Symptotics: A Framework for Analyzing the Scalability of Real-World Wireless Networks

Ram Ramanathan
Raytheon BBN Technologies
ramanath@bbn.com

Abhishek Samanta
Northeastern University
samanta@neu.edu

Tom La Porta
Pennsylvania State University
tlp@psu.edu

ABSTRACT

We present a framework for non-asymptotic analysis of real-world wireless networks that captures protocol overhead, congestion bottlenecks, traffic heterogeneity and other real-world concerns. The framework introduces the definition of *symptotic* [1] *scalability*, and a metric called *change impact value* (CIV) for comparing the impact of underlying system parameters on network scalability. A key idea is to divide analysis into generic and specific parts connected via a *signature* – a set of governing parameters of a network scenario – such that analyzing a new network scenario reduces mainly to identifying its signature.

Using this framework, we present approximate scalability expressions for line, mesh and clique topologies using TDMA and 802.11, for unicast and broadcast traffic. We compare the analysis with discrete event simulations and show that the model provides sufficiently accurate estimates of scalability. Based on the symptotic expressions, we study the change impact value of underlying parameters on network scalability. We show how impact analysis can be used to tune network features to meet a scaling requirement, and determine the regimes in which reducing routing overhead impacts performance.

Categories and Subject Descriptors

C.2 **[Computer-Communication Networks]**: Miscellaneous

Keywords

Scalability, Performance Analysis, Non-Asymptotic.

1. INTRODUCTION

Consider the following problem: a multi-hop wireless network running OLSR [2] routing over 802.11 radios needs to be deployed in a roughly regular structure (Manhattan grid). Each node needs a VoIP stream to another node (say randomly chosen) and is active 20% of the time. Roughly how many nodes N can one deploy?

Such a question, with straightforward variations, may arise in a deploying a wireless network in diverse situations: deploying instant infrastructure after a disaster, a community mesh network in a rural area, a military infrastructure-less network, a sensor network, etc. Often, the answer does not need to be precise, but quick, and for a wide range of parameter combinations. Currently the only reasonable way available to answer such questions is to construct a simulation model and iterate over different values of N to find the upper bound. Not only does this take a prohibitively long time when N is large, but also requires re-running over again to answer follow up questions: what if we used a faster radio, or used different protocols in place of OLSR and 802.11, or a different encoding for VoIP is used? What if it is a different topology or a different traffic pattern? Last, but not least, simulations do not provide insight into the relationships between parameters.

Analytical modeling appears a natural fit. However, much of the analytical research thus far has been along *asymptotic* lines (e.g [3, 4, 5, 6]). While these have provided tremendous insight in the limiting case, asymptotic scalability has limited applicability to finite real-world networks. A network may be asymptotically unscalable, yet scale comfortably to the requisite number of nodes in a given deployment. Further, such work does not consider control protocols, bottleneck phenomena, and the multiplicity of traffic types that are an inherent part of real-world systems.

In this paper, we present *symptotics* – a framework for approximate non-asymptotic scalability analysis of wireless networks. Unlike asymptotic analysis that typically characterizes a network in binary terms (does it scale or not), symptotics seeks to provide a qualified answer (how many nodes does it scale to) by way of closed-form expressions. The framework defines the concept of *symptotic scalability*, and accommodates real-world concerns including protocol overhead effects, congestion bottlenecks and a multiple traffic types such as unicast and broadcast. It provides a unified approach to analytically modeling a suite of network scenarios without re-working the analysis for each, and a systematic way of determining the impact of changing a scenario parameter on performance.

Our thesis is that the performance of a network scenario is dominated by a few major factors, and by focusing only on those and abstracting away all other details, one can gain reasonable accuracy while avoiding complexity. Specifically, we divide the model into two parts – a) a generic equation for a class of network scenarios that captures the performance in terms of a set of major factors termed the *signature* of the scenario; b) instantiation of this equation using the specific signature of the given network scenario to derive a non-asymptotic (symptotic) closed-form expression for this network

[1]The word "symptotic" is not part of the English lexicon. Our inspiration comes from the commentary of Euclid's Elements by Proclus [1] ".. some are asymptotic, namely, those which however far extended never meet, and others that do intersect are *symptotic*.."

PE-WASUN'12, October 21–22, 2012, Paphos, Cyprus.

scenario. Thus, analyzing a new scenario requires doing only part (b) rather than re-working from scratch.

We illustrate the application of our framework by deriving approximate symptotic scalability expressions for 12 network scenarios resulting from combinations of three regular topologies (line, mesh, clique), two MAC protocols (TDMA, 802.11) and two traffic types (unicast, broadcast). A comparison with simulation results show that despite their simplicity, our models are adequate for the rough estimations motivated at the beginning of this section, and lend support to our thesis that reasonable accuracy can be obtained with simple, approximate models.

A valuable part of the symptotic framework is a rigorous and uniform approach to *impact analysis*, that is, which parameters affect the overall system performance the most. As part of the framework, we introduce the concept of *change impact value* (CIV) to quantify the impact of domain parameters in a uniform way. By comparing the CIVs, we can tell, for example, if halving the offered load is better or worse for scalability than halving the routing overhead. Impact analysis of even simple networks yields some interesting, counter-intuitive insights. For example, it turns out that the impact of routing overhead reduction on scalability is dwarfed by the impact of radio rate and load.

In sum, our contributions are as follows

- A new approach to scalability analysis that captures real-world aspects and enables one to roughly estimate scalability of a scenario by simply identifying its "signature".
- Closed-form symptotic scalability expressions for 12 regular network scenarios validated using simulations.
- A new impact analysis methodology that allows one to swiftly tune a network's features to meet a scaling requirement and estimate which parameter is critical in which conditions.

The rest of the paper is organized as follows. Section 3 describes the symptotic framework, and the derivation of a "master template" equation. In section 4 we illustrate the use of our framework by analyzing 12 network scenarios. In section 5 we compare the analytical model with simulation results for the same network scenarios using ns-2. Impact analysis is covered in section 6.

