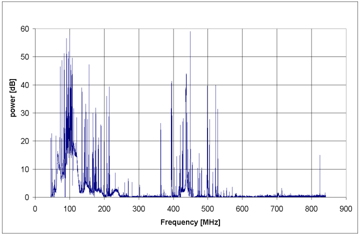
Activity Analysis in Packet-Based Wireless Networks

Shridatt Sugrim, Predrag Spasojevic, Melike Baykal Gursoy

June 12, 2013

Table of Contents

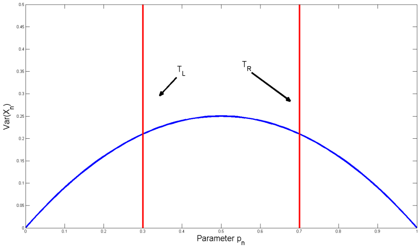
The Problem



- Imagine $N = 256$ channels at a location
- Some are full, but some may be usable
- Suppose I'm driving by in my measurement van and because I'm moving, I'll only get $L = 1000$ measurements before I pass through
- And I can only measure 1 channel per measurement interval
- How do I find the usable ones before I leave the area?
- Can I also find some really bad ones to label as unusable

Variance and Composite Hypothesis

Assume that the activity X_n of a single channel n is modeled by a Bernoulli RV. The variance of a single sample looks like:



- As the parameter gets closer to either end of the interval $[0, 1]$ the variance monotonically decreases
- We'd like to constrain our attention to channels that have either

  $p_n > T_L$ or $T_R < p_n$
- The low variance will allow us to classify their occupancy (p_n) values with fewer samples

Variance and the Naive estimator

The variance of the m_n^{th} sample is a function of the occupancy parameter p_n. It's given by $Var(X_{m_n,n}) = p_n * (1 - p_n)$. Consider a naive estimator $\hat{p_n}$,

$$\hat{p_n} = \frac{1}{M_n} \sum_{m_n=1}^{M_n} X_{m_n,n} \; ; \; Var(\hat{p_n}) = \frac{Var(X_{m_n,n})}{M_n}$$

Where

- M_n is stopping time for channel n,
- m_n is the local time index for channel n
  (Global time obeys the relationship $j = \sum_n m_n$)
- $X_{m_n,n}$ is the m_n^{th} measurement on channel n

We see that:

- The estimator variance $Var(\hat{p_n})$ is directly proportional to the RV variance $Var(X_n)$
- For a fixed variance target V_{req}, low variance channels will reach this target in few samples (smaller M_n)
- Sequential samples will be more consistent (return a confirming value) for low variance channels

A coin toss Analogy

Suppose we have N coins. Each coin's probability distribution has a constant but unknown parameter p_n and each p_n is drawn from a uniform distribution on $[0, 1]$. We wish to discover the $p_n s$ however we are only allowed a total of L coin flips.

- How should we distribute those flips?
- When should we stop flipping an individual coin?
- If $\frac{L}{N}$ is low, How many coins can we characterize?

Suppose further we are interested in coins that biased for heads or tails (unfair coins). Consider 3 bins that we want to place our coins into. The mostly heads bin, the mostly tails bin, and the close to fair bin.

- How does my strategy change when I want to classify?

Composite Hypothesis and Wald's observation

- Simple Hypothesis: The most simple case would be to compare two fixed values θ_1 and θ_0 from our parameter space Θ

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1$$

- Composite Hypothesis: This simple test is only sufficient for checking two simple values, if we want to classify channels, we'll need a composite hypothesis of the form:

$$H_1 : \theta > T$$
$$H_0 : \theta < T$$

  where T is some meaningful threshold

- Wald[1] argues that for a given composite test, if we pick appropriate θ_0, θ_1 such that $\theta_0 < T < \theta_1$, we can use a simple hypothesis test to determine if the true parameter θ is above or below the threshold T, albeit with some errors

