| galois |

I Want Your Flows To Be Lies

Adam Wick

FloCon 2017

Your Network + Zombies

Someone Objects!

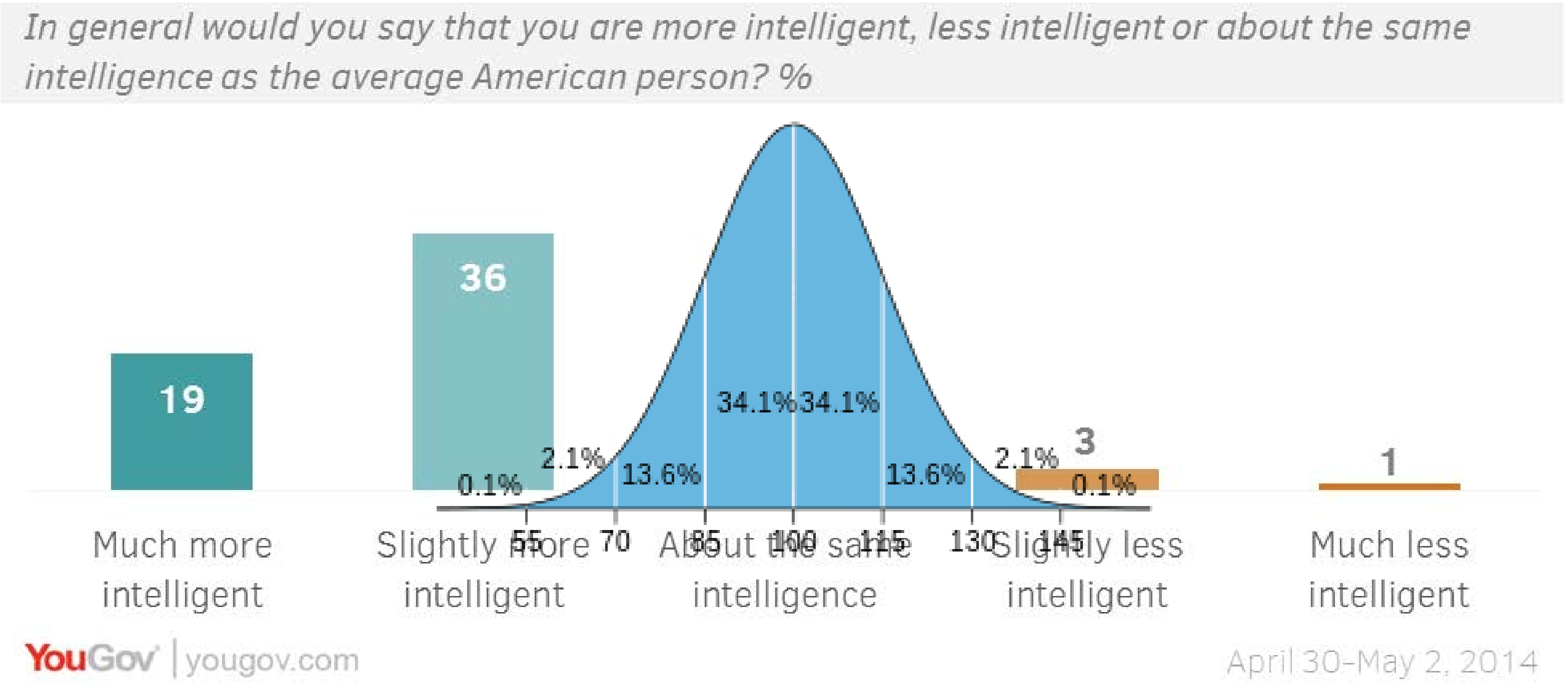
You may be thinking to yourself:

> “Self, these sort of attacks only work on the unwary, unprepared, and foolish. ***I am a FloCon attendee***. I am smart, attractive, fully-versed in the latest in security best practices, and fully capable of protecting my network.
>
> Any anyways, if there was someone bad on my network, I would know.”

To which I say … good luck.

Statistics, Part One: Illusory Superiority

A cognitive bias whereby individuals overestimate their own qualities and abilities, relative to others. Also known as: the above-average effect, superiority bias, leniency error, sense of relative superiority, the *primus inter pares* effect, and the Lake Wobegon effect.



In general would you say that you are more intelligent, less intelligent or about the same intelligence as the average American person? %

YouGov | yougov.com

April 30-May 2, 2014

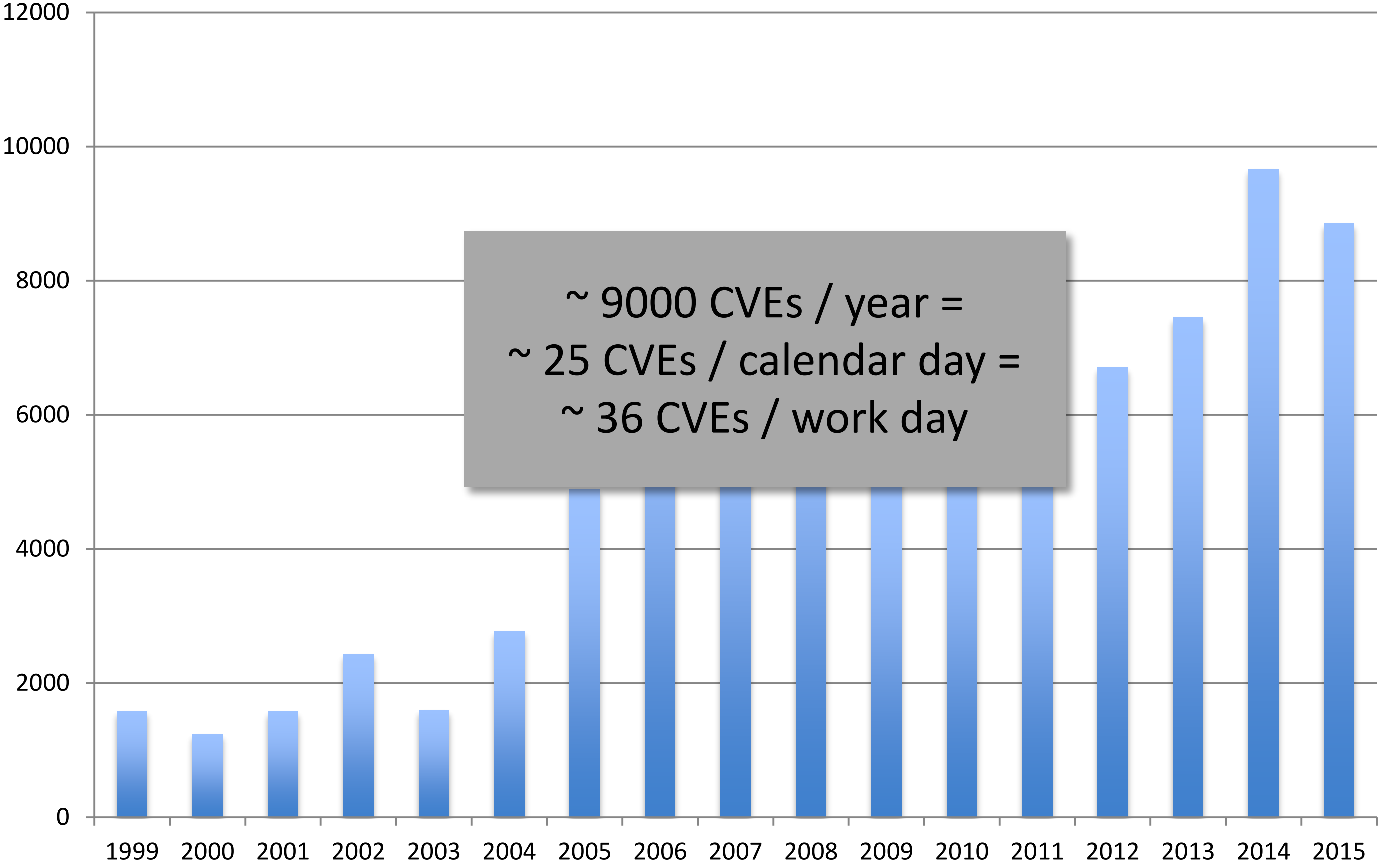
Statistics, Part One: Illusory Superiority

My experience tells me that this also extends to your social and professional grouping.

"Other companies / IT organizations suffer from X, but we're better than average!"

One way to combat this is to remember to tell yourself that you're not worried about your coworkers making mistakes *in general*, but you're worried about *what might happen when your coworker is super busy, tired, and distracted*.

Statistics, Part Deux: CVE Avalanche



~ 9000 CVEs / year =
~ 25 CVEs / calendar day =
~ 36 CVEs / work day

Statistics, Part Tres: Noticing

According to FireEye Security's annual report, what is the mean time between someone gaining access to your network and you detecting them?

