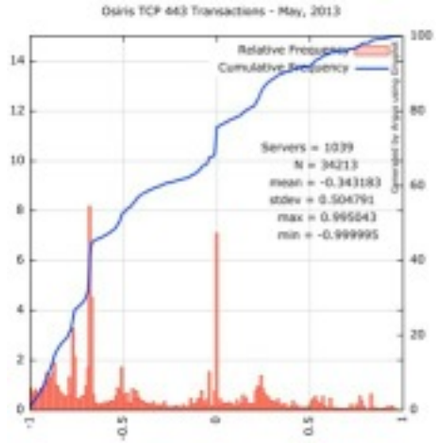


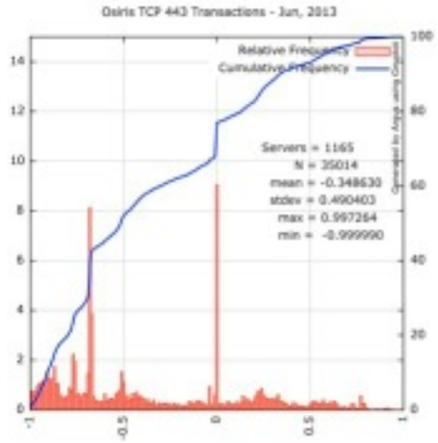


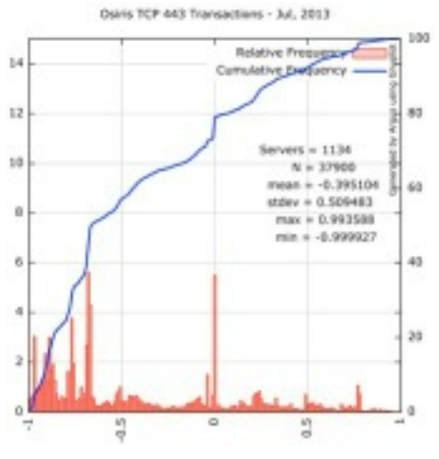


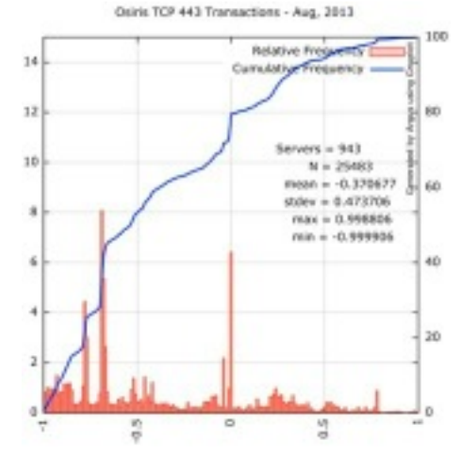