[1] A. Wald, 1947, Sequential analysis

Neyman-Pearson Notation

- Let $x = \{x_1, x_2, ..., x_{max}\}$ be a sequence of observations up to some **fixed and known maximum number of samples**
- To test the simple hypothesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ consider the likelihood ratio

$$\Lambda(x) = \frac{F(\theta_0|x)}{F(\theta_1|x)} \gtrless \eta$$

- The test declares H_0 when $\Lambda(x) > \eta$ and H_1 when $\Lambda(x) < \eta$
- We can define the standard types of error
  - Type I error
    - False rejection of H_0
    - $P(\Lambda(x) \leq \eta | H_0) = \alpha$
    - α is named the size of the test
  - Type II error
    - False acceptance of H_0
    - $P(\Lambda(x) \geq \eta | H_1) = \beta$
    - $1 - \beta$ is the power of the test
- Neyman-Pearson lemma[2] deems this the "most powerful", since the quantity $1 - \beta$ is maximized for an appropriate η

[2] J. Neyman, E. Pearson, 1933, On the Problem of the Most Efficient Tests of Statistical Hypotheses

Our Goals and Approach

Measurement resource is going to be a problem, so we've identified some goals that should scale to the low resource case.

- Get the most utility out of our limited measurement resources
- Classify channels as being "high", "medium" or "low" occupancy
- Identify "useful" channels. Channels with low probability of occupancy (low p_n)
- Identify "useless" channels. Channels with High probability of occupancy (high p_n)

With these goals in mind we'll take the following Approach:

- Build a single channel test that will classify a channel with the *minimum number of samples* and admit a low error rate
- Distribute measurements between channels by prioritizing channels that complete classification sooner (in fewer samples)

Single Channel Optimizations: (Re)Introducing the Sequential Probability Ratio Test

Here we'll depart from the classical Neyman-Pearson formulation. It has been shown[3] that for fixed α, β, the SPRT achieves the lowest expected sequence length. The defining equations are:

- Sequential Probability Ratio:

$$B < \frac{f(X_{1,n}, p_1) \ldots f(X_{m_n,n}, p_1)}{f(X_{1,n}, p_0) \ldots f(X_{m_n,n}, p_0)} < A$$

- If we take as our boundaries:

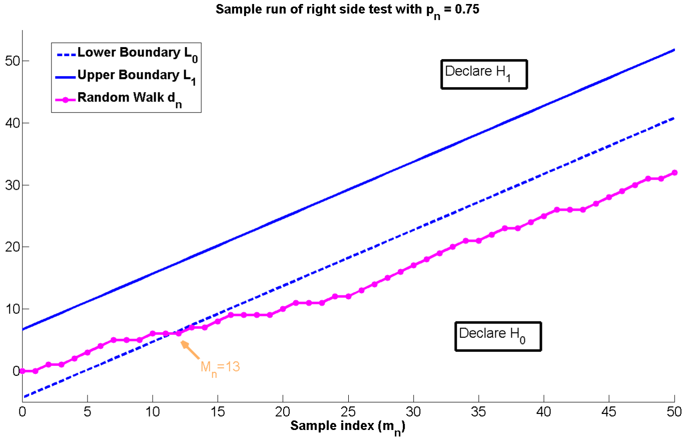
$$A = \frac{1-\beta}{\alpha} \qquad B = \frac{\beta}{1-\alpha}$$

  Wald argues that the actual error rates α', β' are related to the α, β used to build this test by the equation: $\alpha' + \beta' \leq \alpha + \beta$