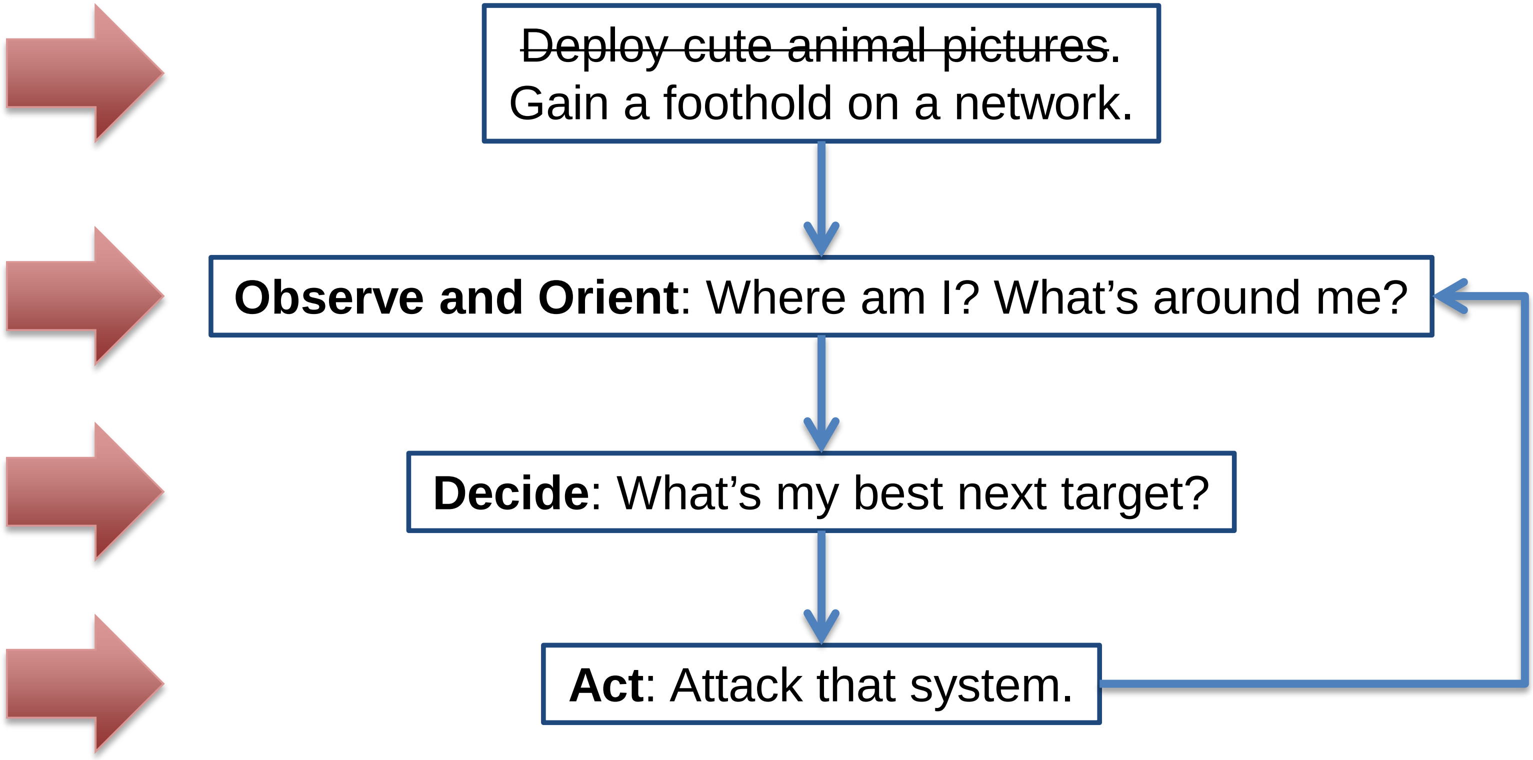
146 days

(or about 4½ months)

So Let’s Get Back To The Point

Let’s Subdivide

Let’s subdivide these steps even further:



~~Deploy cute animal pictures~~.
Gain a foothold on a network.

Observe and Orient: Where am I? What’s around me?

Decide: What’s my best next target?

Act: Attack that system.

Sweet, Sweet Information

So if I’m an attacker that wants to figure out where your most vital systems are, what is the best place for me to look for this information?

Your flow data, of course.

With your flow data, I can quite easily determine:

- Where your most important servers are.
- How your systems interconnect.
- When your staff is on-site.
- When you’re gearing up for a big release.
- Where your honeypots are.

Remember: they have 4½ months to work all this out.

Signal and Noise

The Real Problem: 146 Days of Signal

The real problem is that everything they see in the flow data is truth:

- Traffic patterns identify critical resources
- Traffic patterns identify office hours
- Specific protocols can identify mission applications or goals
- Increased traffic can identify changes in operational tempo
- Traffic from a particular workstation can be used to identify someone’s presence
- A lack of traffic can identify unimportant or trap servers

The critical thing, though, is that all they see is signal.

Prattle: Adding Noise to The Signal

The Prattle project was created to address this problem: to add noise to the signal.

At a high level, our goal is to:

- Level traffic patterns to obfuscate:
  - Which machines are most critical
  - Operational tempo
- Provide false data, including credentials, to mask real activity
- Direct adversaries to honeypots and other traps