2. RELATED WORK

The scalability of wireless networks has mostly been studied along asymptotic, and information-theoretic lines [3, 4, 7, 6, 5]. Such asymptotic analyses seek to determine the fundamental scaling law underlying a network in terms of the order-of-growth. In [3], often considered the seminal paper in this area, the authors show that the per-node transport capacity of arbitrary wireless networks grows as $\Theta(\frac{1}{\sqrt{n}})$ indicating asymptotic un-scalability. In [4], random mobile networks are shown to scale as $\Theta(1)$, or in other words, asymptotically scalable assuming unlimited delay tolerance. Directional antennas are shown not to help in the asymptotic sense [7] whereas distributed MIMO [6] with some assumptions scales as $\Theta(1)$. In this body of work, the assessment of scalability is unqualified (e.g. "Network X does not scale"), whereas symptotics seeks a qualified assessment (e.g. "Network X with parameter set P scales to 1000 nodes"). Further, these works do not consider real-world aspects such as protocol effects or traffic heterogeneity.

There has been some recent work non-asymptotic analysis [8, 9], but these have focused on specific aspects such as spectrum sensing [8] or delay [9], or made assumptions similar to [3] and ignored protocols. Analysis of specific protocols, especially 802.11, has received a lot of attention [10, 11, 12], and specific properties of OLSR [2] have been analyzed [13]. Regular networks have been analyzed in [14, 15] for stability and capacity, albeit with specific focus on capacity regions [14] or impact of buffer sizes [15]. Related work for our impact analysis methodology is sensitivity analysis, which has been studied using simulation [16], and analytically using Automated Differentiation techniques in [17]. Some of these works consider real-world aspects, but focus on specific protocols or issues (e.g. delay) rather than the system as a whole, and are not focused on scalability. They also do not target closed-form expressions, which we seek due to their ability to provide insights and assist in impact analysis.

3. THE SYMPTOTIC FRAMEWORK

The fundamental entity for analysis in our framework is a *network scenario*, which we define as a particular instantiation of the 4-tuple: network topology (e.g. mesh, random), traffic type (e.g unicast, broadcast), node attributes (e.g. rate, number of transceivers), and control mechanisms (e.g. 802.11, OLSR).

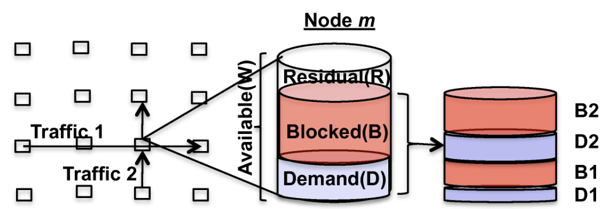

Figure 1: The different components of capacity occupation at a representative node in the network.

Consider a representative node m in the network, along with its immediate neighborhood, as shown in Figure 1. We begin with a few definitions. The *available capacity* $W(m)$ indicates the amount of data that can be handled by m. The *demanded capacity* $D(m)$ denotes the amount of data load at m required to transport a required set of flows. The *blocked capacity* $B(m)$ denotes the capacity that is unusable by node m (for example, due to contention). The *residual capacity* $R(m)$ is the difference $W(m)$ - $D(m)$ - $B(m)$.

Both D and B have components stemming from different traffic types, each of which may have a different "cast" (unicast, multicast, broadcast) and/or a different "scope" (e.g., a one hop "Hello" vs a network wide broadcast). Data, MAC control, network control, network management control etc., each have a different combination of cast and scope and are modeled as a different component of D and B. Denoting the j^{th} component with a subscript j we have

$$R(m) = W(m) - \sum_j D_j(m) - \sum_j B_j(m) \tag{1}$$

In steady state, the immediate neighborhood of m in a large network experiences approximately the same dynamics as m. Note that we do *not* assume that *all* nodes have the same load on average. For instance, in a grid network, the center node and the corner node experience quite different loads but nodes in the neighborhood of the center (or corner) have similar loads for sufficiently large networks. Thus,

$$B_j(m) = \Gamma_j(m) \cdot D_j(m) \tag{2}$$

where $\Gamma_j(m)$ is the number of nodes in the neighborhood of m that either cause interference to, or in some other way cause m to defer when they are active.

We term Γ_j the *contention factor* for component j. The contention factor $\Gamma_j(m)$ depends on the topology around m, on the medium access control protocol in use, and the kind of transmission (e.g. unicast or broadcast), which in turn depends on the traffic type j. The contention factor is related to the spatial reuse achieved in a wireless network. The higher the contention factor, the lower is the spatial reuse.

Thus, from equations 1 and 2, we have

$$R(m) = W(m) - \sum_j (1 + \Gamma_j(m)) D_j(m) \tag{3}$$

Now consider $D_j(m)$. This is the contribution to m's demanded capacity from traffic component j. Suppose that every node sources on average the same amount of traffic L_j for a given traffic component j. The contribution to $D_j(m)$ is the traffic sourced by m plus the traffic from other sources relayed by m. We call the latter (the relayed traffic) the *transit factor* of j, and denote it by Υ_j. For example, in a 5 node line network with each node flooding one packet each, 4 packets are relayed by the central node, and so its transit factor is 4. A beacon signal sent periodically by a node has a transit factor of 0 since it is not relayed. The expected demanded capacity is then $D_j(m) = L_j \cdot (1 + \Upsilon_j(m))$.

Finally, many medium access control schemes (e.g. CSMA/ CA) have non-trivial *inefficiency*, that is, the effective rate is lower than the actual rate. Thus, the actual capacity available for multi-access communications is a fraction of W. Assuming MAC efficiency is indepenent of traffic type, and denoting it by η, and based on the above discussion, we rewrite equation 3 as

$$R(m) = \eta W(m) - \sum_j (1 + \Gamma_j(m)) L_j \, (1 + \Upsilon_j(m)) \tag{4}$$

A number of assumptions have been made above, notably a) contention is the only source of "blocked" capacity; b) load in a 2-hop neighborhood is roughly the same on average; and c) every node sources the same load. These assumptions allow us to reduce the complexity of the model without compromising on the key elements to accuracy – as attested to by our validation results in section 5. Relaxing these and other assumptions could make the predictions even better and is a topic for future work.

Using equation 4, we now consider the definition of symptotic scalability and the generation of a "master template" for symptotic analysis.