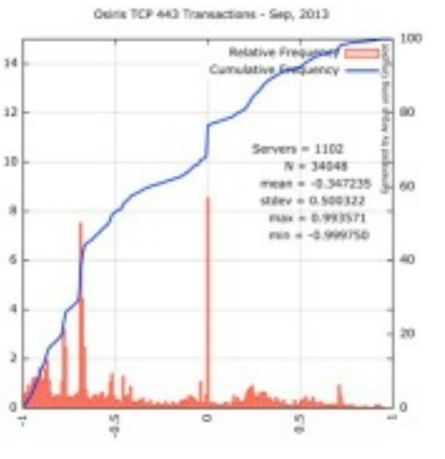


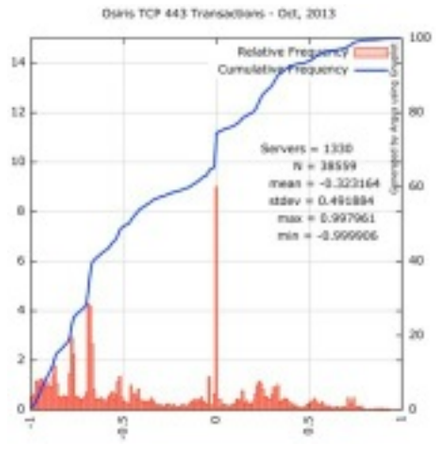






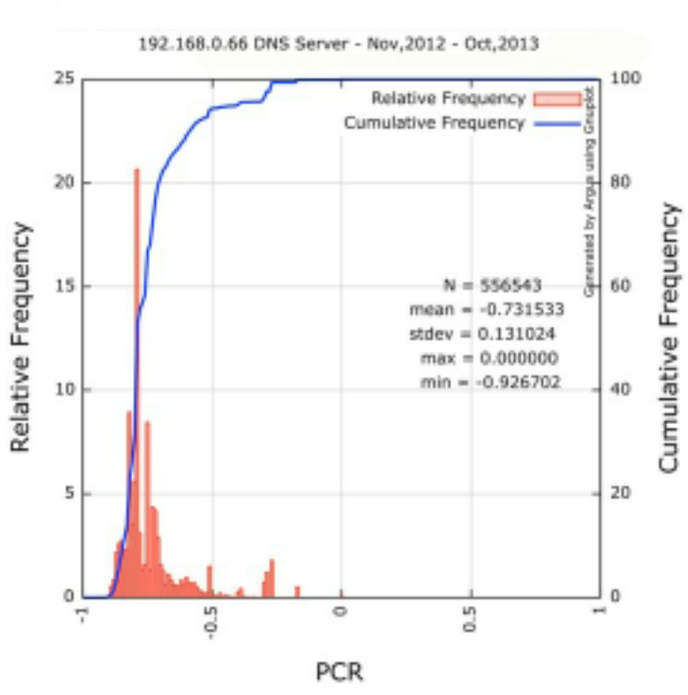


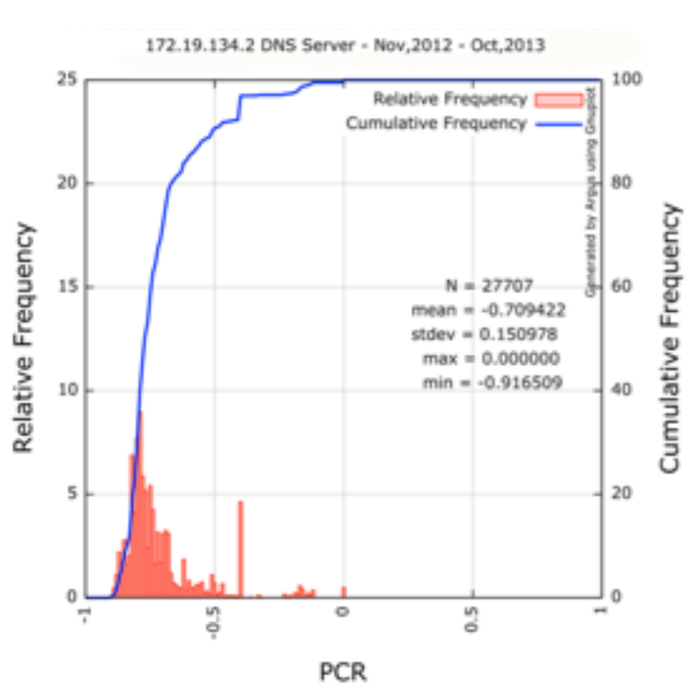


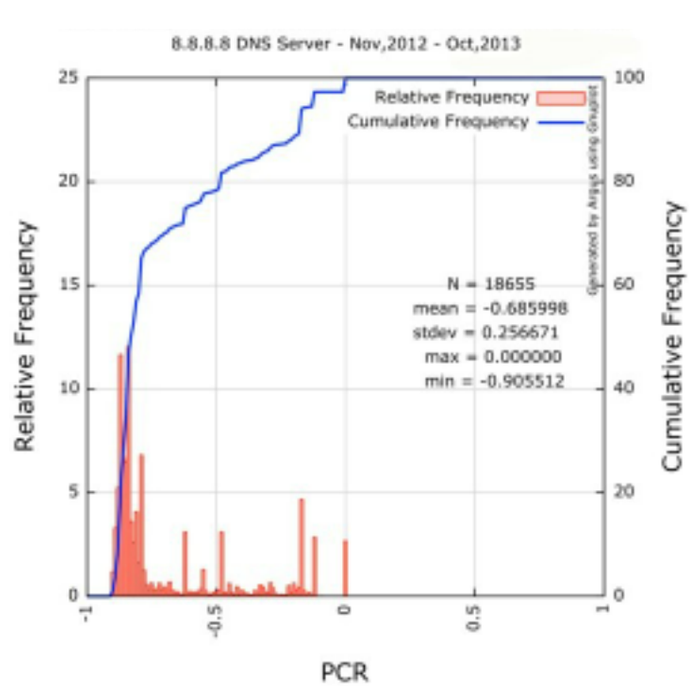