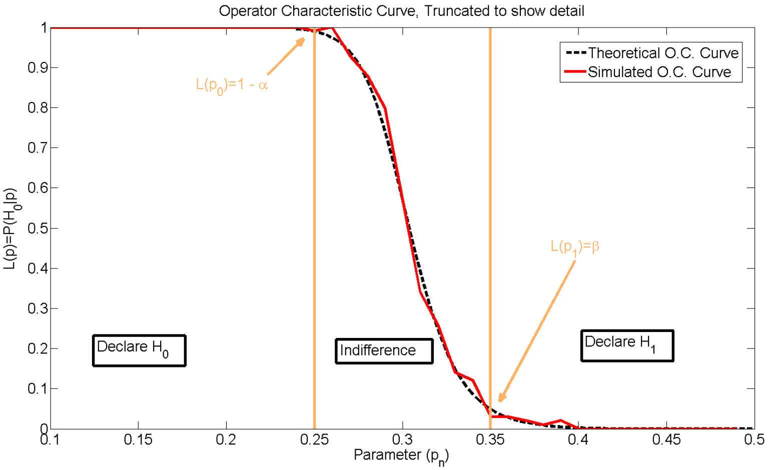
---

[3] A. Wald and J. Wolfowitz, 1948, Optimum Character of the Sequential Probability Ratio Test

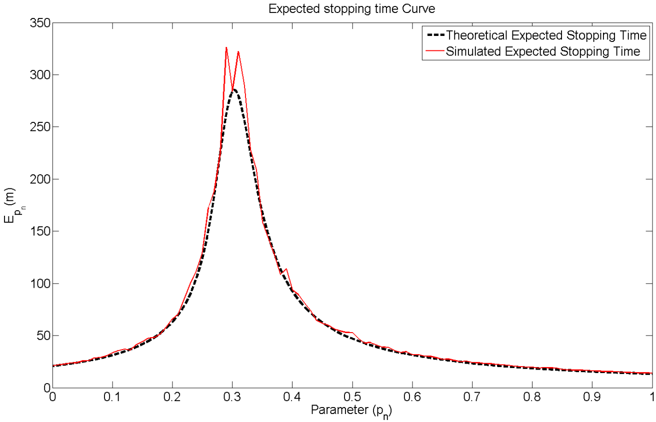
SPRT Sample Run



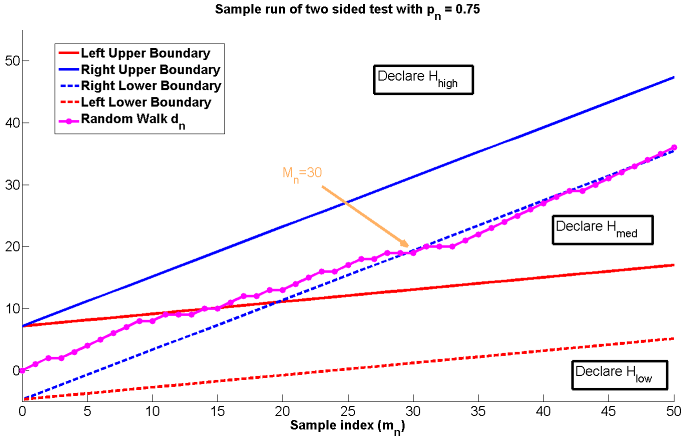
Operator Characteristic Curve: $L(p_n) = P(declare\ H_0 | p_n)$



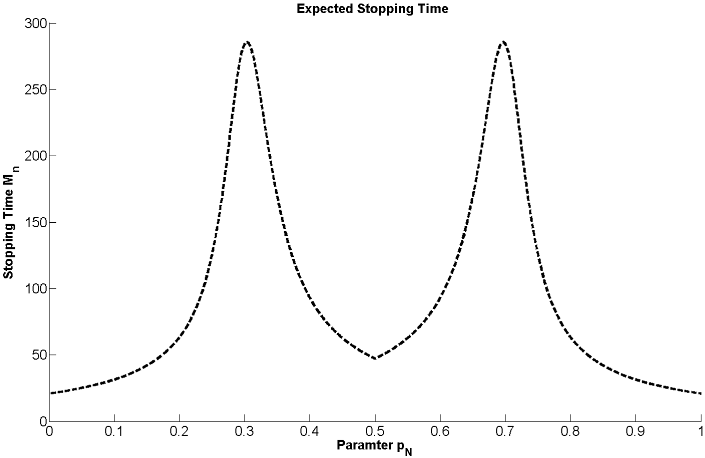
Expected Stopping Time $E_{p_n}(m) = F(L(p))$