3.1 Symptotic Scalability

As motivated in section 1, we seek non-asymptotic scalability, that is, to provide an answer to a question such as "how many nodes will my network scale to"? We begin by observing that such a question only makes sense for *expandable* network scenarios, that is, those that have a "scale agnostic" specification. Examples of expandable network topologies include *regular* topologies such as line, ring, mesh etc., as well as *irregular* stationary topologies based on some probabilistic model (e.g random unit-disk graphs, scale-free networks etc.), and mobile scenarios that have an expandable mobility model (e.g. the random waypoint model).

An arbitrary network with a specific set of nodes and a specific set of links between them is not expandable because there is no "rule" to generate higher-sized versions. A similar differentiation can be made with respect to traffic as well. Thus, the question "how many nodes will a network scale to" is well formed only for expandable networks, and not for arbitrary networks. On the other hand, note that computing the *capacity* of a given network is a reasonable proposition for arbitrary (non-expandable) networks.

We consider the class of expandable networks for which the residual capacity monotonically decreases as the size N increases. This class clearly includes regular networks with uniform traffic model. It also includes irregular expandable networks averaged over multiple instances, although a *particular* random network of size N may happen to have a higher residual capacity than a particular random network of size $N - 1$. An example of a network scenario not in this class is when the set of nodes sourcing traffic is constant (say 1). Network scenarios not in this class are arguably asymptotically scalable and so are not of interest for symptotics.

For such networks, there is a point at which the monotonically decreasing residual capacity of an expanding network scenario transitions to negative. This is the maximum number of nodes supportable, or the "symptotic" scalability.

More precisely, let $R_N(m)$ denote the residual capacity of node m in a network scenario with N nodes.

DEFINITION 3.1. *The* symptotic scalability *of a network scenario is the number of nodes X such that for all $N \leq X$, and for all m, $R_N(m) \geq 0$, and for all $N > X$, there exists a node m_b such that $R_N(m_b) < 0$.*

We call m_b a *bottleneck node*. Note that a network scenario may have multiple bottlenecks, that is, nodes with equally lowest residual capacity, in which case m_b is any one of them.

Setting $R(m_b) = 0$ per definition 3.1, and dropping the reference to m_b with the notion that hereinafter it is implicit, we have from equation 4

$$\eta W = \sum_j (1 + \Gamma_j) L_j \cdot (1 + \Upsilon_j) \tag{5}$$

where η is the efficiency, W is the available capacity (radio rate), Γ_j is the contention factor for traffic type j, L_j is the average offered (sourced) load (in bps) per node for type j, and Υ_j is the transit factor for traffic j.

Equation 5 may be considered the basic "master template" that we further instantiate on a per-scenario basis. This requires us to further expand the above parameters to generate the expression characterizing the performance for that system. Specifically, the contention factor Γ_j and the transit factor Υ_j for each j play a critical role in the performance, and will be referred to as the *signature* of the system. To estimate the performance of a given system, one simply has to identify the signature and plug it into the master template. For asymptotically unscalable[2] networks, the transit and/or the contention factors are a function of N and hence equation 5 is of the form $\eta W = f(N, L_j)$, which can then be solved for N. In section 4, we give several detailed examples of how to identify the signature of a network scenario and instantiate and solve the master template.

In contrast to previous works (e.g. those mentioned in section 2), equation 5 can capture a multiplicity of traffic types that a typical real-world system has. For instance, consider a network in which each node generates VoIP as well as web-browsing traffic, each with a different source rate and destination distribution, and additionally a network-wide "situational awareness" broadcast traffic. Each of these can be modeled with a separate signature (Γ_j, Υ_j). Further, control traffic is accommodated simply as yet another traffic type.

[2] The typical multi-hop wireless network is asymptotically unscalable [3]. While our framework can accommodate asymptotically scalable networks as well, these are largely uninteresting from a symptotic viewpoint as the scalability is infinite nodes.

4. SYMPTOTIC ANALYSIS OF NETWORK SCENARIOS

We consider 12 network scenarios generated from all possible combinations of the following: three topologies – *line*, degree-4 *mesh* (Manhattan grid) and *clique* (complete graph); two traffic types – *unicast* and *flooding* (network-wide broadcast); and two MAC protocols – *TDMA* (Time Division Multiple Access) and IEEE *802.11* DCF.

In all cases, we only have one type of data traffic and two types of control traffic – routing advertizements and beacons ("Hello" broadcasts). In other words, with reference to equation 5, $j = 3$. Thus, for the purposes of this section, the specific master equation we consider is

$$\begin{aligned}\eta W &= (1+\Gamma_d)L_d(1+\Upsilon_d) + (1+\Gamma_l)L_l(1+\Upsilon_l) + \\ &\quad (1+\Gamma_h)L_h(1+\Upsilon_h) \qquad (6)\end{aligned}$$

where Γ_d, $\Gamma_l$and $\Gamma_h$denote the contention factors for data, OLSR link-state updates (LSUs) and OLSR Hellos respectively, Υ_d, Υ_l, and $\Upsilon_h$denote the transit factors for data, LSUs and Hellos respectively, and L_d, L_l and L_h denote the offered load per node for each traffic type respectively.

The loads L_d, L_l and L_h are calculated based on the sourced packets per second (pps) for each of data (λ_d), LSU frequency (λ_l) and Hello frequency (λ_h) respectively, the payload sizes, and the network and MAC layer headers. The header lengths and variables we use for the rest of this section are summarized in Table 1. To avoid too many leading or trailing zeros, all rates are in units of kbps (10^3 bits/sec).

For network-wide broadcast (flooded) packets such as OLSR link state advertizements and flooded data, we assume a single broadcast transmission at the link layer by each node[3].

The symptotic scalability expressions for the 12 scenario combinations are summarized in Table 2 (for TDMA-based scenarios), and Table 3 (for 802.11 based scenarios). In each table, the first column indicates the specific scenario, the next six columns are the signature for the scenario and the last column is the scalability expression derived by substituting the signature into the master template (eq 6) and solving symbolically for N using the Sage mathematical and symbol manipulation software [18]. Note that the transit factors are identical because the MAC protocol does not affect how many paths go through a node.