Noise-Canceling Headphones, Inverted

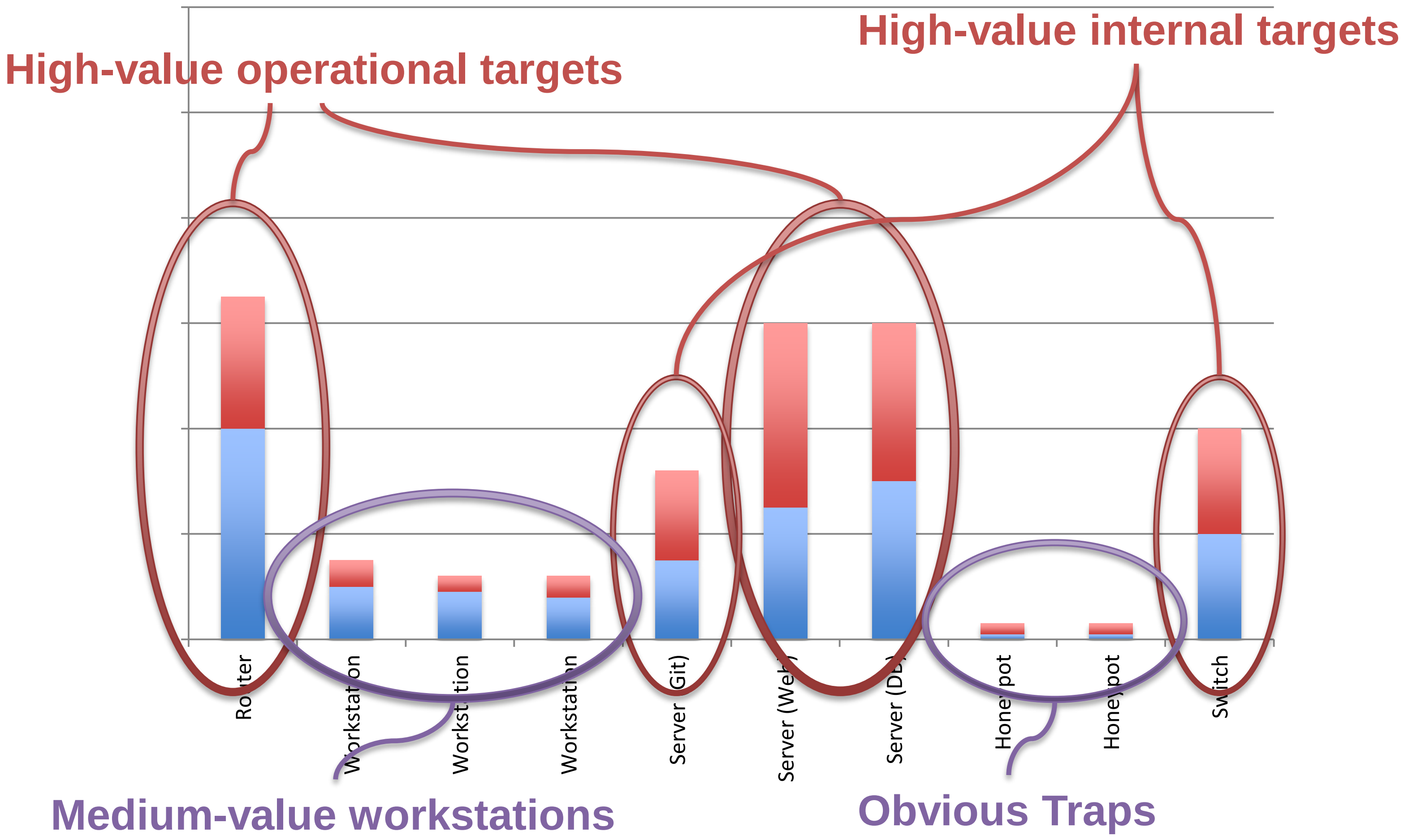
My Simple Network

Let’s say I have a ten node network with

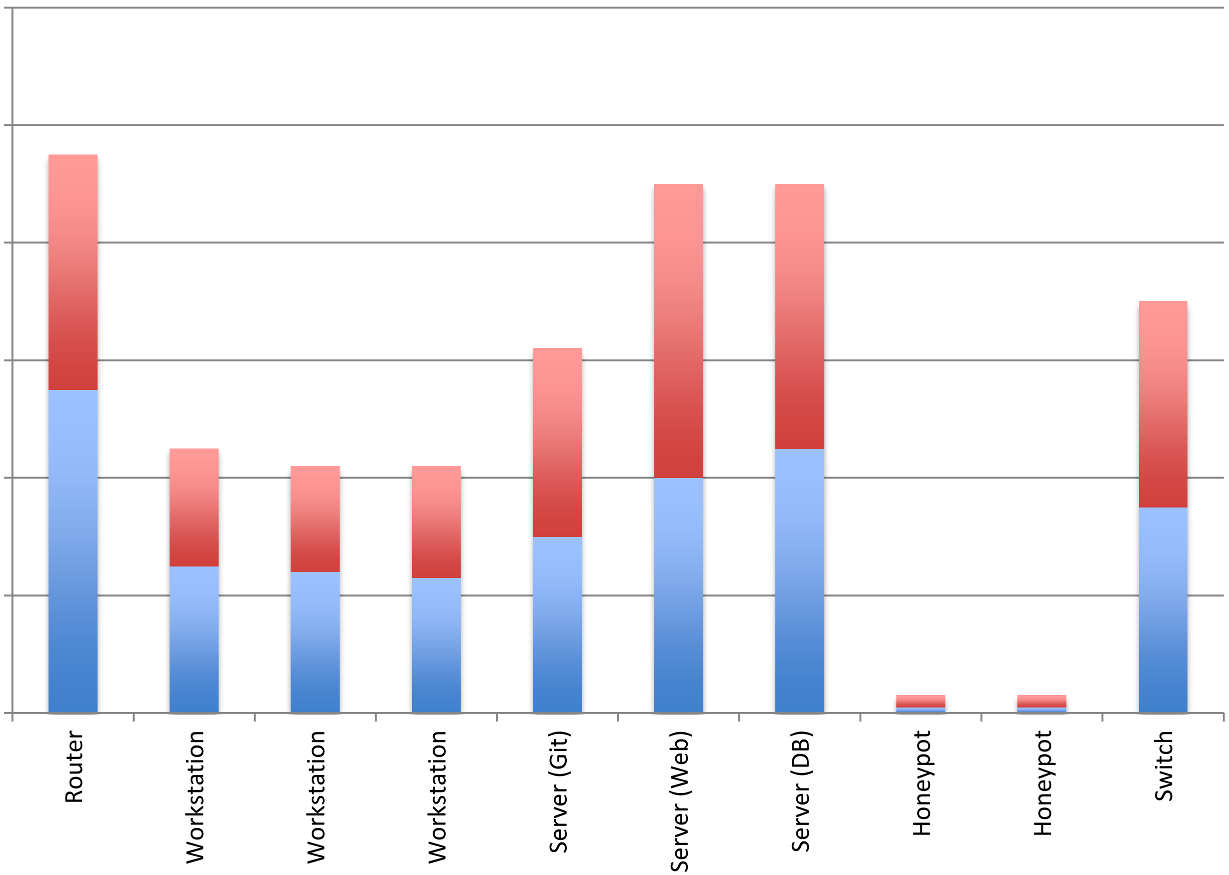
- A router
- Three workstations
- Three servers: git, web, and DB
- Two honeypots
- Switch

What happens if I grab a snapshot of traffic over some period of time, and graph it?