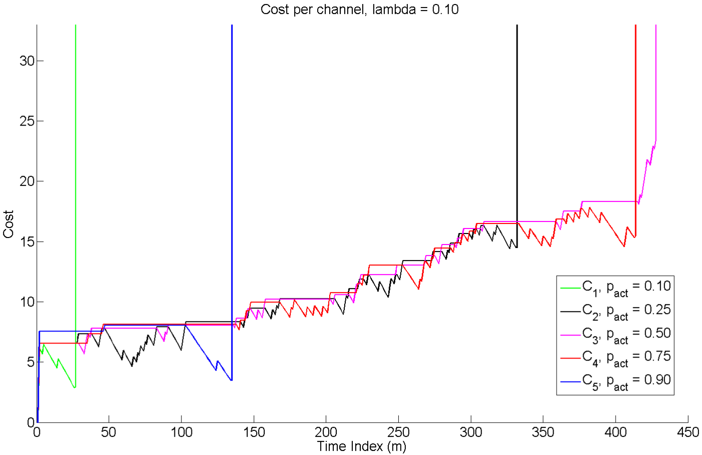
Running 2 tests at Once



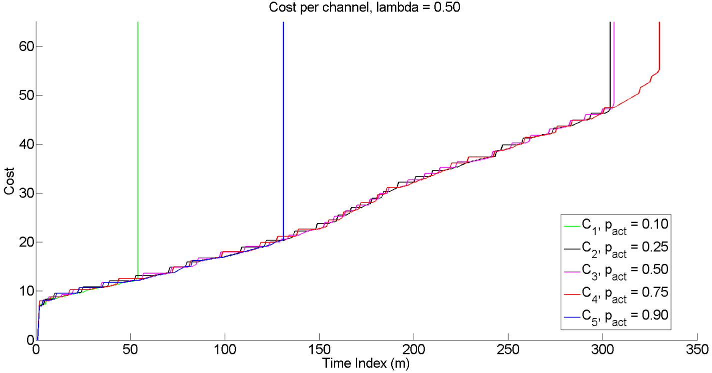
Two Test Expected Samples: $E_{p_n}(M_n) = max\{(E_{p_n}(M_n))_L, (E_{p_n}(M_n))_R\}$



Low λ 5 Channel Cost



High λ 5 Channel Cost



Scaling Performance

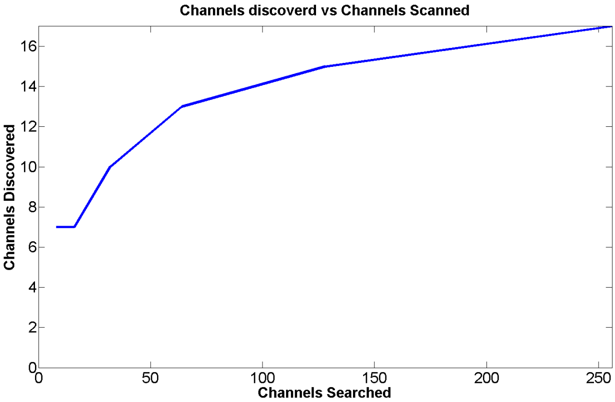
We ran several scaling simulations to determine the scaling of:

- Number classified before the $L = 1000$ deadline
- Number of samples required to completely classify all channels
- Number of samples require to find 4 good(In H_{low}) channels

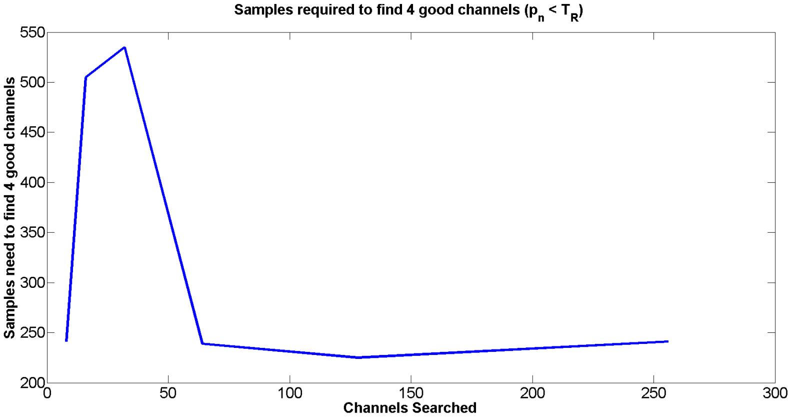
The experiment setup was as follows:

- Select N ($N \in \{8, 16, 32, 64, 128, 256\}$)channels uniformly from $[0, 1]$
- Pass these channels through the classifier
  - Calculate each channel's cost
  - identify the set $\min_n(c_n)$
  - if $|\min_n(c_n)| > 1$ choose one randomly
  - sample and update
- Preform 100 trials for each N

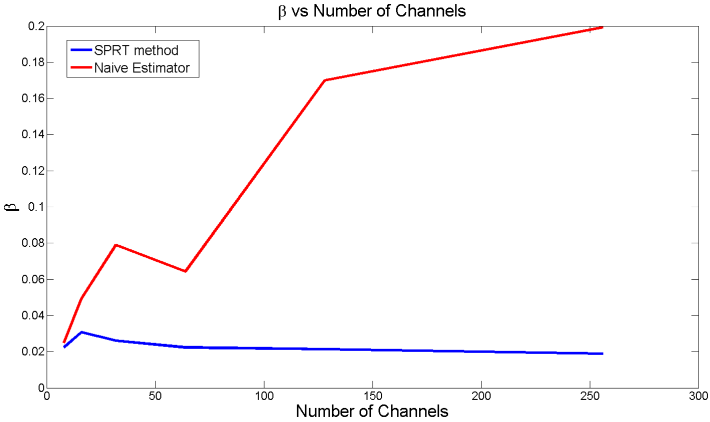
Scaling: Channels Classified vs N, $L = 1000$



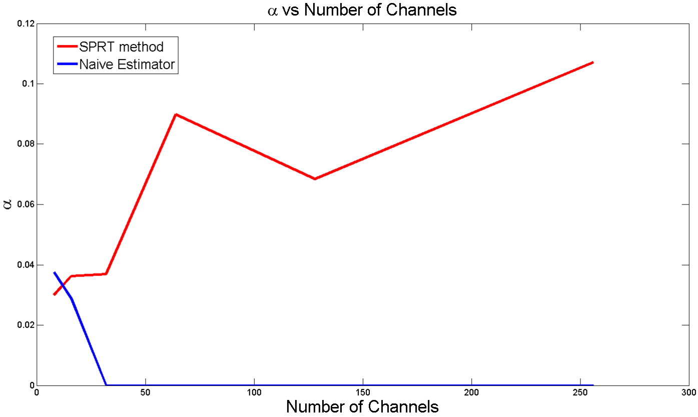
Scaling: Number of samples required to find 4 good channels vs N



Error Performance - β - Probablity that I collide



Error Performance - α - Probablity that I miss something useful



Error Performance Observations

- The SPRT's errors are tied to p_n proximity to center of the indifference region
- As the pool of good channels grows, because we prioritize channels that are closer to the edge, we make fewer mistakes before the deadline.
- The classification error rate of the full set is higher because it traverses the entire parameter space
- After we pass 16 channels the pool of available good channels is much greater than the number of channels we can discover with $L = 1000$ measurements
- As $\frac{L}{N} \searrow$, We won't be able to find more channels, but we will enumerate channels with lower p_n

Future Directions

- Tune λ, α, β for different meta distributions
- Identify values of $\frac{L}{N}$ where our method is optimal, and where it should not be used
- Compare computational complexity with the naive estimator
- Evaluate alternative cost functions tuned for different goals (e.g Be more aggressive for lower p_n by setting $e_n = (|d_{m_n,n}(m_n) - L_{L,0}(m_n)|))$

Thank You

Questions?