Deriving the signature consists of first identifying the bottleneck node, if relevant, and then analyzing the contention and transit factors. Below, we illustrate the derivation of the first two rows in Table 2 (namely, TDMA Line network flooding and unicast). For lack of space, we are unable to describe the other derivations, but the complete details can be found in [19].

4.1 Example: Line network running TDMA

We consider a spatial-reuse TDMA model with node scheduling, also referred to a broadcast scheduling [21] and used in operational systems, for example [22]. That is, time is slotted and slots are grouped into repeating frames. Every node is assigned a slot in a frame in which it is allowed to transmit and its neighbors receive. Thus, nodes that are neighbors or share a common neighbor should be assigned different slots. The goal of a TDMA protocol is to perform conflict-free assignment using the least possible number of slots. Our model captures the control overhead for doing so by means of an additional "control slot" which we assume is large enough to hold any necessary messages for managing the slot assignment process.

A line network can be node scheduled using 3 slots, for example, using slot numbers 1, 2, and 3 repeating from left to right on the line. Thus, a typical node has to defer for nodes transmitting in other slots than its own and the control slot, and thus the contention factor is 3 for both link-layer unicast and broadcast. Since all traffic uses one of these modes, all of the contention factors are 3.

Consider the transit factor (TF). For broadcast flows (flooding), all nodes except the one at the ends have equal load. For unicast flows the node at the center (two nodes if N is even) is the bottleneck. Without loss of generality, assume N is odd and take the center node (say b) as the node of interest for all cases.

The TF of Hello's $\Upsilon_h$is clearly zero as it is single hop. The TF for LSUs $\Upsilon_l$is $N-1$ since every other node's LSU is transmitted by this node. Similarly, for flooded data, Υ_d= $N-1$. For the TF of unicast data, we need to compute the expected number of paths that go through b. The probability that a given node routes through b is the probability that the destination lies on the "other side" of b, that is, $p(B)$ = $\frac{(N-1)/2}{N-2}$. Thus, the expected number of paths is $p(B)\cdot(N-1)$ as shown.

We now perform the next step in the symptotics approach, namely to substitute the values from the first two rows of Table 2, along with the constants from Table 1 into equation 6. For flooding (row 1), we get

$$\frac{4}{125}(B+48)N\lambda_d + \frac{64}{25}N\lambda_l + \frac{304}{125}\lambda_h = W\eta \qquad (7)$$

Solving for N, we get the symptotic scalability of a line network running TDMA, OLSR, with network-wide broadcast traffic as

$$N = \frac{125\,W\eta - 304\,\lambda_h}{4\,(B\lambda_d + 48\,\lambda_d + 80\,\lambda_l)} \qquad (8)$$

Similarly, for unicast traffic (row 2), the equation is

$$\frac{2}{125}(N+1)(B+48)\lambda_d + \frac{64}{25}N\lambda_l + \frac{304}{125}\lambda_h = W\eta \qquad (9)$$

Solving for N, the symptotic scalability of a line network running TDMA, OLSR, with random unicast traffic is

$$N = \frac{125\,W\eta - 2\,B\lambda_d - 96\,\lambda_d - 304\,\lambda_h}{2\,(B\lambda_d + 48\,\lambda_d + 160\,\lambda_l)} \qquad (10)$$

The above illustrates the approach and steps involved in deriving the expression within our framework for the simple example of a line. The derivations for other rows in the two tables are similar to the above in spirit, although the computation of transit factor can get a bit involved for more complicated networks. Unfortunately, due to lack of space we are not able to describe those derivations. We

OLSR	802.11	Variables
LSU : 52 bytes	RTS : 20 bytes	Data pps: λ_d
Hello : 48 bytes	CTS : 14 bytes	LSU pps: λ_l
Net Hdr: 20 bytes	ACK : 28 bytes	Hello pps: λ_h
	MAC Hdr: 28 bytes	Datarate: W
		Efficiency: η
		Datasize: B

Table 1: Protocol header sizes from [2, 20] and notation for free variables used in analysis.

[3]An alternate model/assumption would be multiple unicast transmissions at the link layer, and can also be easily analyzed with our framework if necessary.

	Contention Factor			Transit Factor			Scalability Expression
	Γ_d	Γ_l	Γ_h	Υ_d	Υ_l	Υ_h	$N =$
Line Flooding	3	3	3	N-1	N-1	0	$\frac{125\,W\eta-304\,\lambda_h}{4\,(B\lambda_d+48\,\lambda_d+80\,\lambda_l)}$
Line Unicast	3	3	3	$\frac{(N-1)^2}{2(N-2)}$	N-1	0	$\frac{125\,W\eta-2\,B\lambda_d-96\,\lambda_d-304\,\lambda_h}{2\,(B\lambda_d+48\,\lambda_d+160\,\lambda_l)}$
Mesh Flooding	5	5	5	N-1	N-1	0	$\frac{125\,W\eta-456\,\lambda_h}{6\,(B\lambda_d+48\,\lambda_d+80\,\lambda_l)}$
Mesh Unicast	5	5	5	$0.4(1+\frac{2}{\sqrt{N}})(N^{\frac{3}{4}}+4N^{\frac{1}{4}})$	N-1	0	SEE CAPTION
Clique Flooding	N-1	N-1	N-1	0	0	0	$\frac{125\,W\eta}{B\lambda_d+48\,\lambda_d+76\lambda_h+80\,\lambda_l}$
Clique Unicast	N-1	N-1	N-1	0	0	0	$\frac{125\,W\eta}{B\lambda_d+48\,\lambda_d+76\lambda_h+80\,\lambda_l}$

Table 2: Signature set and symptotic scalability expressions for regular networks using TDMA (see below and [19] for details). The scalability for mesh unicast is $N = \frac{\left(3\,B\lambda_d-\sqrt{P+Q}\sqrt{3}+144\,\lambda_d\right)^2}{230400\,\lambda_l^2}$ **where** $P = 3\,B^2\lambda_d^2 + 20000\,W\eta\lambda_l - 3840\,(12\,\lambda_d + 19\,\lambda_h)\lambda_l$ **and** $Q = 96\left(3\,\lambda_d^2 - 10\,\lambda_d\lambda_l\right)B + 6912\,\lambda_d^2$