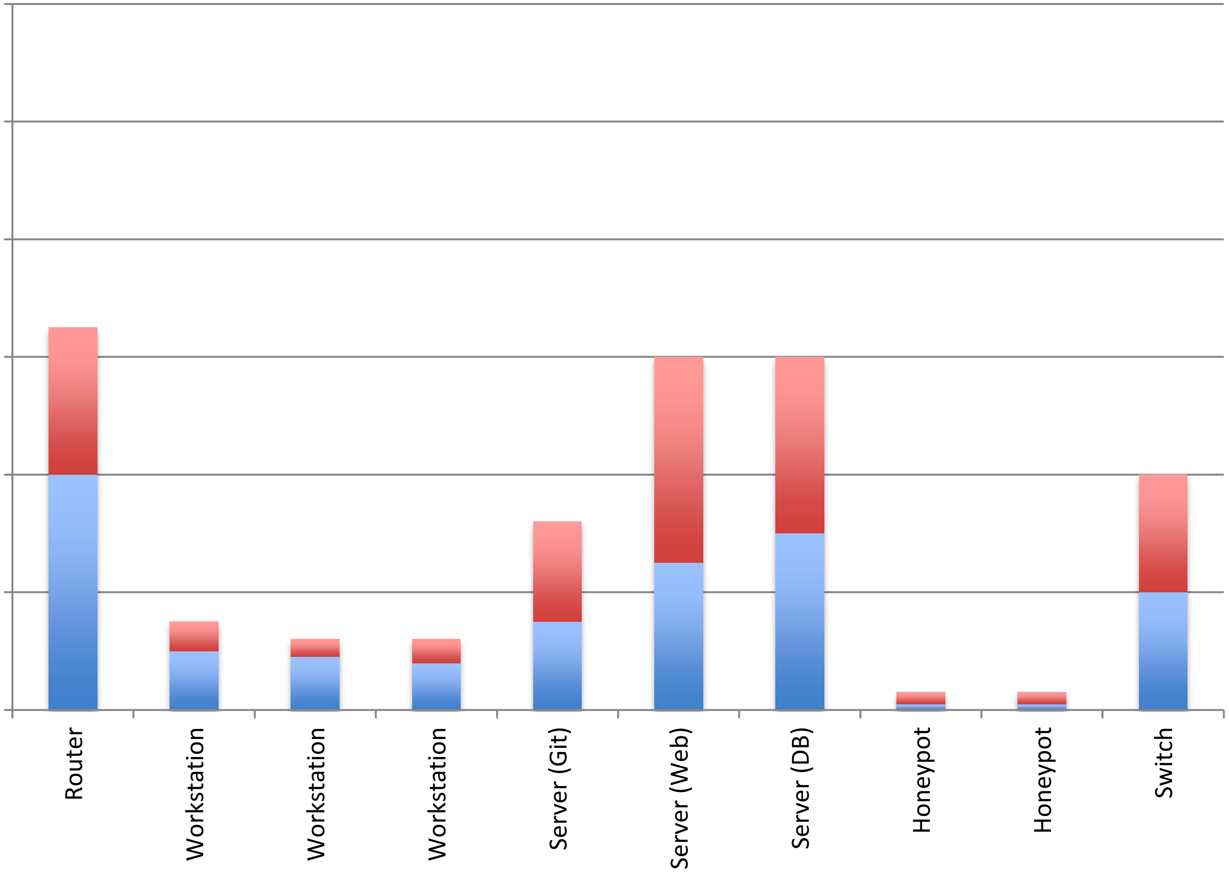
Truth, and Its Inconvenience



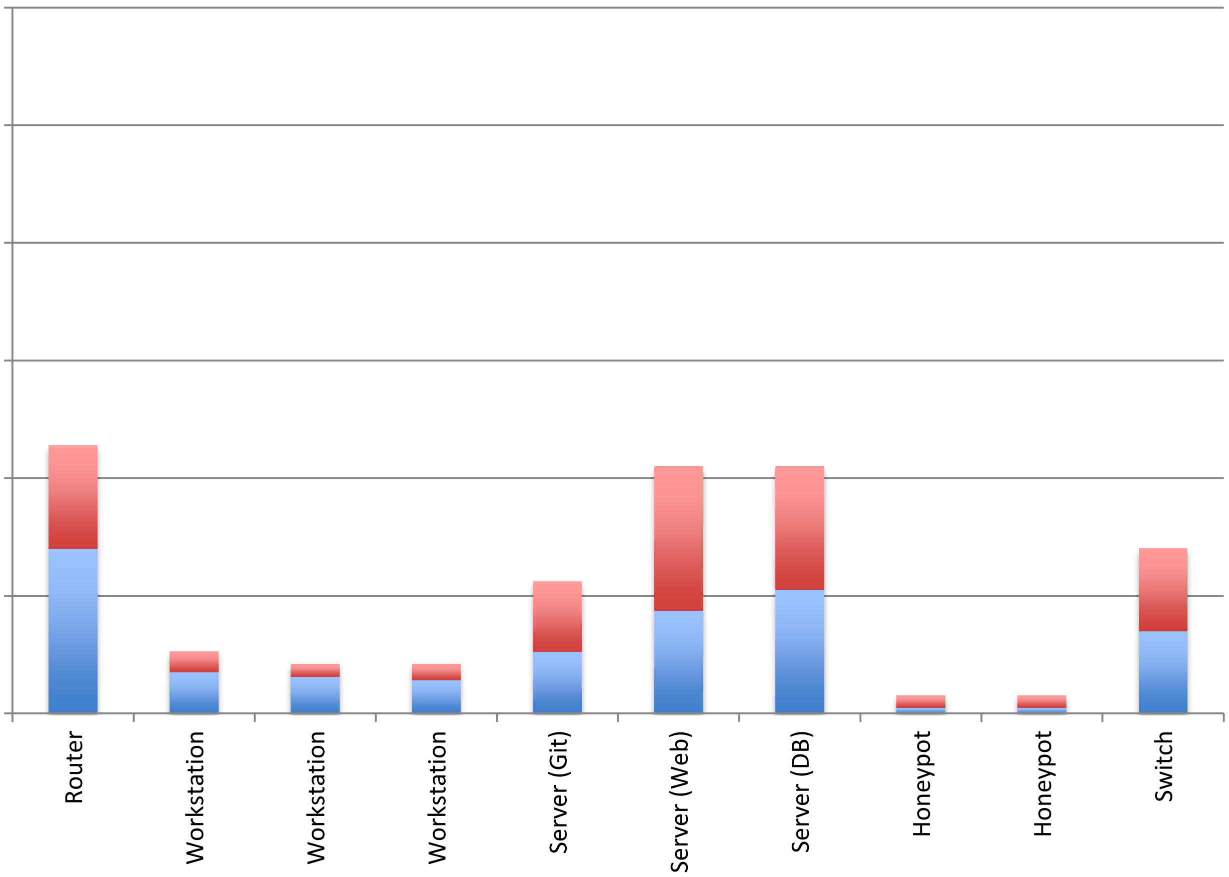
Truth, and Its Inconvenience



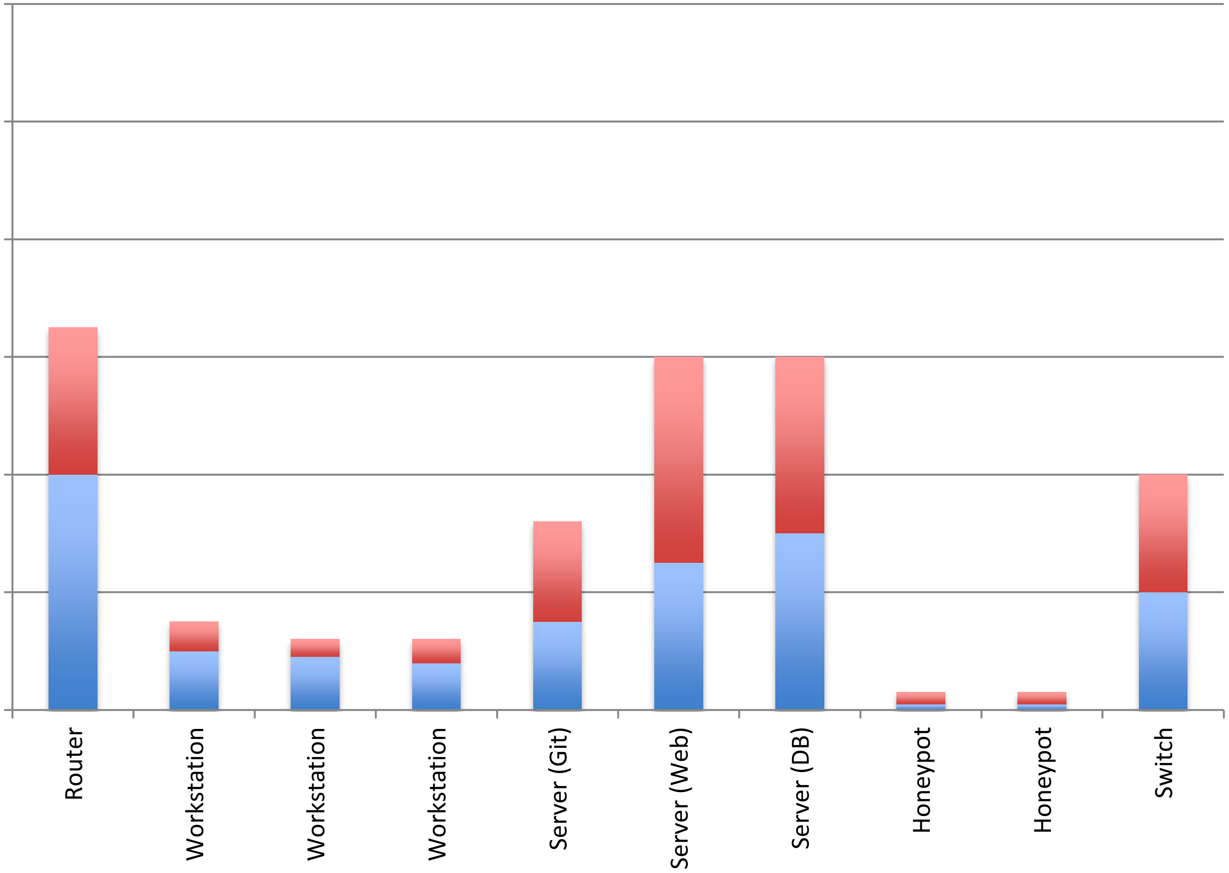
Truth, and Its Inconvenience



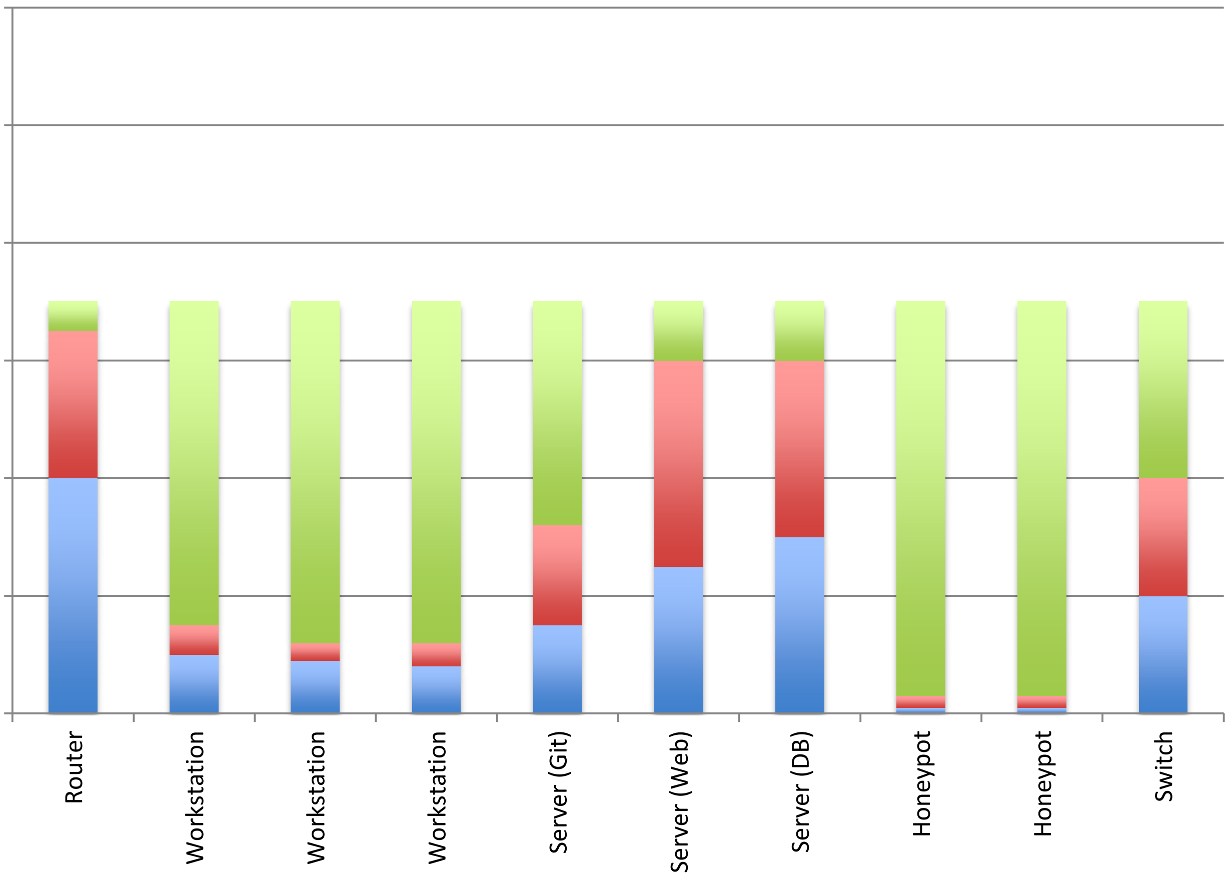
Truth, and Its Inconvenience



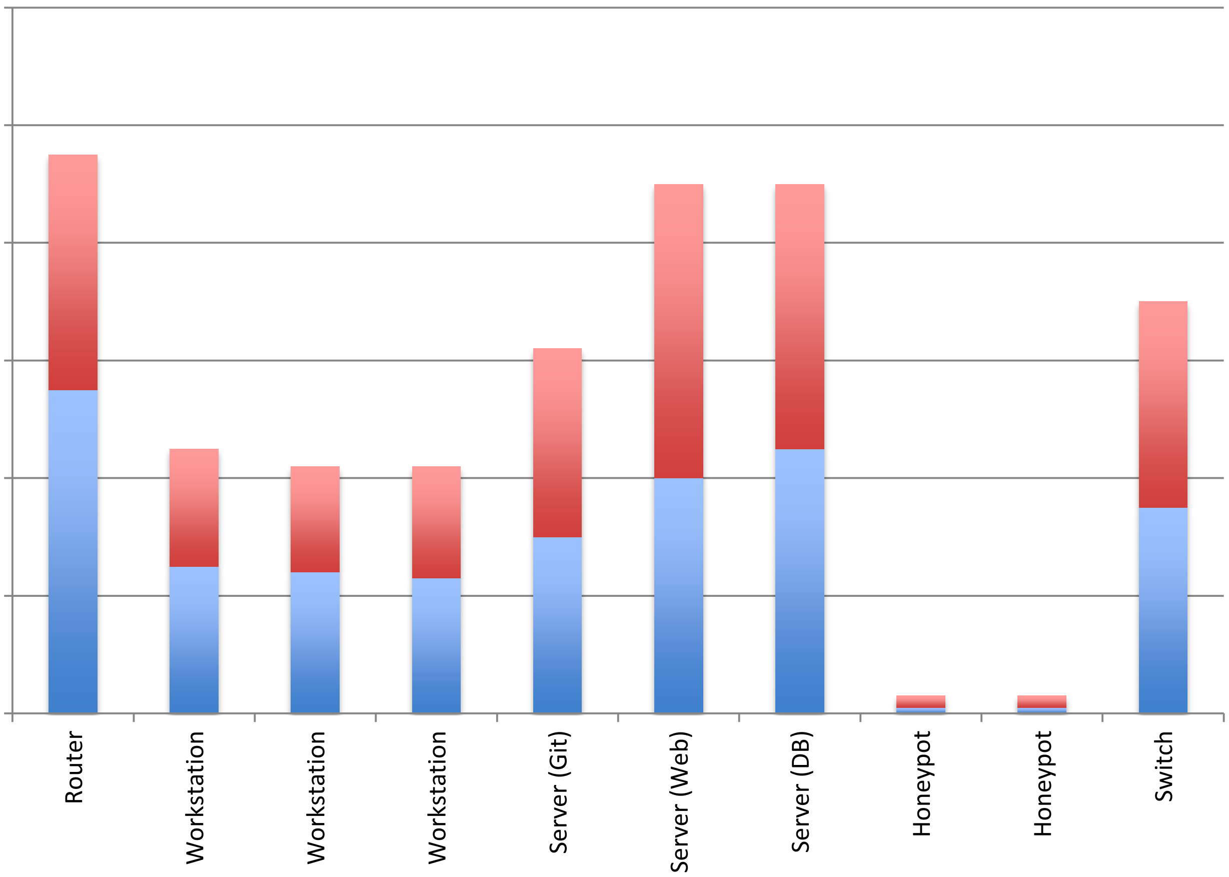
Lying for Good



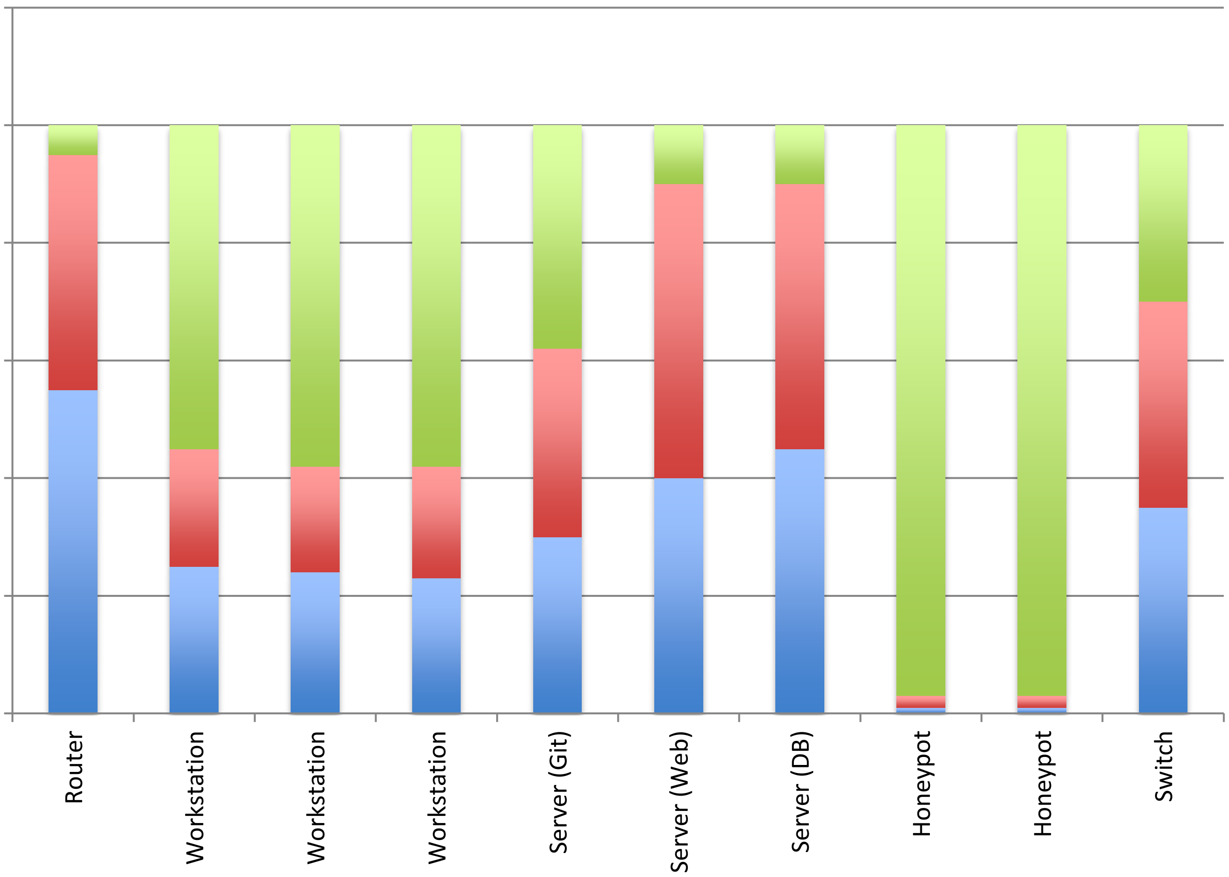