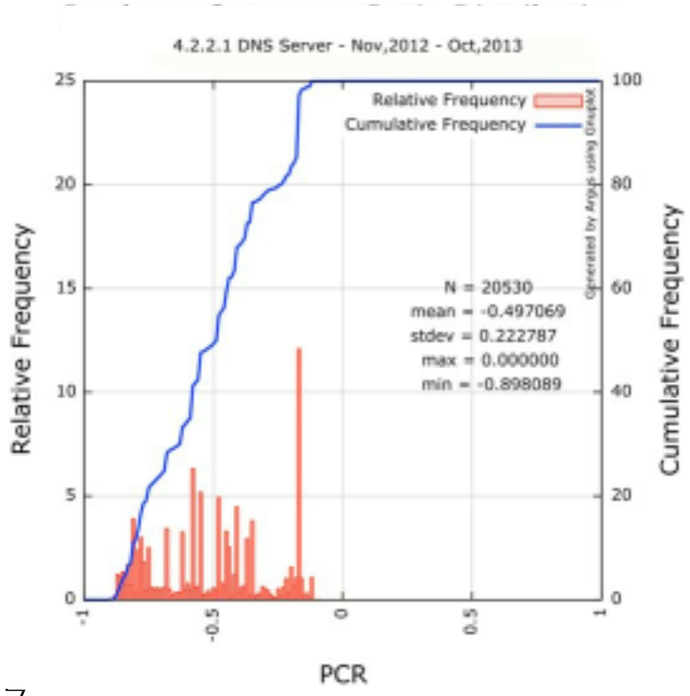
PCR Application Characterization

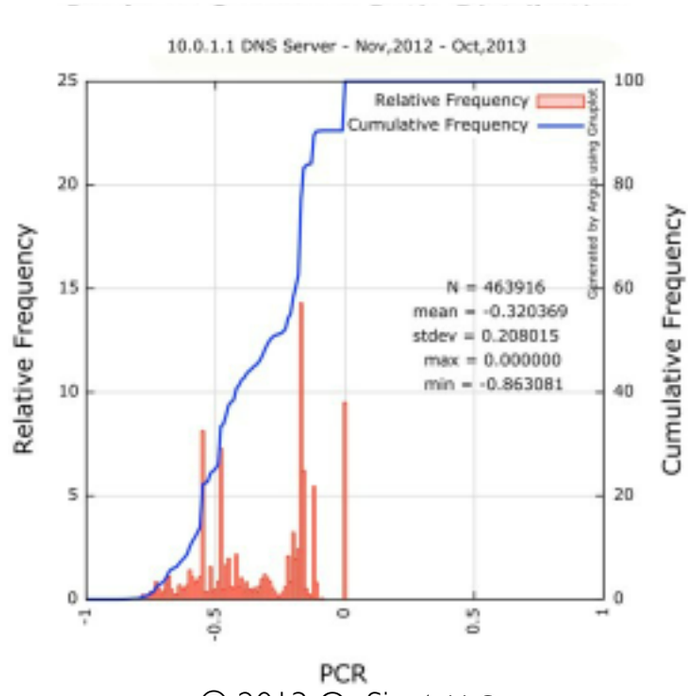
Domain Name Servers





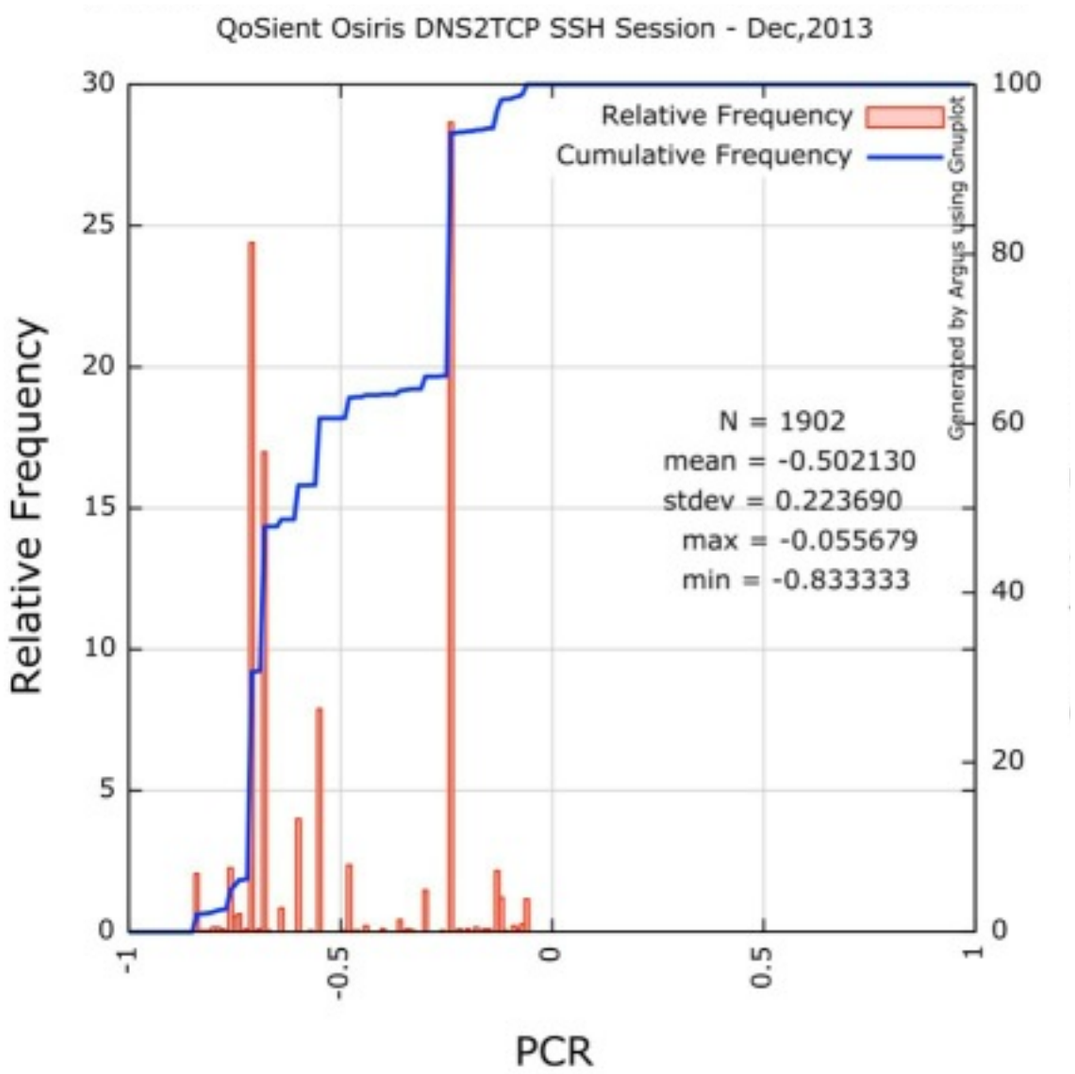