	Contention Factor			Transit Factor			Scalability Expression
	Γ_d	Γ_l	Γ_h	Υ_d	Υ_l	Υ_h	$N =$
Line Flooding	2	2	2	N-1	N-1	0	$\frac{125\,W\eta-228\,\lambda_h}{3\,(B\lambda_d+96\,\lambda_d+80\,\lambda_l)}$
Line Unicast	3	2	2	$\frac{(N-1)^2}{2(N-2)}$	N-1	0	$\frac{2\,B\lambda_d+125\,W\eta-192\,\lambda_d-228\,\lambda_h}{2\,(B\lambda_d+96\,\lambda_d+120\,\lambda_l)}$
Mesh Flooding	4	4	4	N-1	N-1	0	$\frac{25\,W\eta-76\,\lambda_h}{B\lambda_d+96\,\lambda_d+80\,\lambda_l}$
Mesh Unicast	7	4	4	$0.4(1+\frac{2}{\sqrt{N}})(N^{\frac{3}{4}}+4N^{\frac{1}{4}})$	N-1	0	SEE CAPTION
Clique Flooding	N-1	N-1	N-1	0	0	0	$87.41\,\left(\frac{W\eta}{B\lambda_d+96\,\lambda_d+76\,\lambda_h+80\,\lambda_l}\right)^{0.92}$
Clique Unicast	N-1	N-1	N-1	0	0	0	$87.41\,\left(\frac{W\eta}{B\lambda_d+96\,\lambda_d+76\,\lambda_h+80\,\lambda_l}\right)^{0.92}$

Table 3: Signature set and symptotic scalability expression for regular networks using 802.11 (see [19] for details). The scalability expression for mesh unicast is $N = \frac{\left(B\lambda_d+96\,\lambda_d-\sqrt{P+Q}\right)^2}{10000\,\lambda_l^2}$ **where** $P = B^2\lambda_d^2 + 3125\,W\eta\lambda_l - 100\,(192\,\lambda_d + 95\,\lambda_h)\lambda_l$ **and** $Q = 8\left(24\,\lambda_d^2 - 25\,\lambda_d\lambda_l\right)B + 9216\,\lambda_d^2$

refer the interested reader to [19] where the derivation is described in detail.

5. VALIDATION

In this section we present results of ns-2 simulations of some of the scenarios analyzed in section 4. We instantiate the corresponding symptotic expressions given there and compare them with simulation results for the same set of parameter values (refer Table 1).

The TDMA protocol is an extension of the TDMA model available in the ns-2 distribution. Specifically, we have extended it to a spatial reuse TDMA, that is, one that allows multiple nodes to transmit in the same slot. We then implemented a 3-slot assignment for line networks and a 5-slot assignment for mesh networks as described in section 4.1. The TDMA in ns-2 already models a control slot, which we have retained.

For 802.11, we have used the ns-2 model from an overhaul that significantly improves on the original model [23]. The key features include cumulative SINR computation, preamble and PLCP header processing and capture, and frame body capture. The MAC accurately models the basic IEEE 802.11 CSMA/CA mechanism, as required for credible simulation studies. The model implements models of four modulation schemes – BPSK, QPSK, 16-QAM, 64-QAM – with 1/2 coding rate for the first three and 3/4 for 64-QAM to provide four data rates: 6 Mbps, 12 Mbps, 24 Mbps, and 54 Mbps.

The physical layer propagation model is the *two-ray ground* model that is part of the WirelessPHYExt protocol for both TDMA and 802.11. The physical layer settings gives a transmission range of 450m. To create a line topology, we place nodes separated by a distance of slightly less than 450m such that only adjacent nodes are within range. Similarly, for a mesh, only nodes adjacent in 4 directions are within range. A clique is formed by ensuring that all nodes are within 450m of each other. The simulation duration is 30 seconds, sufficient for static networks.

We have studied the symptotic scalability N_{max} as a function of pps $\lambda_d$for each of the 12 network scenarios described in section 4. Per definition 3.1, at $N > N_{max}$ the residual capacity of at least one node is less than zero. At this point, the input rate on the node's transmit queue is more than the output, and the queue becomes unstable (that is, starts growing continuously). The ns-2 simulation system has a finite queue length – therefore, queue instability is detected as packets being dropped due to queue being full. In our simulations, the queue length is 50 packets, and we deem the network saturated if there are non-trivial queue drops, in particular, if there are 50 or more packets dropped.

Thus, we run simulations with increasing size N_i till we encounter two consecutive N_i and N_{i+1} such that there are no queue drops in N_i and there are non-trivial (> 50) queue drops in N_{i+1}. We then measure the symptotic scalability as the average of N_i and N_{i+1}. For example, if there are no queue drops for N=40, and non-trivial drops for N=50, the symptotic scalability is N=45. For the line and clique networks, the increment was 10 nodes and for a mesh, the side was incremented by 1 (i.e., the sizes were 16, 25, 36, 49 and so on).

We have considered alternate measures such as a sudden drop in throughput or increase in delay. The former is fairly unreliable especially for 802.11 networks where collisons cause loss. The latter correlates well with the queue drop measurement, but harder to objectively measure, and hence we have used queue drops as an indication of saturation.