Lying for Good



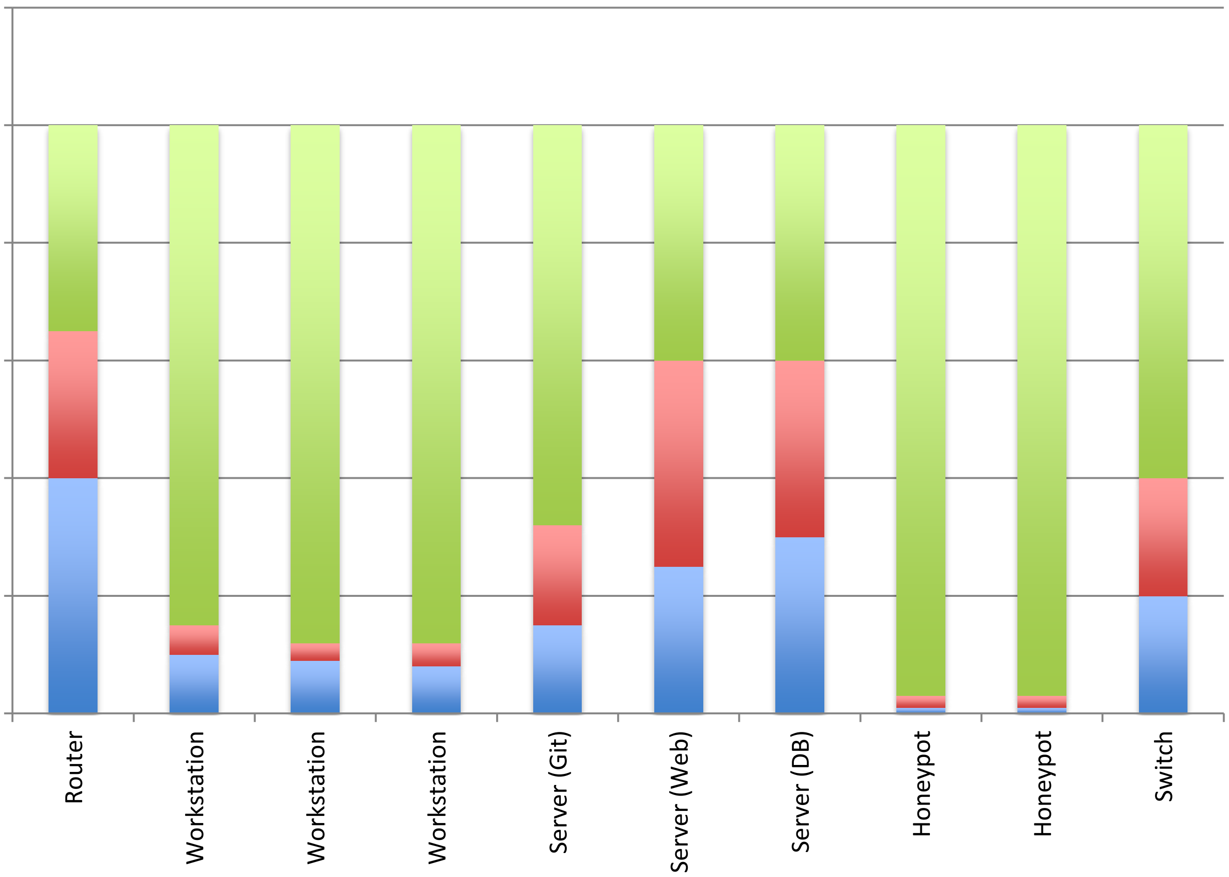
Lying for Good



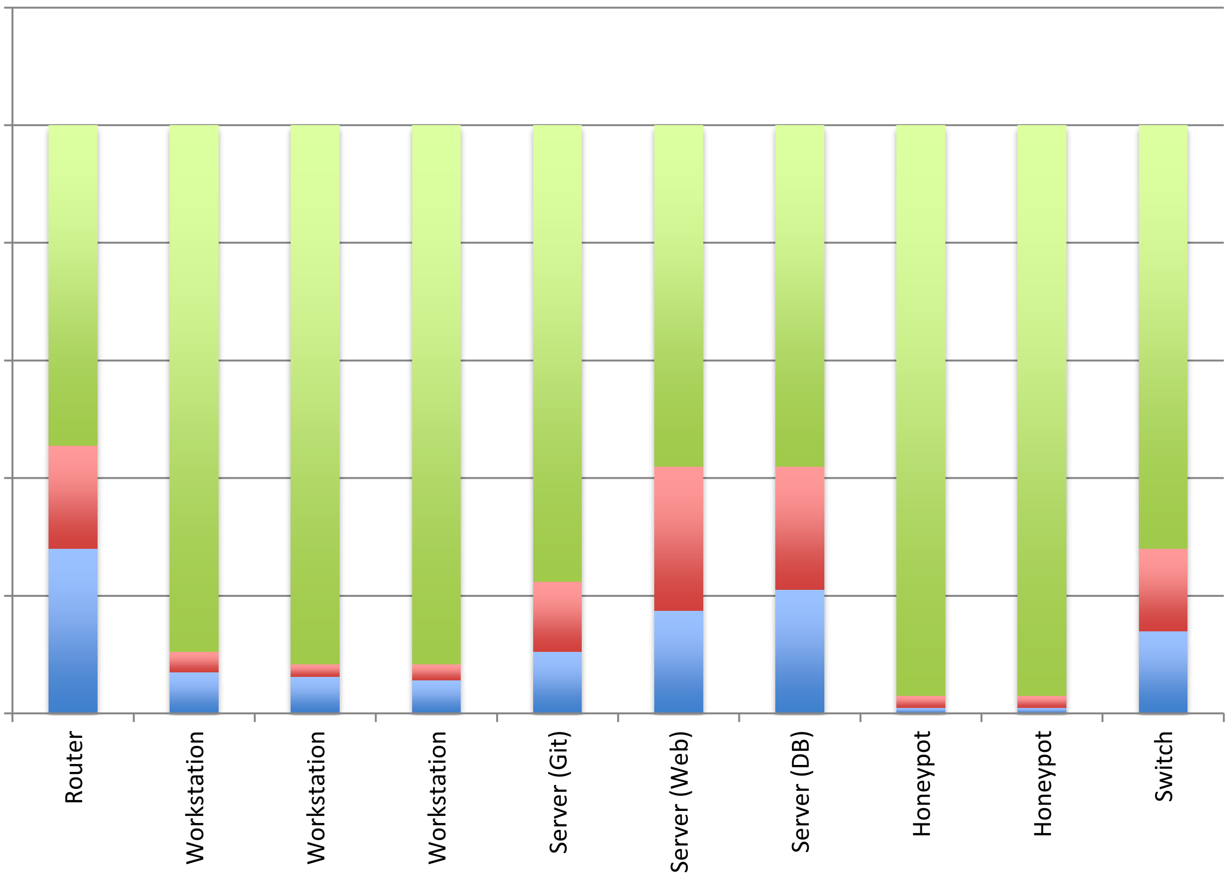
Truth, and its Inconvenience



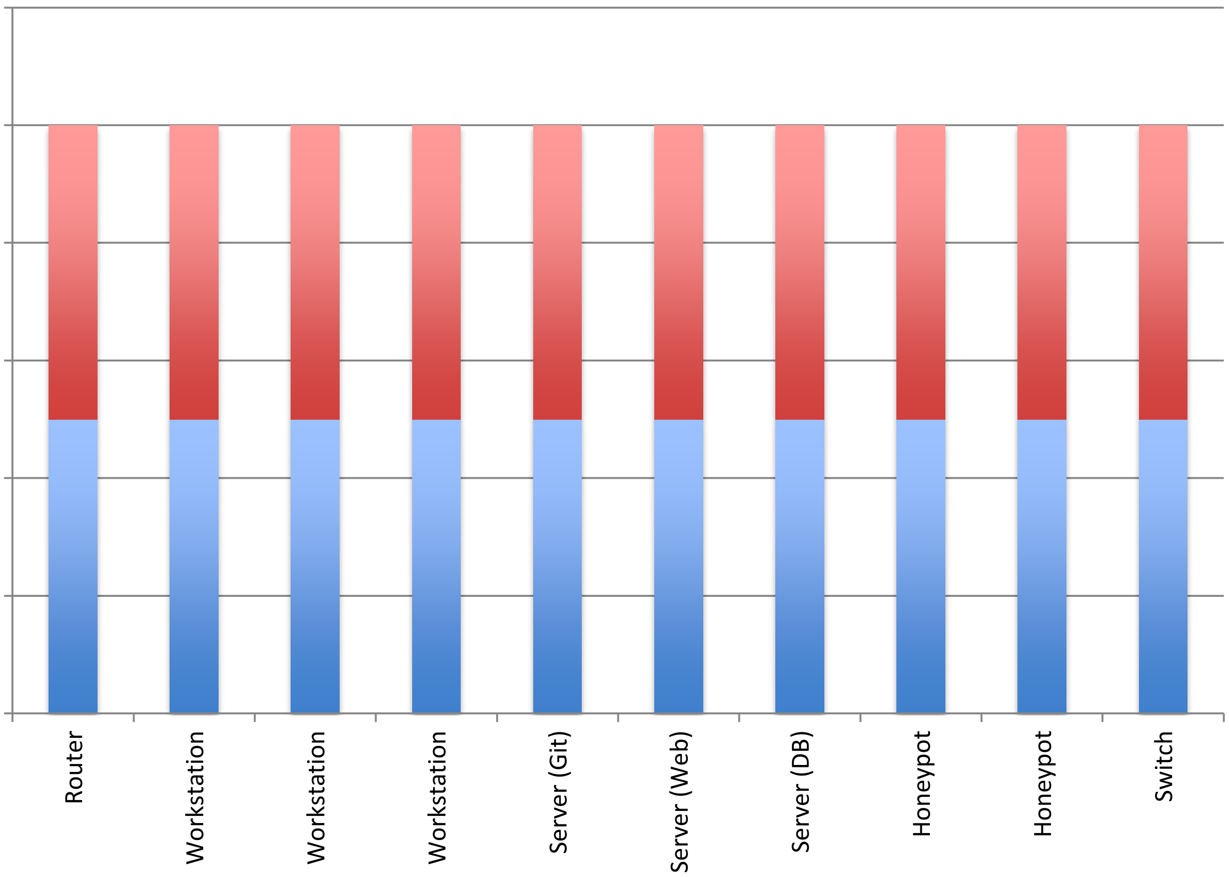
Truth, and its Inconvenience



Truth, and its Inconvenience



Truth, and its Inconvenience



The Caveats

The Four Challenges of Prattle

1. **How do we make sure that our noise isn't easily detectable by an adversary?**
2. **How do we make sure that our noise is easily detectable by an admin?**
3. **How do we deal with too much data? (Version 1)**
4. **How do we deal with too much data? (Version 2)**

Challenge #1: Real Good Fakery

Unlike other traffic generators, Prattle must emit *traffic that is indistinguishable from real traffic* to a relatively knowledgeable adversary.