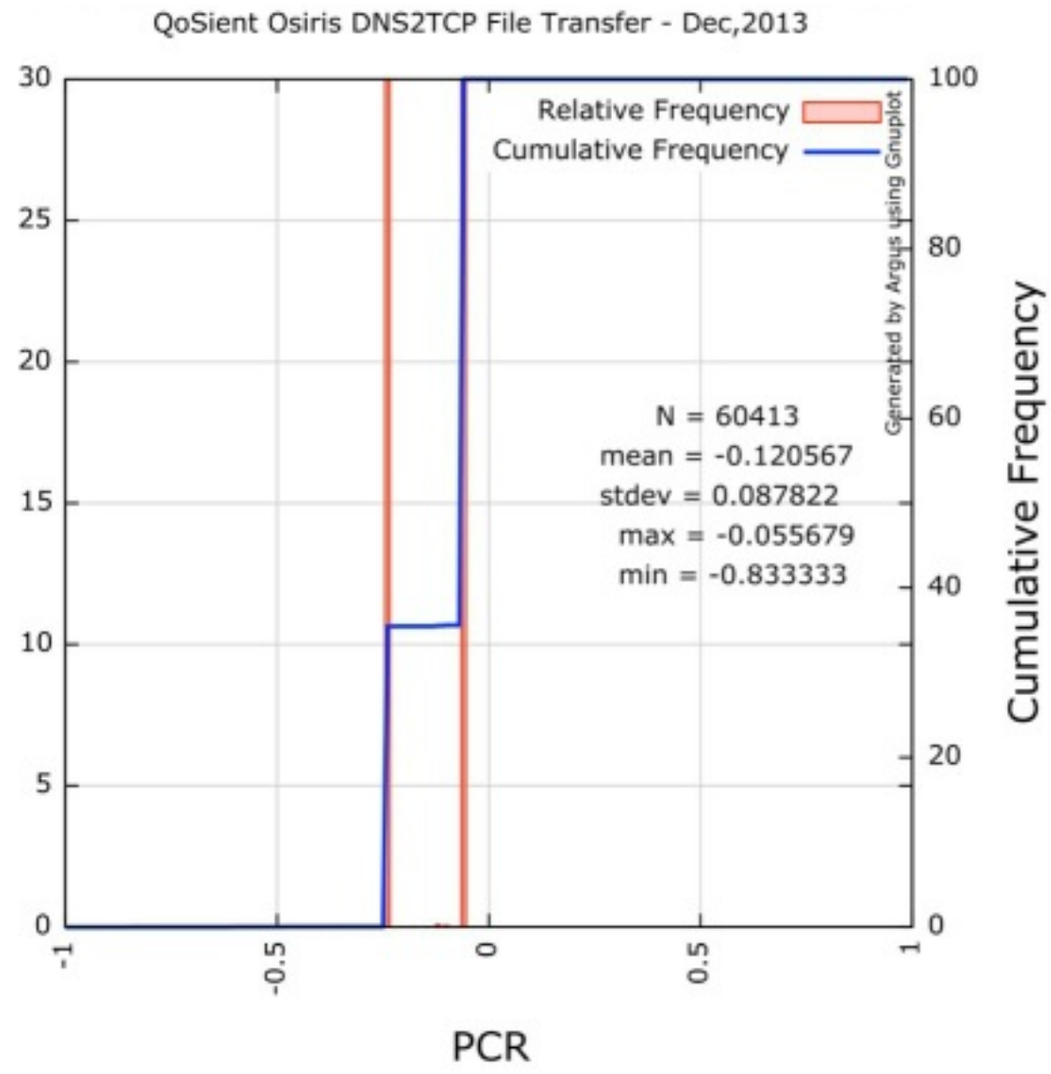




PCR Covert Channel Analysis

Domain Name Service - dns2tcp





Farnham, G. and Atlasis, A. *Detecting DNS Tunneling,* SANS Institute InfoSec Reading Room, Feb, 2013.
https://www.sans.org/reading-room/whitepapers/dns/detecting-dns-tunneling-34152

Supporting Slides

PCR Stability Characterization

PCR vs Total Application Bytes