Figures 2, 3, 4, and 5 compare the symptotic scalability predicted by our model with simulation results. Since the simulations take a long time to run for larger sizes, we picked the radio rate and packets

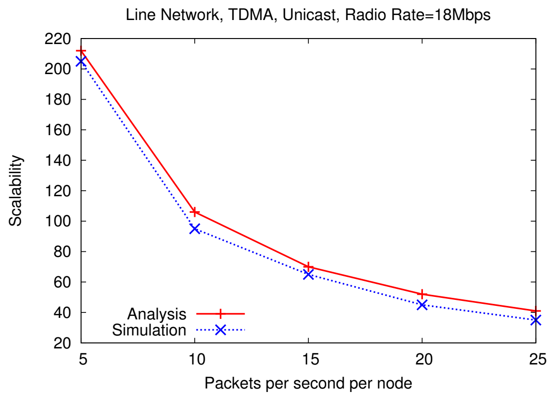

Figure 2: TDMA, Line, Unicast, with 18 Mbps radio

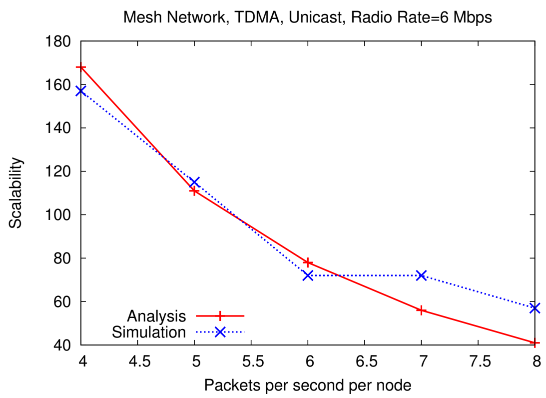

Figure 3: TDMA, Mesh, Unicast, 6 Mbps radio

per second such that the y-axis maximum is not prohibitively large. Further, due to space constraints, we have only shown a subset of the plots, picking a subset such that each topology, MAC and traffic type is represented at least once.

Our results show that despite its simplicity and abstraction of details, the scalability predicted by our model matches that predicted by simulations fairly well, and adequately for practical purposes of estimating the rough order of magnitude as motivated in section 1.

The slight discrepancy between the analysis and simulation is due to several factors. First, the analysis is approximate by design, and in particular due to simplifications made in the course of the derivations (e.g. assuming $N-1 \approxeq N$), the accuracy is lower for smaller N. On the other hand, we cannot compare with high N because simulations cannot scale to large sizes. Second, for unicast mesh results,

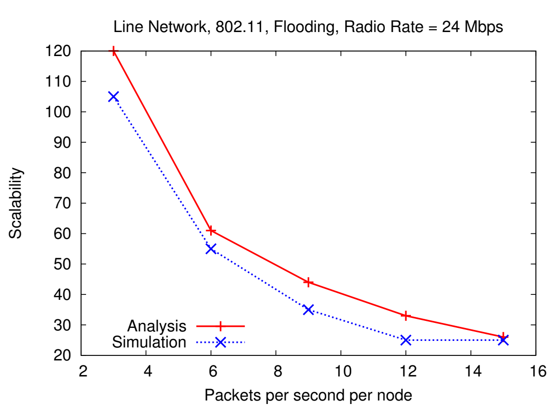

Figure 4: 802.11, Line, Flooding, 24 Mbps radio

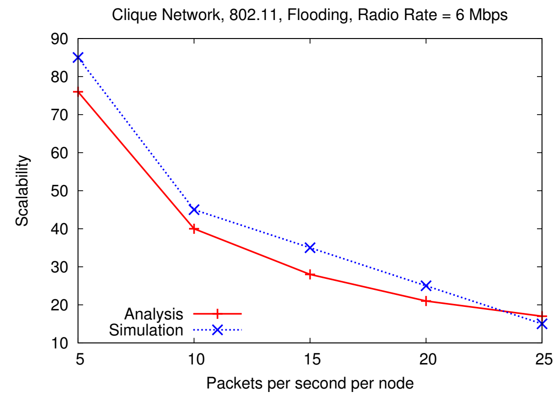

Figure 5: 802.11, Clique, Unicast, 6 Mbps radio

we have assumed that OLSR picks randomly from amongst shortest paths, which may not be the case. Third, for CSMA results, the random processes are only captured in the aggregate. Finally, due to simulation running time constraints, the node step size granularity is high. In particular, in Figure 3 the derivation for mesh unicast in Appendix A is really valid only for large N and therefore we see higher discrepancy at lower N.

The simulation results attest to the fact that the approach taken and approximations and assumptions made do not excessively compromise accuracy. The differences seen are within reason, and the slopes parallel for the most part. Our simulation study gives us confidence that our model, despite its simplicity, is adequate for rough order of magnitude scalability predictions. In the next section, with a validated model in hand, we turn our attention to impact analysis.

6. IMPACT ANALYSIS

In this section, we introduce a methodology for analyzing the impact of parameters such as routing overhead, radio rate, and offered load on scalability, and illustrate how it can be used to drive design choices to meet a scaling requirement. We then study how the impact changes with nominal values of network parameters.

6.1 Change Impact Value

In section 4, we derived several expressions of the form $N = f(X)$ where $X = (x_1, x_2, \ldots, x_n)$ is the parameter vector on which N depends (see for example equation 8 in section 4).

Let V be the *nominal instantiation* of X, with $(x_1 = v_1, x_2 = v_2, \ldots, x_n = v_n)$. Let $V_{x_j=k}$ represent that parameter x_j is instantiated with value k in V

DEFINITION 6.1. *The* Change Impact Value $CIV(x_j, \alpha)$ = $\frac{f(V_{x_j=\alpha \cdot v_j})}{f(V)}$.

In other words, $CIV(x_j, \alpha)$ is the factor change in scalability between using a particular (nominal) value for x_j and using α times that value.

For example, in the expression $N = W - C$, where V = (W = 100, C = 10), $CIV(W, 2)$ denotes the impact of doubling W on N and is the ratio between $2W - C$ and $W - C$ for nominal parameters V and is 190/90 = 2.11.

The CIV depends upon the choice of α for the particular parameter. Increasing a parameter may increase or decrease the scalability. A parameter whose increase increases the scalability is called *positively aligned* and one that decreases the scalability is called *negatively aligned*. To compare the impact uniformly, we shall choose $\alpha_p > 1$ for positively aligned parameters and $1/\alpha_p$ for negatively aligned parameters. In particular, for the remainder of this section, we will use $\alpha = 2$ for W, and $\alpha = 0.5$ for λ_l. For brevity, we shall often use "impact" to refer to the CIV.

A natural question is: isn't this equivalent to simply taking the differential coefficient and evaluating/instantiating it? To see why this won't serve our needs, consider two expressions $y = 1 + x$, and $y = 1000 + x$, and the nominal value of x is 10. In both cases, dy/dx is 1, but the impact of doubling x is much lower for the second equation (CIV=1.01) than the first (CIV=1.91), since the nominal value of the constant is much bigger in the second. The magnitude of nominal values play a crucial role in impact, and our approach is tailored to accommodate that in the simplest possible manner.