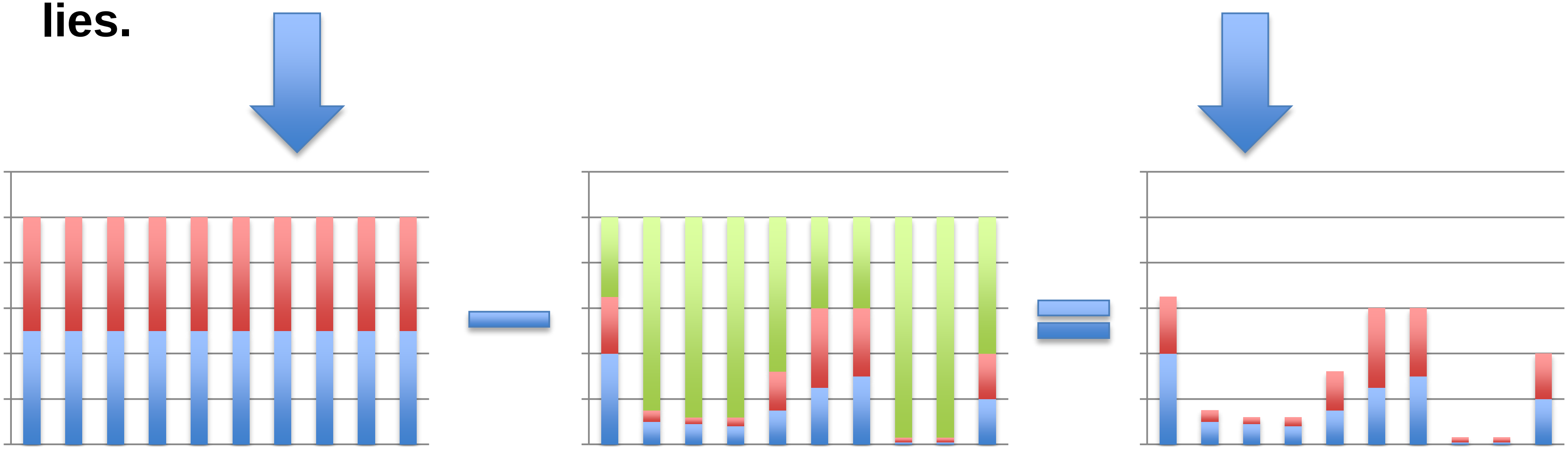
This is actually the main research problem of the project, and we’ve had some success:

- Prattle can generate browsing sessions that are very hard to distinguish from real users.
- We can generate encrypted traffic (via SSL or SSH) that generate packet sizes and timings that are statistically identical to real workflows.
- We are adding more and more protocols all the time, to mimic real workstations.

Challenge #2: Subtraction

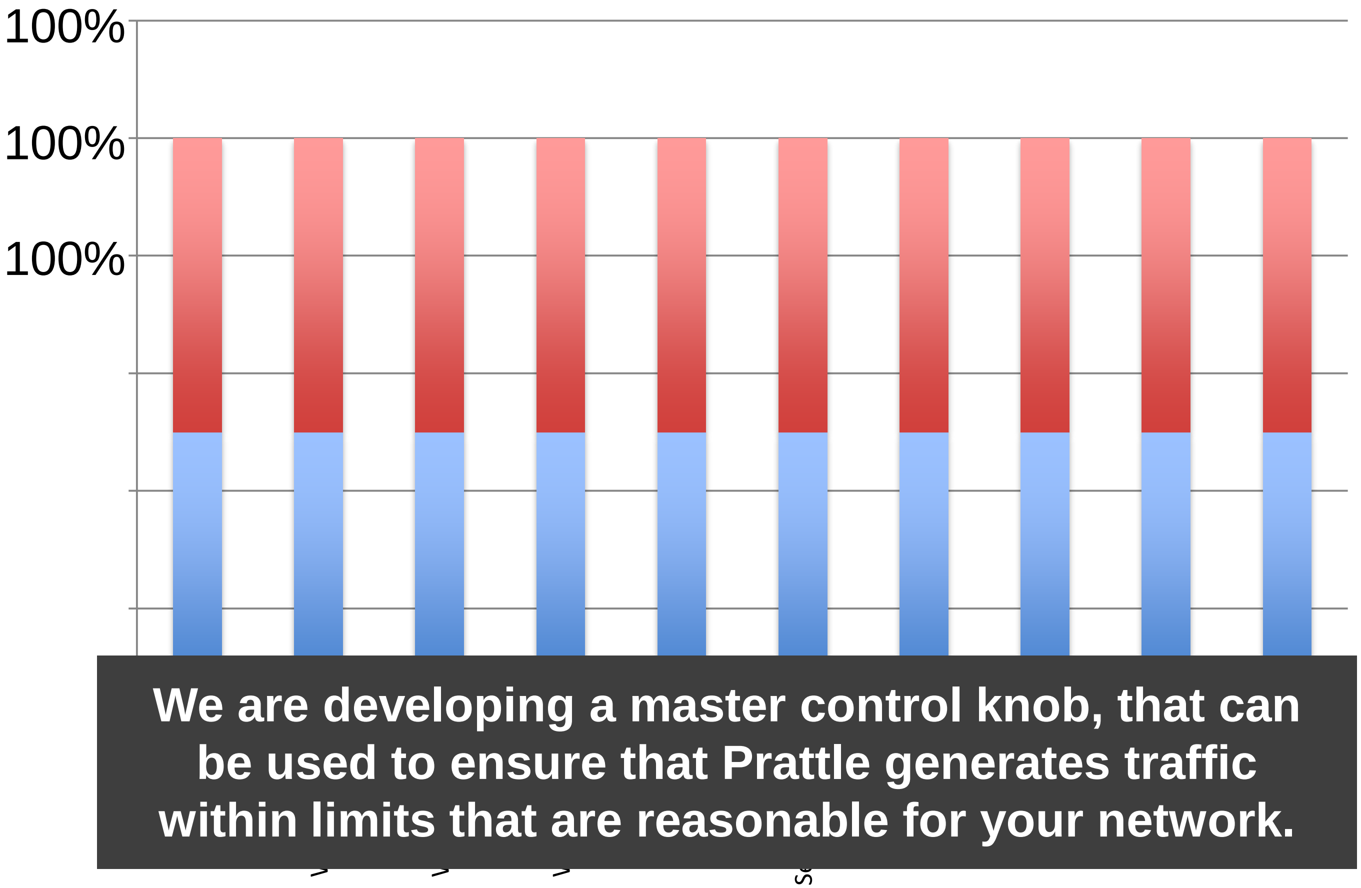
Me: I want your data to be lies.

You: I want my data to be useful!