6.2 Using CIV for Network Design

We begin by illustrating how impact analysis can be used to tune a system's features to meet a scalability requirement. Consider the following example scenario: A mesh network has to be deployed in a very remote area to provide VoIP phones, mostly for a community of about 2000 houses to talk amongst themselves. The going-in architecture is to use OLSR over 802.11b radios (max rate 11 Mbps), a G.711 VoIP codec (174 kbps duplex, 132 pps, 160 byte packets including network- layer headers) [24]. For lack of any other information, traffic is assumed to be random unicast, and we assume a conservative active time of 20%.

Plugging these parameters into the symptotic equation in Table 3 for 802.11-based mesh unicast (row 4), the scalability turns out to be only 116 nodes. To increase the scalability, we can either go to higher rate radios, a more efficient codec, or reduce routing overhead via a better routing protocol. We apply impact analysis from section 6.1, and compute the CIVs. Using $\alpha = 2$ for radio rate (W) and $\alpha = 0.5$ for source packets-per-second (λ_d) and routing overhead pps (λ_l)), we have the CIVs as: [4]

CIV(W) = 2.965; CIV(λ_d) = 2.811; CIV(λ_l) = 1.020

Thus, radio rate W and source rate $\lambda_d$are about equally dominant, providing approximately a 3x increase in scalability. However, source rate is more easily changed by using a better codec for very little tradeoff in clarity. Suppose we pick the G.723 codec (33 kbps duplex, 66 pps, 64 byte packets) instead with everything else remaining the same. This is factor of 5 lower in load and therefore we should expect it to scale much more than 2.811*116 = 326 nodes. Recalculating, the scalability is now 630 nodes, which is a vast improvement, but still well below the target 2000. Now, the CIVs are

CIV(W) = 2.588; CIV(λ_d) = 2.382; CIV(λ_l) = 1.000

Notice that depending upon the values of other nominal parameters, the impact of changing a parameter is different. In particular, the impact of changing W is now non-trivially higher. Given that we have already reduced λ_d, we turn to 802.11g radios with a maximum rate of 54 Mbps (a factor of $\tilde{5}$ improvement). With this, the scalability is about 2507 nodes which meets the requirement. Now the CIVs are

CIV(W) = 2.502; CIV(λ_d) = 2.231; CIV(λ_l) = 1.014

[4]Technology choices offer a range of factor-of improvements, and rather than pick different α's for each, we have simply picked the smallest integer factor, namely 2, for simplicity. The relative CIVs for $\alpha = 2$ should adequately capture the relative CIVs with other α's for our purposes.

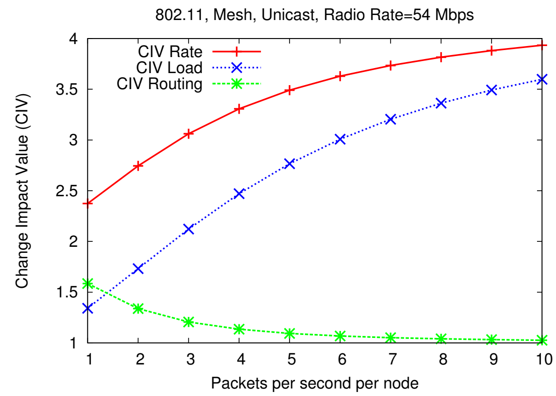

Figure 6: Comparing the impact of changing radio rate, source load and overhead on scalability (higher CIV means more impactful)

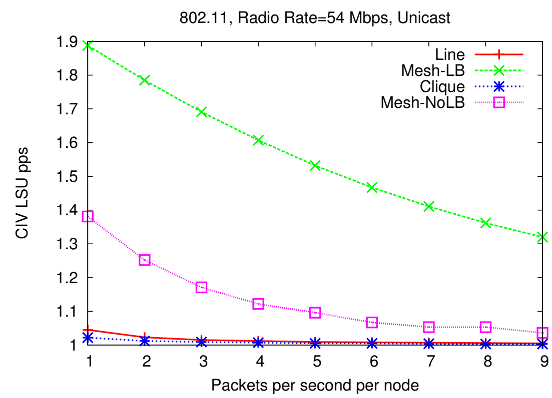

Figure 7: Impact of reducing routing overhead on scalability as a function of load for line, clique and mesh topologies, with load balanced and non-load-balanced versions (higher CIV means more impactful).

This suggests that if we could further double the radio rate, we might reach 2.5*2507 = 6250 nodes.

In all of the above, we notice that the CIV(λ_l), that is, the impact of routing load is negligible. While it is not surprising that radio rate and load are more impactful, we were surprised by the *magnitude* of the difference.

Given the amount of work that has gone into innovative techniques for reducing routing overhead in a multi-hop network, it would be helpful to know in what regimes, if any, routing overhead has a high impact on scalability. Accordingly, we present a study of the CIV as a function of various nominal parameters.

Figure 6 shows the CIVs of radio rate, source load and overhead over a range of traffic loads. For a vast majority of this space, the impact of doubling radio rate or halving source load is vastly higher than halving overhead. The gap narrows at very low loads where reducing overhead becomes about equally important as halving offered load. This is because at low loads, the network scales to larger sizes in which routing overhead (which grows as $O(N)$ per node) occupies a larger fraction of the capacity. On the other hand, this is the regime in which scalability is at its highest, and one may not need the increase quite as much as at higher loads.

Figure 7 shows that the impact of overhead reduction is much higher for a regular mesh than for line or clique, and higher when load balancing is used. This is because the lower the effective load on the bottleneck, higher is the fraction of LSUs, and consequently more impactful it is.

We have gained a number of insights from impact analysis. First, as we have shown, computing the CIVs can assist in deciding which of several options provides the designer the most benefit for a given cost. Second, it appears that for most network scenarios the impact of radio rate and load on scalability is far higher than that of control overhead. Further, the absolute value of CIV for typical networks is close to 1, which means reducing it is largely ineffective for scalability[5] Third, topologies with more opportunities to spread out traffic (e.g. mesh with unicast) offer more gain from reducing overhead, reducing load or increasing radio rate. In particular, load balancing can significantly amplify the impact of routing protocol efficiency gains.

In the past, much emphasis has been placed on ideas for reducing routing overhead. In the big picture, though, improving the radio rate and codec technologies is much more impactful than overhead reduction in real-life wireless networks. On the other hand, load balancing has received scant attention relatively and holds the potential, in certain network scenarios for better impact.

7. CONCLUDING REMARKS

Considerable effort has been expended over the last couple of decades in designing wireless networks such as sensor, mesh and ad hoc networks. However, the deployment of such networks requires an understanding of their non-asymptotic scalability, and tools for performance prediction and impact analysis. To build useful tools, we first need a validated framework that can work for a large number of potential real-world scenarios in a unified manner. We have presented such a framework and shown that it can capture a variety of topologies, protocols and traffic types and congestion phenomena thereof. We have validated our model's predictions using simulations. Our novel impact analysis enables designers to pick the most effective technology features, and has given us insight on which research pursuits might be more impactful.

In ongoing work, we are extending the framework to *irregular* expandable network scenarios such as random and mobile networks. Other future directions include refining the framework by relaxing some assumptions, and deriving signatures for other topologies (e.g. hierarchical mesh), other MAC and network protocols (e.g. busy tone, AODV), and node architectures (e.g. multi-radio cognitive networks) and combinations thereof. The rudimentary impact analysis approach presented here, while useful, ought to be developed into a more sophisticated impact analysis theory. An implementation of these models in a publicly available tool for performance prediction is another fruitful future direction. Finally, we would hope for a collaborative extension of this framework toward one that is more powerful, and that includes a growing "library" of signatures and the associated analyses. The open source Sage environment [18], in which we have coded all of the framework, and which we shall publish on the Internet, can facilitate such "crowd sourced" evolution toward a publicly available tool.

8. ACKNOWLEDGMENTS

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

[5] Although our impact analysis has not considered mobile networks, we have used an LSU source rate of 0.2 LSUs per second which captures mobilities with link dyanmics of up to once every 5 seconds.

9. REFERENCES

[1] G. R. Morrow, *Proclus: A commentary on the first book of Euclid's Elements*. Princeton University Press, 1970.

[2] T. Clausen and P. Jacquet, "Optimized link state routing," in *IETF RFC 3626*, 2003.

[3] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388–404, 2000.

[4] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, pp. 477–486, 2002.

[5] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *Information Theory, IEEE Transactions on*, vol. 53, pp. 1009 –1018, March 2007.

[6] A. Ozgur, O. Leveque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3549–3572, 2007.

[7] S. Yi, Y. Pei, and S. Kalyanaraman, "On the capacity improvement of ad hoc wireless networks using directional antennas," in *In 4th ACM MobiHoc*, pp. 108–116, 2003.

[8] F. Penna, R. Garello, D. Figlioli, and M. Spirito, "Exact non-asymptotic threshold for eigenvalue-based spectrum sensing," in *Proc. CROWNCOM '09*, pp. 1 –5, June 2009.

[9] J. Liebeherr, A. Burchard, and F. Ciucu, "Non-asymptotic delay bounds for networks with heavy-tailed traffic," in *Proc. IEEE INFOCOM*, pp. 1 –9, March 2010.

[10] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *Selected Areas in Communications, IEEE Journal on*, vol. 18, pp. 535 –547, mar 2000.

[11] J. Jangeun, P. Peddabachagari, and M. Sichitiu, "Theoretical maximum throughput of ieee 802.11 and its applications," in *Proc. of NCA 2003*, June 2003.

[12] A. Jindal and K. Psounis, "The achievable rate region of 802.11-scheduled multihop networks," *IEEE/ACM Trans. Netw.*, vol. 17, pp. 1118–1131, August 2009.

[13] D. Nguyen and P. Minet, "Analysis of mpr selection in the olsr protocol," in *Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on*, vol. 2, pp. 887 –892, may 2007.

[14] G. Mergen and L. Tong, "Stability and capacity of regular wireless networks," *Information Theory, IEEE Transactions on*, vol. 51, pp. 1938 – 1953, June 2005.

[15] G. Barrenetxea, B. Berefull-Lozano, and M. Vetterli, "Lattice networks: capacity limits, optimal routing, and queueing behavior," *IEEE/ACM Trans. Netw.*, vol. 14, pp. 492–505, June 2006.

[16] N. Zhang and A. Anpalagan, "Sensitivity of swan qos model in manets with proactive and reactive routing: a simulation study," *Telecommunication Systems*, vol. 44, pp. 17–27, 2010.

[17] J. S. Baras, S. Perumal, V. Tabatabaee, K. Somasundaram, and P. Purkayastha, "Loss network models and multiple metric performance sensitivity analysis for mobile wireless multi-hop networks," in *Proc. WICON '08*, 2008.

[18] W. Stein *et al.*, *Sage Mathematics Software (Version 4.7.2)*. The Sage Development Team, 2011. `http://www.sagemath.org`.

[19] R. Ramanathan, A. Samanta, and T. L. Porta, "An approximate model for analyzing real-world wireless network scalability," Tech. Rep. TM-2031, Raytheon BBN Technologies, 2012. http://www.ir.bbn.com/ ramanath/pdf/symptotics-techreport.pdf.

[20] M. Gast, *802.11 Wireless Networks: A Definitive Guide*. O'Reilly Publications, 2005.

[21] S. Ramanathan, "A unified framework and algorithm for channel assignment in wireless networks," *Wirel. Netw.*, vol. 5, pp. 81–94, March 1999.

[22] C. Young, "Usap: A unifying dynamic distributed multichannel tdma slot assignment protocol," in *Proc. IEEE MILCOM*, Oct 1996.

[23] Q. Chen, F. Schmidt-Eisenlohr, D. Jiang, M. Torrent-Moreno, L. Delgrossi, and H. Hartenstein, "Overhaul of ieee 802.11 modeling and simulation in ns-2," in *Proceedings of MSWiM '07*, pp. 159–168, 2007.

[24] "Tech info: Voip codecs." [Online; accessed 15-January-2012].