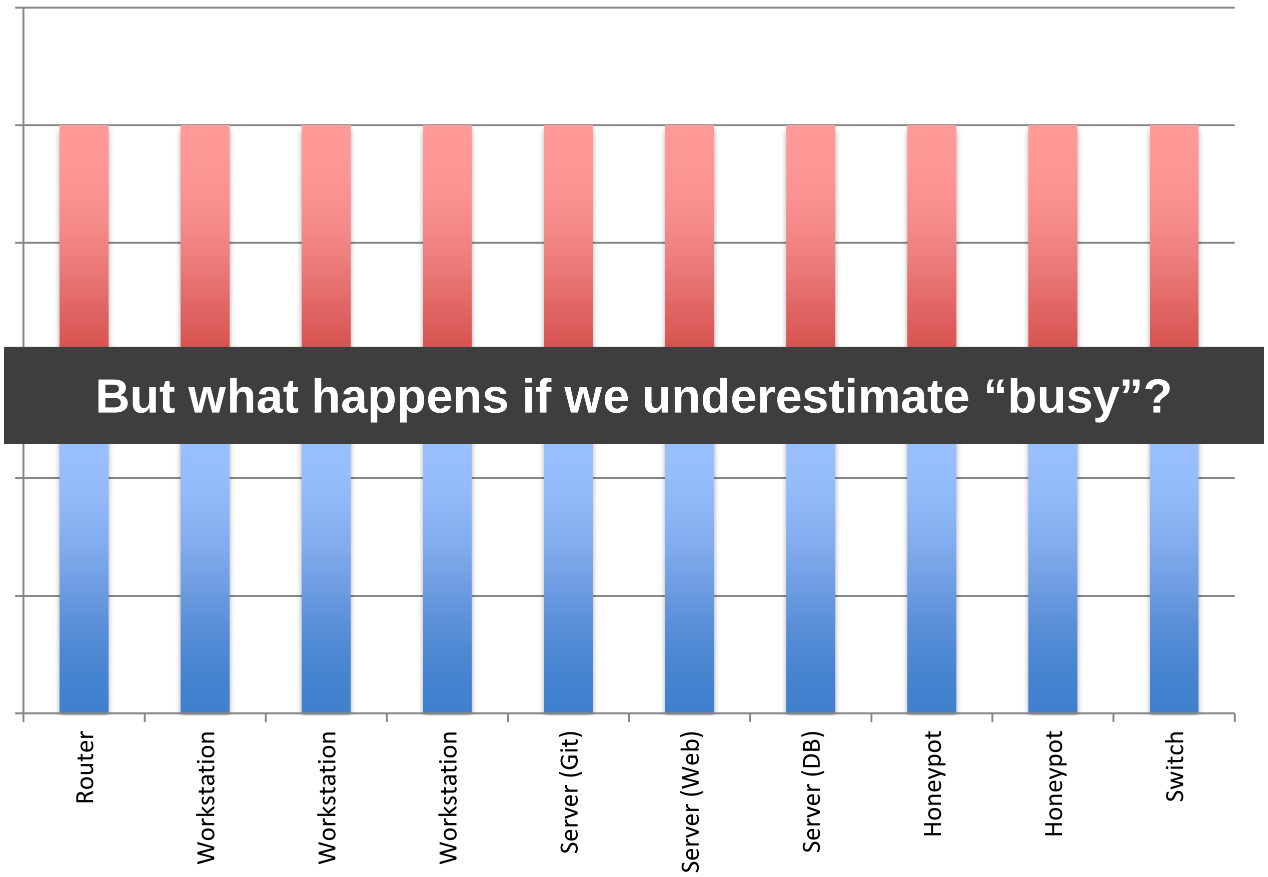
Prattle can generate flow databases that contain only generated traffic, which can then be used to rediscover the real traffic patterns

Challenge #3: Too Much Data (Version 1)



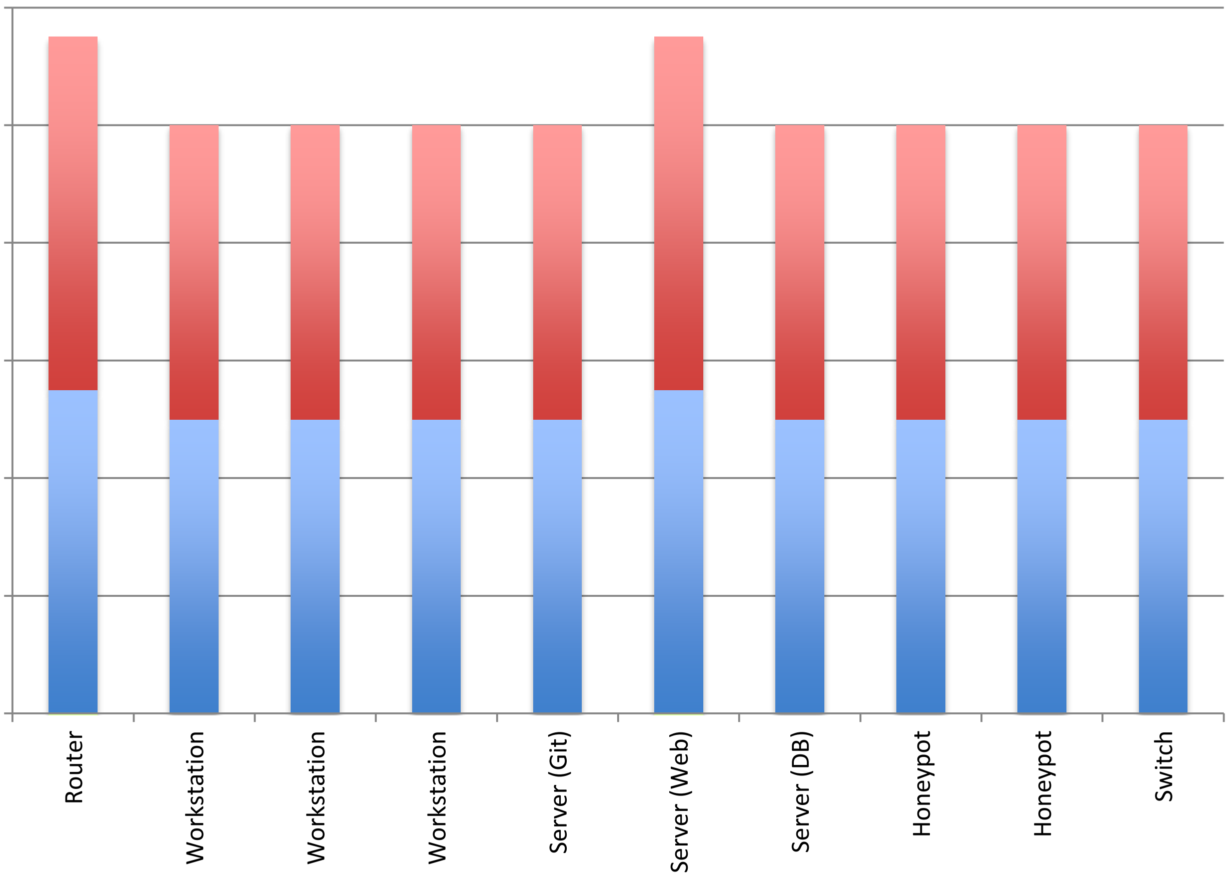
We are developing a master control knob, that can be used to ensure that Prattle generates traffic within limits that are reasonable for your network.

Challenge #4: Too Much Data (Version 2)

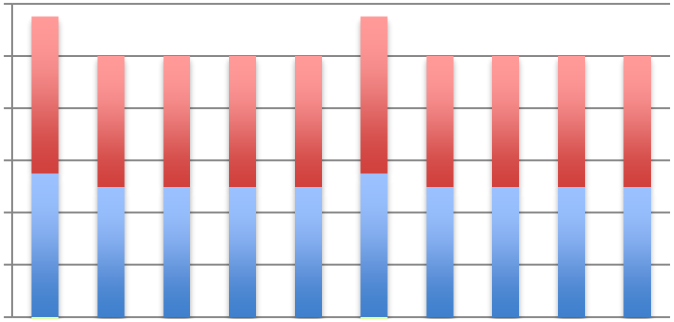


But what happens if we underestimate “busy”?

Challenge #4: Too Much Data (Version 2)



Challenge #4: Too Much Data (Version 2)



This isn’t the end of the world.

You are still:

- Hiding your honeypots
- Obfuscating your employee’s work
- Masking the true traffic patterns of even the busy server

In fact, one of our roadmap items is to use the new peak to set the new maximum level, so everything goes up at once (to a user-defined maximum).

In Summary

I Want Your Data To Be Lies

- **Flow data is incredibly useful**
  - It’s useful to you to understand and monitor your network
  - It’s useful to your adversaries to understand your network
- **Let’s add noise!**
  - Generate traffic to mask your critical systems
  - Generate traffic to mask your operational tempo
  - Generate traffic to make your defenses more effective
- **At the same time, let’s be sure to be good citizens.**
  - Keep the bandwidth within reason.

I Want To Lie To *Your* Network

Prattle is under development as part of an Air Force Research Project.

We are working on it *right now*.

However, we'd love to start piloting Prattle on real systems, and commercial systems, so that we can make sure that the technology is useful, practical, and effective.

In other words: pilot partners wanted!

If you'd be interested in trying it out, or have any requirements you think would be important,
please reach out!

Contact Information

Adam Wick
Research Lead / Prattle PI
awick@galois.com
@acwpdx

All trademarks, service marks, trade names, trade dress, product names, and logos appearing in these slides are the property of their respective owners, including in some instances Galois, Inc. This project is sponsored by the Air Force Research Laboratory (AFRL) under contract number FA8750-16-C-0305. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force.